



HAL
open science

Caractérisation du facteur de transcription Shavenbaby par approches génomiques chez la drosophile

Alexandra Mancheno-Ferris

► **To cite this version:**

Alexandra Mancheno-Ferris. Caractérisation du facteur de transcription Shavenbaby par approches génomiques chez la drosophile. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université Paul Sabatier - Toulouse III, 2021. Français. NNT : 2021TOU30208 . tel-03700974

HAL Id: tel-03700974

<https://theses.hal.science/tel-03700974>

Submitted on 21 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

Délivré par l'Université Toulouse 3 - Paul Sabatier

**Présentée et soutenue par
Alexandra MANCHENO-FERRIS**

Le 22 octobre 2021

**Caractérisation du facteur de transcription Shavenbaby par
approches génomiques chez la Drosophile**

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **BIO-INFORMATIQUE, GENOMIQUE ET BIOLOGIE DES SYSTEMES**

Unité de recherche :

MCD - Molecular, Cellular and Developmental Biology Unit

Thèse dirigée par

Cedric POLESELLO et François PAYRE

Jury

Mme Mounia Lagha, Rapporteur
M. Frédéric Bantignies, Rapporteur
M. Krzysztof Jagla, Rapporteur
M. François Payre, Examineur
Mme Julie Caramel, Examinatrice
M. Benoit Ballester, Examineur
Mme Kerstin Bystriky, Examinatrice
M. Cedric POLESELLO, Directeur de thèse

« - ¿ Eligir no vivir por el miedo a la caída

es de sensatos o de cobardes ?

- Es de sensatos, pero no veo mayor

cobardía que la sensatez. »

– David Martínez Álvarez

« Science sans conscience n'est que ruine de l'âme »

– Rabelais

Table des matières

Table des illustrations	3
Résumés	4
RESUME DE THESE FRANÇAIS	4
ABSTRACT	5
RESUME GRAND PUBLIC.....	6
Remerciements	7
Introduction.....	13
Chapitre 1 Les facteurs de transcription de la famille OvoL : rôles et fonctions.....	13
1.1 La formation des gonades et établissement de la lignée germinale : de la souris à la drosophile en passant par le ver à soie	16
1.2 La formation de l'œil : l'exemple du ver plat et la dystrophie polycornéenne.....	17
1.3 Le système pileux : quand une dérégulation d'OvoL t'hérise le poil.....	18
1.4 Les facteurs OvoL/Shavenbaby dans la Transition Epithélio-Mésenchymateuse.....	20
Chapitre 2 Shavenbaby : Mode de régulation et Fonctions	25
2.1 Régulation de Shavenbaby.....	25
2.1.1 Contrôle de l'expression de Shavenbaby par une pléiade de régions cis-régulatrices	25
2.1.2 Régulation post-traductionnelle de l'activité de Svb	29
2.1.2.1 Identification des peptides smORF Polished-rice.....	29
2.1.2.2 Shavenbaby : deux protéines pour le prix d'une.....	30
2.2 Rôle de Svb dans l'épiderme.....	32
2.2.1 Régulation centrale des gènes effecteurs du remodelage épidermique	32
2.3 Shavenbaby dans le système digestif	34
2.3.1.1 Le système digestif.....	34
2.3.1.1.1 Caractéristiques structurelles et fonctionnelles du système digestif.....	34
2.3.1.1.1.1 L'intestin moyen	36
2.3.1.1.1.2 Le lignage cellulaire	36
2.3.1.2 Fonction antagoniste de Shavenbaby pour le maintien de l'homéostasie de l'intestin	37
Chapitre 3 Au croisement entre l'informatique et la biologie : la bio-informatique pour la génomique chez la drosophile	39
3.1 Le séquençage : De Sanger au Séquençage Nouvelle Génération.....	39
3.1.1 Le principe du séquençage de l'ADN par Sanger	42
3.1.2 Séquençage et obtention des séquences de référence.....	43
3.1.2.1 Le principe du séquençage global	44
3.1.2.1.1 Assemblage du génome	44
3.1.2.1.1.1 Nettoyage	44
3.1.2.1.1.2 Criblage.....	44
3.1.2.1.1.3 Chevauchement.....	45
3.1.2.1.1.4 Échafaudage	45
3.1.2.1.1.5 Résolution récurrente : roches, pierres, galets	45
3.1.2.1.1.6 Consensus.....	45
3.1.3 Séquenceurs Nouvelle Génération	46
3.1.4 Le génome de <i>Drosophila melanogaster</i>	49
3.2 Éléments de régulation de la transcription : les enhancers.....	51
3.2.1 Mise en évidence des régions cis-régulatrices.....	51
3.2.2 Mécanismes de fonctionnement des enhancers	51
3.2.2.1 Le pistage ou tracking	52

3.2.2.2	Le chaînage	52
3.2.2.3	La formation de boucle ou looping	52
3.2.3	Les techniques d'identification de régions cis-régulatrices	55
3.2.3.1	Principe de gènes rapporteurs - promoter bashing	55
3.2.3.2	Approche globale : Principe du STARR-seq « self-transcribing active regulatory region sequencing »	56
3.2.3.3	Approche globale : Identification des sites de fixation des facteurs de transcription Principe du ChIP-seq.....	58
3.3	Organisation de la chromatine et accessibilité des enhancers	60
3.3.1	Les marques des histones :	60
3.3.1.1	L'acétylation et la désacétylation des histones.....	61
3.3.1.2	La méthylation et la déméthylation des histones	61
3.3.2	L'organisation de la chromatine chez la Drosophile	63
3.3.2.1	Les couleurs de la chromatine	63
3.3.2.1.1	La chromatine dite « non permissive »	65
3.3.2.1.1.1	La chromatine noire	65
3.3.2.1.1.2	La chromatine verte	66
3.3.2.1.1.3	La chromatine bleue	66
3.3.2.1.2	La chromatine « permissive ».....	69
3.3.2.1.2.1	La chromatine jaune	70
3.3.2.1.2.2	La chromatine rouge.....	70
3.3.2.2	Les histones comme signatures des régions cis-régulatrices	71
	Objectifs	74
	Résultats.....	75
	Chapitre 4 Shavenbaby protège les cellules souches rénales en interagissant avec la voie Hippo	75
4.1	Contexte de la publication	75
4.2	Résumé de la publication.....	78
4.2.1	Contribution.....	78
	Chapitre 5 Les smORF peptides Pri sont de médiateurs de la signalisation de l'ecdysone, contribuant à façonner les réponses spatio-temporelles	91
5.1	Contexte de la publication	91
5.2	Résumé de la publication.....	92
5.2.1	Contribution.....	93
	Chapitre 6 OvoL/Shavenbaby régule des ensembles distincts de gènes cibles en fonction du type cellulaire dans lequel il opère	109
6.1	Résumé de la publication.....	109
	Discussion	156
	Chapitre 7 Les acteurs régulant la fixation de Svb.....	157
7.1	Le rôle des facteurs de transcription pionniers.	157
7.2	Polycomb et Shavenbaby ont une localisation similaire au sein du noyau : peuvent-ils être partenaires ?	161
7.3	Rôle des effecteurs du remodelage de la chromatine ?	163
	Chapitre 8 Implication de SvbRep dans la pause des enhancers.....	167
	Chapitre 9 Shavenbaby dans les cellules intestinales.....	171
9.1	Vers une fonction spécifique de SvbRep dans le lignage intestinal	173
	Conclusion	176
	Références	177
	Table des abréviations	195

Table des illustrations

Figure 1 : Le locus de shavenbaby/ovo code pour 3 formes différentes.....	13
Figure 2 : Le gène ovo est nécessaire à la formation et au maintien de l'oeil chez <i>Schmidtea mediterranea</i> (ver plat).....	17
Figure 3 : Localisation des formes OvoL1 et OvoL2 au sein du bulbe pileux.....	19
Figure 4 : La régulation de l'expression de svb passe par l'utilisation de plusieurs régions cis-régulatrices.....	27
Figure 5 : La régulation de l'expression de svb au sein des cellules centrales épidermiques se fait par l'intégration des voies de signalisations Notch, Wg/Hh et EGF.....	28
Figure 6 : La recherche sur les smORFs est en pleine expansion.....	29
Figure 7 : Classification des smORF.....	30
Figure 8 : Processus de maturation de Shavenbaby.....	31
Figure 9 : Lignes stables exprimant les deux formes de Svb issus de la modification de cellules S2.....	31
Figure 10 : Shavenbaby contrôle le remodelage des cellules épidermiques embryonnaires.....	32
Figure 11 : Détermination des sites de fixations de Shavenbaby dans les cellules épidermiques embryonnaires.....	33
Figure 12 : Composition et comparaison des systèmes digestifs humain (gauche) et de drosophile (à droite).....	35
Figure 13 : Lignage en cellule unique des cellules intestinales.....	37
Figure 14 : Prise de notes du carnet de Francis Crick sur le Dogme Central de la Biologie Moléculaire (Cobb, 2017).....	41
Figure 15 : Algorithme de Celera Adaptée de (Myers et al., 2000).....	46
Figure 16 : Nombre de publications contenant le terme "bioinformatics" source PUBMED.....	47
Figure 17 : Frise chronologique du développement de la bio-informatique.....	48
Tableau 18 : Caractéristiques de l'assemblage du génome de <i>Drosophila melanogaster</i> issu de (Hoskins et al., 2015).....	49
Figure 19 : Assemblage du génome de la <i>Drosophile</i>	50
Figure 20 : Modèles de communication enhancer-promoteur.....	54
Figure 21 : Principes de caractérisation de régions cis-régulatrices in vivo par approche de microscopie.....	56
Figure 22 : Principe de la technique du STARR-seq.....	57
Figure 23 : Localisation des enhancers au sein de la <i>Drosophile</i>	58
Figure 24 : Principe de la technique de CHIP-seq.....	59
Figure 25 : Modifications des histones.....	62
Figure 26 : Principe du DamID.....	64
Figure 27 : Composition des cinq types de chromatine en protéines associés à de la chromatine et en histones.....	65
Figure 28 : Les protéines du complexe protéique Polycomb.....	68
Figure 29 : Les différentes couleurs de la chromatine et sa composition.....	69
Figure 30 : Les chromatines jaune et rouge sont différentes.....	70
Figure 31 : Les différentes couleurs de la chromatine.....	71
Figure 32 : Les marques d'histones comme signatures des éléments géniques.....	72
Figure 33 : Signature en histones des enhancers développementaux et de ménage.....	73
Figure 34 : Tubules de Malpighi et lignage cellulaire des cellules souches rénales.....	77
Figure 35 : L'hormone ecdysone et ses pulsions au cours du cycle de vie de la <i>Drosophile</i>	92
Figure 36 : Modèles de fonctionnements des régions enhancers.....	158
Figure 37 : Comparaison des sites de fixation de Grh et Svb.....	160
Figure 38 : Logo de Grainyhead (source : base de données Jaspar).....	161
Figure 39 : Observation par immunomarquage de Svb et Polycomb dans les cellules exprimant les formes de Svb.....	162
Figure 40 : Svb et Polycomb anticorrèlent sur les régions de fixation de Svb.....	163
Figure 41 : Enrichissement de différentes protéines architecturales sur les régions de fixation de Svb.....	165
Figure 42 : Pipsqueak pourrait être partenaire de Svb.....	166
Figure 43 : Les pics Svb (totaux et associés aux gènes dérégulés) présentent outre les motifs ovo des motifs lov/BTB, Trl et srp.....	167
Figure 44 : Principe du Hi-C.....	168
Figure 45 : La présence de SvbRep induit la formation de boucle enhancer-promoteur.....	170
Figure 46 : Algorithme du clustering hiérarchique ascendant.....	171
Figure 47 : Algorithme du clustering de k-moyennes.....	172
Figure 48 : Détermination du groupe de gènes cibles spécifiques de SvbRep par clustering.....	173
Figure 49 : Gene Ontology (Processus Biologique) des gènes cibles de SvbRep (GORilla).....	175

Résumés

RESUME DE THESE FRANÇAIS

Les facteurs de transcription à doigt de zinc de la famille OvoL, au nombre de trois chez les mammifères (OvoL1, 2 et 3), sont des acteurs majeurs de la régulation épithéliale durant le développement des vertébrés. Chez l'adulte, ils participent à la fois à la différenciation terminale des cellules épithéliales et au maintien de ce caractère pour l'homéostasie tissulaire. Pour ce faire, les facteurs OvoL inhibent l'action des facteurs pro-EMT (Transition Epithélio-Mésenchymateuse) ZEB et TWIST qui eux poussent vers l'état mésenchymateux. Ainsi, lors de cancer humain, OvoL1 empêche la transformation des cellules tumorales épithéliales en cellules invasives mésenchymateuses métastatiques. De plus, des maladies héréditaires, dont des pathologies sévères des yeux et de la peau, ont été associées à la dérégulation des gènes OvoL.

À ce jour et malgré des intérêts biomédicaux potentiels, les mécanismes moléculaires qui soutiennent l'action des facteurs OvoL et notamment leurs gènes cibles spécifiques restent mal compris. Il est possible que la redondance fonctionnelle des facteurs OvoL chez les vertébrés complique l'accès à leur fonction.

Chez la Drosophile (*Drosophila melanogaster*) les facteurs OvoL, fortement conservés au sein du règne animal, sont seulement représentés par le gène *ovo/shavenbaby (ovo/svb)*, membre fondateur de cette famille. Les deux formes germinales, OvoA et OvoB soutiennent l'ovogenèse alors que la forme somatique, Svb, contrôle le remodelage tridimensionnel des cellules épithéliales. Svb est également nécessaire pour le maintien, la prolifération, mais également la différenciation des cellules souches digestives adultes. Svb est synthétisée sous une forme longue, appelée SvbREP qui, sous l'action conjuguée des petits peptides Polished-rice (Pri) et de l'Ubiquitine-E3-ligase Ubr3, est clivée par le protéasome pour obtenir une forme plus courte, SvbACT, transcriptionnellement active.

Pour comprendre le mode d'action individuel de chacune des deux formes de Svb, nous avons développé un modèle cellulaire (S2) permettant l'expression de l'une ou l'autre des deux formes. Nous avons effectué des expériences de génomique pour déterminer de façon globale le comportement de SvbAct et SvbRep. Pour cela, nous avons analysé leur profil de fixation à l'ADN (ChIP-seq) d'une part, et identifié les gènes cibles (RNA-seq) pour chaque forme d'autre part. L'intégration de ces données a permis de mettre en évidence que SvbREP et ACT se fixent sur les mêmes régions génomiques cis-régulatrices (enhancers) pour contrôler de manière antagoniste l'expression d'environ 250 gènes cibles directs. Nous avons également montré que Svb régule des groupes de gènes différents en fonction du type cellulaire au sein duquel il évolue et mis en évidence que l'environnement chromatinien contraint son activité.

Mes travaux ont non seulement permis de mieux appréhender le mode de fixation de facteurs OvoL sur le génome, mais apportent une nouvelle perspective sur la nature de leurs gènes cibles impliqués dans le contrôle du cycle cellulaire et/ou l'adhésion cellulaire.

ABSTRACT

The OvoL family of zinc finger transcription factors, three in number in mammals (OvoL1, 2 and 3) are major players in epithelial regulation during vertebrate development. In the adult, they participate in both the terminal differentiation of epithelial cells and the maintenance of this character for tissue homeostasis. To do this, OvoL factors inhibit the action of the pro-EMT (Epithelial-Mesenchymal Transition) factors ZEB and TWIST, which push towards the mesenchymal state. Thus, in human cancer, OvoL1 prevents the transformation of epithelial tumour cells into invasive metastatic mesenchymal cells. In addition, hereditary diseases, including severe eye and skin diseases, have been associated with the deregulation of OvoL genes. Currently, despite potential biomedical interests, the molecular mechanisms that underpin the action of OvoL factors, including their specific target genes, remain poorly understood. It is possible that the functional redundancy of OvoL factors in vertebrates complicates access to their function.

In *Drosophila (Drosophila melanogaster)*, OvoL factors are highly conserved in the animal kingdom and are represented only by the *ovo/shavenbaby (ovo/svb)* gene, a founding member of this family. The two germline forms, OvoA and OvoB, support oogenesis while the somatic form, Svb, controls three-dimensional epithelial cell remodelling. Svb is also required for the maintenance, proliferation and differentiation of adult digestive stem cells.

Svb is synthesised in a long form, called SvbREP, which, under the combined action of the small Polished-rice peptides (Pri) and the Ubiquitin-E3 ligase Ubr3, is cleaved by the proteasome to obtain a shorter, transcriptionally active form, SvbACT. To understand the individual mode of action of each of the two forms of Svb, we developed a cell model (S2) that allows the expression of either form. We performed genomic experiments to determine the overall behaviour of SvbAct and SvbRep. To do this, we analysed their DNA binding profile (ChIP-seq) on the one hand, and identified the target genes (RNA-seq) for each form on the other. Integration of these data revealed that SvbREP and ACT bind to the same cis-regulatory genomic regions (enhancers) to antagonistically control the expression of approximately 250 direct target genes. We have also shown that Svb regulates different groups of genes depending on the cell type in which it evolves and that the chromatin environment constrains its activity.

My work has not only provided a better understanding of how OvoL factors bind to the genome, but also provides a new perspective on the nature of their target genes involved in cell cycle control and/or cell adhesion.

RESUME GRAND PUBLIC

Les facteurs de la famille OvoL sont connus chez les vertébrés pour contrôler la destinée épithéliale en luttant contre le processus de Transition Epithélio-Mésenchymateuse (EMT). La dérégulation de ces facteurs est associée aussi bien à différentes pathologies humaines comme des cancers. S'il est probable que la redondance fonctionnelle des facteurs OvoL chez les vertébrés complique l'accès à leur fonction, chez la Drosophile, le gène OvoL/Shavenbaby (*svb*), est l'unique membre de cette famille. Il régule l'acquisition tridimensionnelle des cellules épidermiques lors du développement ainsi que le maintien des cellules souches digestives adultes.

Mon travail de thèse a permis d'appréhender le mode de fixation de Svb à l'échelle du génome et d'identifier les gènes dont il régule l'expression dans trois types cellulaires distincts. Grâce à des analyses bio-informatiques, j'ai mis en évidence que la fixation de Svb est contrainte par l'environnement chromatinien ce qui a pour conséquence de modifier le répertoire de ses gènes cibles en fonction du contexte cellulaire au sein duquel il évolue.

PUBLICLY ACCESSIBLE ABSTRACT

OvoL family factors are known in vertebrates to control epithelial fate by controlling the Epithelial-Mesenchymal Transition (EMT) process. Dysregulation of these factors is associated with various human pathologies such as cancers. While the functional redundancy of OvoL factors in vertebrates probably complicates access to their function, in *Drosophila*, the OvoL/Shavenbaby (*svb*) gene is the only member of this family. It regulates the three-dimensional acquisition of epidermal cells during development and the maintenance of adult digestive stem cells.

My thesis work allowed me to understand how Svb is bound on a genome-wide scale and to identify the genes whose expression it regulates in three distinct cell types. Using bioinformatics analyses, I demonstrated that Svb binding is constrained by the chromatin environment, which results in a modification of its target gene repertoire depending on the cellular context in which it evolves.

Remerciements

Cet exercice des remerciements n'est pas l'exercice dans lequel je me sens la plus à l'aise, mais il est important pour moi de remercier toutes les personnes qui m'ont tant apporté tout au long de cette aventure qu'est le doctorat¹. Tant de chemin parcouru en si peu de temps.

Ce travail a été financé durant quatre ans par la Ligue Nationale Contre le Cancer, merci à eux pour la confiance accordée. Je remercie également mes rapporteurs les Drs Mounia Lagha, Frédéric Bantignies et Krzysztof Jagla d'avoir accepté d'évaluer ces travaux ainsi mes examinateurs les Drs Julie Caramel et Benoît Ballester pour leurs retours.

Ce manuscrit représente le travail de 4 ans de thèse (6 mois de stage et une pandémie) au sein de l'équipe de François Payre. Il est le résultat des conseils, des discussions que j'ai pu avoir et recevoir de tous les membres actuels et passés de cette équipe si particulière, mais pas que !

Je suis arrivée au sein de l'équipe avec mon bagage de bio-info et ma valise de biologiste expérimentatrice, il a fallu apprendre à communiquer ensemble. Ça n'a pas toujours été facile. Je ne compte pas le nombre de réunions où nous fûmes frustrés les uns et les autres de ne pas avoir réussi à nous comprendre. Je suis heureuse aujourd'hui de pouvoir dire sans sourciller que je ne ressens plus ce fossé de compréhension qui existait entre la bio-informaticienne que j'étais en arrivant et le reste de l'équipe comme j'ai pu le ressentir au début de ce travail. J'espère que ce sentiment est réciproque.

Cédric et François vous avez été des directeurs de thèse assez fous pour m'encadrer mais aussi me permettre de me construire et me transmettre votre amour de la recherche ! MERCI !

En 2017, **François**, tu as fait le pari de prendre au sein de ton équipe une étudiante de M2 Bio-informatique et Biologie des Systèmes. À ce moment-là, la bio-informatique ne faisait pas partie des compétences internes de l'équipe et tu as pourtant parié sur ma venue dans ton groupe et je t'en remercie ENORMEMENT. Cela a marqué le début de cette aventure ! Tes conseils, les pauses cigarettes sur les marches du Centre de Biologie du Développement (CBD) et maintenant du bâtiment de TP ont contribué en partie à faire de ce projet ce qu'il est.

Cédric, il y a 4 ans que tu as amorcé ton passage du côté obscur de la bio en mettant un pied dans le monde des histones en acceptant de m'encadrer en doctorat. Je ne peux que te remercier (et le mot ici est faible) pour tout ce que tu m'as apporté. J'ai reçu au début de la rédaction un message d'un futur étudiant de M2 cherchant un stage me demandant comment c'était de travailler avec toi et quel type d'encadrant tu es. J'ai répondu que travailler avec toi c'est extrêmement enrichissant, que tu es juste, tu es le genre d'encadrant qui en te disant les choses te pousse à aller plus loin ! Alors Cédric Merci pour tout ! La scientifique^{en perpétuelle construction} que je suis ne serait pas la même sans toi ! L'autre partie du projet c'est à toi que je la dois !

Passons aux autres membres de l'équipe Payre, il faut dire qu'on en a vécu des péripéties ensemble, toutes ne pouvant évidemment pas figurer ici.

Hélène, tu as été ma première encadrante avec François lors du stage de M2. Les ordi et toi vous n'êtes pas très copains (on se souvient tous de ce fameux Mac faisant des aller-retours) et pourtant tu t'es accrochée et m'a soutenue pour que tout se passe au mieux et que mon adaptation au sein de l'équipe se fasse. Les discussions à la pièce à mouche lors de ma visite mensuelle lors de

¹ Julie et Guigui pas taper si des fautes il y a, je refuse que vous lisiez cette partie en avance

passage des stocks étaient bien drôles. Je retiens de toi ce conseil que tu m'as donné un jour : « Alex construit ta bulle de sciences et de « tranquillité au labo » tu ne te rends pas compte, je pense, de l'impact qu'a eu cette simple phrase sur moi. Merci !

Jenny ton franc-parler, tes remarques toujours bien placées, ta rigueur scientifique et ta gentillesse font que travailler avec toi c'est se sentir soutenue et au sein d'une vraie équipe. Les pauses café, les voyages (enfin surtout le meeting en Suisse), les retraites CBD, tout cela n'aurait sûrement pas été les mêmes sans toi et ta joie de vivre. Je compte sur toi pour la soirée de thèse. Merci vraiment ! Merci pour la relecture du manuscrit également.

Phiphi ! Merci de m'avoir pris sous ton aile pour mon retour à la paillasse ! Quand je pense à toi je pense aussi à ta musique dans le labo ! Sache qu'avec la nouvelle disposition il y a des jours où ça me manque vraiment² ! Avoir partagé le bureau avec toi jusqu'au déménagement m'a fait avoir de nombreux souvenirs et discussions et cela n'est pas négligeable, bien au contraire, dans cette aventure. Tes déguisements lors des retraites CBD font partie de ces souvenirs que je vais chérir. Merci Phiphi et j'espère pouvoir assister à cette soutenance si attendue !

Simon, ta gentillesse et ta douceur compensent et apaisent les caractères volcaniques présents dans l'équipe (dont je n'exclus en rien le mien), je te remercie pour tes retours sur mon travail, les discussions à propos de musique ou d'informatique, le prêt incessant de câble pour recharger mon téléphone³. Je te souhaite d'être heureux avec tes filles. Merci pour tout.

Anne, les discussions avec toi ont toujours été un plaisir, on peut commencer à parler science, puis partir sur du personnel, pour revenir à de la science et discuter à nouveau de livre ou de films. Ta curiosité, ta gentillesse m'ont tant apporté durant ce doctorat. J'aurais vraiment aimé t'avoir comme prof à la fac, j'espère que ce manuscrit t'apportera les schémas sur la bioinfo que tu attends. Ne doute jamais de ta capacité à transmettre la science, tu as réussi à faire que certains processus de bio du dev ne me semblent plus si mystérieux et incompréhensibles. Merci Anne.

Clément travailler avec toi Cédric et Alexia sur le projet RNA-seq dans les cellules de l'intestin a été l'une des périodes où j'ai le plus ressenti que je faisais vraiment partie de l'équipe à part entière. Et pourtant on était en plein confinement lors de la réception des premières données ! Mais outre cet aspect professionnel, tu es une personne foncièrement gentille avec qui il est agréable de parler de tout, de jeux de société (va falloir que j'apprenne à jouer aux Magics) et un exceptionnel mage de l'air ;)

Hadi, on s'est connu bien avant que tu n'intègres l'équipe lors de mon tout premier meeting DrosoFrance pour ensuite se revoir à nouveau en meeting à l'EDRC à Lausanne, quels souvenirs ! Discuter avec toi est toujours motivant, ta joie de vivre et ta gentillesse sont un moteur lors des coups de stress. Tu vas être un très bon PI j'en suis sûre !!! J'ai gardé en mémoire un tweet ne l'oublie pas xD

Damien (DamDam), ta gentillesse, ton soutien, ta bonne humeur et ton humour m'ont toujours rendu le sourire dès que je te voyais au labo. Ne perd jamais cela, c'est ta force ! Prend garde avec le groupe de filles du CBD, elles vont réussir à te faire aimer les enfants, elles commencent avec Salva. Caresse à Iron et courage à toi pour la dernière ligne droite.

Aurore tu l'auras eu ta thèse et tu la mérites amplement ! Tu es une fille en or, gentille et compréhensive. Je me souviens de la fois où Jenny m'a demandé de te parler de l'équipe, je savais déjà que cela n'avait pas tant d'importance car tu étais déjà bien plus capable que je ne pensais l'être moi-même. Tu vas le trouver ce phénotype j'en suis sûre !!! Belle vie avec Guilhem !!!

Alexia, ma petite stagiaire. Je souhaite que le jour de la soutenance on puisse dire que tu prends le relai de la bio-info chez les Payre ! Tu auras été ma première stagiaire de M2, j'espère avoir pu t'encadrer au mieux. Ne sous-estime jamais tes capacités et garde toujours cette niaque. Quel autre stagiaire peut dire qu'il a dormi dans une tente avec sa co-encadrante hein ? Bon par contre pour Loulou je ne suis pas sûre qu'il n'y ai pas de cacahuète lors du pot, mais je ferai tout pour.

Je n'oublie pas les anciens membres de l'équipe !

² A l'époque du 4R3 je n'aurai pas cru dire cela un jour pour ceux qui savent de quoi je parle

³ Foutu téléphone qui ne tient pas la charge

Jérôme, j'espère un jour partager à nouveau une bière et une corde d'escalade avec toi. Qui sait peut-être un jour ?

Maylis, on a partagé pas mal de choses pendant tes années de thèse. Des fous rires, des concerts (non, non je ne suis pas du tout devenue aphone lors l'un d'eux), des sorties avec AlphaT. Je te souhaite plein de bonheur avec Pierre et une belle continuation.

Sandy (Didy) ! Il y a des rencontres professionnelles qui font que l'amitié ressentie va bien plus loin. Je n'ai pas honte de dire que tu es ma sœur libanaise. Tu mérites tellement d'être heureuse avec ton Khalil⁴. Tes adaptations des expressions françaises me manquent ! Les discussions scientifiques ou non dans le bureau jusqu'à pas d'heure aussi ! Ne laissent jamais personne te dire que ton accent libanais tu dois l'effacer, sinon même en étant loin je vais lui expliquer moi ! Un jour, je te le promets, tu me feras visiter ta maison du Liban et je rencontrerai ta famille ! Et surtout, n'oublie jamais que tu es Princesse Didy de la Night !

Sachez que passer les stocks de mouches va me manquer à Tromsø !

Mais le labo ce n'est pas uniquement les membres de l'équipe Payre. C'est aussi les membres de l'ancien CBD et maintenant MCD. Et un petit grand intrus ancien du CRCA⁵.

Adeline, je ne pouvais pas commencer cette partie sans parler de toi ! On en a partagé des choses ensemble, des meetings, une chambre d'auberge de jeunesse suisse, des anecdotes, des retraites CBD sans dormir ! On ne peut pas raconter cela ici, mais sache que cette thèse sans ton amitié aurait été bien plus compliquée et décidément plus fade. Je te souhaite pleins de bonheurs et je suis sûre que tu seras une excellente prof !

Ben, il y a des amitiés qui commencent en Licence et qui se retrouvent en Doctorat et cette amitié c'est la nôtre ! J'ai hâte d'à nouveau partager un parquet de danse avec toi, que tu m'apprennes à patiner sur la glace et de partager l'aventure scandinave avec toi. On se retrouve en Février à Tromsø !

Les filles **Dehlia** et **Marie** vous vaincrez la thèse ! Delhia croit en toi vraiment, tu en es plus que capable ! Et longue vie au 10^{ème} étage ! Marie, toujours en train de courir le sourire aux lèvres. Ne laisse personne abuser de ta gentillesse, elle est ta force ! Je dis bien personne !

Isma ! Juste en écrivant ton prénom le smile me vient ! Tu es une personne d'une gentillesse et d'une empathie comme rarement on en croise dans une vie. Je te souhaite tant de bonheurs avec Marie, Charlie, Lou et Salva.

Clair, mon beau **Clair**. J'espère que tu es prêt à twerker comme jamais ! Je tiens à dire que ton canapé est de plus confortable, sache-le ! Bonne continuation et courage à toi aussi !

Vicente, notre rencontre a été cocasse⁶. Tu m'as vu débouler dans ton labo les bras chargés de cadeaux. J'espère avoir le temps avant mon départ de continuer à te connaître. Et si un jour je vais à Lisbonne, je ferai en sorte que ça soit en même temps que toi.

Mumu, Jean-Louis, Laurence merci pour votre gentillesse à mon écart, les rires, le partage de souvenirs avec nous. Moi quand je serai grande je veux être un peu comme vous ! MOLKY !

Julie, tu es la pile du CBD. Merci pour tes mots de soutien, ta gentillesse et tout ce que tu fais pour nous les doctorants. C'est toi la star !

Je pense aussi aux anciens : Angie, Cyril, Mélanie, Momo, Alexandre, Arnaud, Thomas, Aurore, Mélanie, Sarah, Mégane. Mais aussi à tous ceux dont je n'ai pas cité les noms merci pour cette aventure à vos côtés.

Ces années au CBI (CBD) resteront marquées dans ma mémoire. Bonne continuation à tous !

⁴ Dit Didy comment on dit Tatie Alex en libanais ?

⁵ En fait je ne sais pas si te mettre ici ou après car on s'est connu avant la thèse mon Ben

⁶ Je t'écris en français car comme tu l'as dit tous les deux nous communiquons ainsi ;)

L'aventure de la thèse c'est aussi des rencontres⁷ lors de meeting, de formations avec qui on reste en contact et on se lie d'amitié. **Julian** fuerzas tú puedes, te deseo lo mejor y ya sabes si necesitas tomar un cafe o un maté desde dónde sea. Espero un día cumplir con ir a Argentina (con un guía) !

Cette thèse a été rédigée durant l'été entre le CBI, mon appart' sur Borderouge et la terrasse du DiCapo de la Halle des Machines. Merci à toute l'équipe du DiCapo pour leur accueil et de m'avoir permis de squatter la terrasse. Cela aura contribué au stock de vitamine D avant le départ dans le Grand Nord !

Le soutien de ma famille a été essentiel durant cette période. Je vais me faire un peu violence pour écrire certaines choses ici.

Abuelo, Abuela esperó que leáis esto un día. ¡Pediré a Thibault que os lo traiga ! Sois los mejores abuelos del mundo y eso nadie me lo negara. Estoy tan orgullosa de ser vuestra nieta. ¡Este trabajo va por vosotros ! ¡ Lo hice lo mejor que pude pero lo hice !

También pienso a **Mamie Dolores** y **Pepe David**. Os llevo siempre conmigo. ¡ También va para vosotros !

Thibault, mon frangin, tu es le petit frère mais tu es toujours là pour me soutenir. Alors on se le dit rarement, je vais te l'écrire : Je t'aime ! No dejes nadie decirte de no vales porque tú vales más que nadie ! Tu liras ceci au meilleur neveu du monde qui d'un sourire fait s'estomper les soucis : « **Maël**, sois curieux tous les jours, pose des questions, n'ai pas honte de demander pourquoi les choses sont telles qu'elles sont ! Les plus belles expériences viennent avec la curiosité, la gentillesse et une dose de culot ».

Papa, tu n'as jamais compté tes heures du travail, tu m'as appris à me relever toujours malgré les échecs et de toujours avancer malgré les soucis. Ton soutien et les petites étoiles dans tes yeux quand les gens te demandent ce que je fais dans la vie sont un moteur. Ta princesse.

Maman, tu es un modèle de maman battante, travailleuse et courageuse. Ta générosité et ton altruisme sont un moteur pour moi. Tu as vu je l'ai fait, bien qu'en réalité tu y as cru bien plus que moi. Je te confie Yelda et Bagheera, prend bien soin de tes petits-enfants poilus comme tu l'as si bien fait pour Lilly en son temps.

Je pense aussi à vous tous **Los Ferris por el mundo** et la **famille Mancheño** je ne serai pas là où j'en suis aujourd'hui sans vous tous.

Vous allez me manquer en Norvège.

⁷ « Moi, si je devais résumer ma vie aujourd'hui avec vous, je dirais que c'est d'abord des rencontres. Des gens qui m'ont tendu la main, peut-être à un moment où je ne pouvais pas, [...] Et c'est assez curieux de se dire que les hasards, les rencontres forment une destinée »

Según una leyenda japonesa, las personas estamos unidas por un hilo rojo. Un hilo que nos enlaza con la persona a la que estamos destinados. No creo en el destino. Creer en el destino te relaja, te predispone a la espera, te convierte en un pasivo inerte aguardando el futuro. Me niego a arrastrar una cadena buscando a quien engancharme. Sin embargo creo en los vínculos. Esas hebras o tentáculos que te unen a las personas que ya conoces. Los vínculos se estiran, se tensan... Y a veces se rompen. El tiempo, el desgaste y el olvido corroen el lazo.

Pero algunos, inexplicablemente, perviven. ¿Qué nos une a alguien con la firmeza necesaria ? – Valeria Ferris - Valeria

Les amis vous croyez en moi bien plus que je ne le fais au quotidien et cette confiance que vous m'insufflez a contribué à ce que vous tenez entre vos mains.

Julie, ma **Juju**, ma chewie, ma meilleure amie, ma femme. Qui l'eut cru il y a quelques années⁸ dans ce self d'un collège carcassonnais en ZEP que je serai en train de finaliser la rédaction d'un manuscrit de thèse ? Certainement pas la fille de 11ans que j'étais qui demanda timidement à une maman et sa fille, dont le sac pesait 5 tonnes, si elle pouvait s'asseoir avec elles. Ça a marqué le début de notre amitié ! On en a partagé des aventures, des rires et des pleurs aussi. Et malgré la distance qui nous sépare aujourd'hui, un tout petit océan de rien du tout, ton soutien et ton amitié sont et resteront les piliers de ce que je suis. Merci d'être toujours là ! Belle vie avec Franky et Sparrow !

Guillaume, mon **Guigui** ! Alors là, ça y est, on y est ! Enfin ! Je pourrais écrire des pavés sur notre amitié et ce qu'elle m'apporte au quotidien. Tu te souviens qu'on a révisé le sujet de l'épreuve d'SVT du bac grâce à toi 1h avant ? Moi oui ! Cette anecdote est le reflet de ce que tu es ! Toujours là au bon moment ! Merci ! Tu es le meilleur des meilleurs amis qu'une fille comme moi puisse avoir. Je crois bien que ces remerciements commencent à devenir un peu comme tes idées de nouvelles. Je vais te dire ce que tu me dis souvent : tu es bien plus capable que ce que tu ne crois. Alors va Guigui ! Lance-toi ! Et termine ce Sceptre Ancestral !!!!! Ah et aussi on va le marquer ici TU AS TOUJOURS RAISON⁹ !

S'il n'y a pas trop de fautes dans ce manuscrit c'est grâce à ces deux-là¹⁰ !

François, **Mushu** ! Sans toi, je n'aurai pas appris à coder et sans code pas de thèse en bio-info ! Les sessions Skype en plein été après ma journée de boulot sur le banc d'un village à plus de 1200m d'altitude avec une connexion plus qu'aléatoire alors que tu étais déjà en thèse, je ne te remercierai jamais assez ! Tu es la force tranquille, toujours présent pour écouter, analyser et conseiller ! Merci ! Bon par contre va falloir revoir cette règle complètement injuste qui dit que le Fluxx est un jeu banni auquel on ne jouera plus¹¹.

J'ai hâte de pouvoir repartir en concert avec vous !

Romain, **Rominou**, tes boutades, nos sorties ciné, tout cela a permis de rendre ces dernières années de thèse bien plus douces ! On se ressemble bien plus qu'on ne pourrait le croire, alors sache que pour moi tu es un véritable ami et prend soin de toi de temps en temps ! Merci pour ton soutien. Je vais essayer de reprendre la guitare (quand j'aurai trouvé le moyen de la ramener).

⁸ On taira le chiffre exact mais certaines sources disent que si c'était un humain il pourrait passer son permis B

⁹ Ça restera plus qu'un message sur une appli de messagerie

¹⁰ S'il en reste c'est aussi de leur faute et pas du tout à cause de ma dyslexie, mauvaise foi quand tu nous tiens

¹¹ L'histoire veut qu'aucune partie n'ai été gagnée par François dans l'aéroport de Milan

Chacha, Chachaaaaaaa ! Les tatouages sur mon corps c'est elle ! Mais avant d'être ma tatoueuse pleine de talent tu es mon amie. Celle qui te rentre dedans mais pour te faire sortir de ta zone de confort ! Tu es pleine de talentS et applique les conseils que tu m'as souvent donnée le long de la thèse à toi et on dit Scregneugneu à Jean-Mi Le Syndrôme de l'Imposteur¹² ! A toi et ton envol !

Bastien, toujours présent quand il s'agit de me filer un coup de main, pour discuter et rigoler. Tes questions et interrogations ont toujours permis de me faire avancer alors merci Bastou ! Je t'attends avec plaisir pour observer les aurores boréales en refaisant le monde !

Flore, que dire ? Quelle belle rencontre ! J'aime quand tu me pousses à vaincre ma timidité, quand tu finis par dire tout haut les bêtises auxquelles j'ai pensée. Et surtout je te remercie pour ta capacité à me soutenir et me canaliser. Merci pour ces fous rires, les escapes games qu'on a partagées, les soirées jeux de société. Merci d'accepter la garde partagée de la N'Alex avant son départ en Norvège.

Clément ! Notre capacité de résonance m'étonnera toujours¹³ ! François qu'as-tu créé en faisant qu'on se rencontre ?! Merci pour ta gentillesse à tout point de vue, tes Pisco Sour, tes « petits » plats asiatiques, ton ordi, les sessions FFXIV et Genshin, la découverte des chroniques sur France Inter et toutes ces autres choses ! Je te souhaite une merveilleuse continuation !

Il y a la famille de sang et la famille que la vie nous fait connaître. Vous êtes de ma famille les copains ! Je vous attends avec impatience sur Tromsø.
Rdv pour le Nouvel An 2022-23 !!!

Je remercie également tout ceux dont je n'ai pas cité le nom mais aussi tous ceux qui se sont déplacés pour la soutenance et fêter la fin de cette aventure, mais aussi ceux qui y ont assisté en visio.

¹² Cf. Coliandre

¹³ La patator du parc de nos débuts reste quand même une de nos plus belles idées, avec le cadeau d'anniversaire de François cette année !

Introduction

Chapitre 1 Les facteurs de transcription de la famille OvoL : rôles et fonctions

Les protéines Ovo-like (OvoL) sont des facteurs de transcription à doigts de zinc (tétrade de C2H2). Ce domaine de fixation à l'ADN (Acide Désoxyribonucléique), situé dans la région C-terminale de la protéine est connu pour être fortement conservé au cours de l'évolution. Il est ainsi identifiable du ver plat aux mammifères en passant par la drosophile (Kumar et al., 2012; Saxena et al., 2020).

Le gène *Ovo*, identifié initialement chez la Drosophile, est le membre fondateur de cette famille. *Ovo* est spécifiquement exprimé dans la lignée germinale femelle et son activité est déterminante pour l'ovogenèse et la différenciation sexuelle (Oliver et al., 1987; Mével-Ninio et al., 1991; Garfinkel et al., 1992). La présence de deux promoteurs alternatifs permet la production de deux protéines nommées OvoA et OvoB qui agissent de manière antagoniste sur le développement de la lignée germinale. OvoA agit en qualité de répresseur transcriptionnel et OvoB en tant qu'activateur (Andrews et al., 2000). Les grands cribles génétiques réalisés par Wieschaus et Nüsslein-Volhard en 1984 (Wieschaus et al., 1984) ont permis l'identification de deux allèles nommés *Shavenbaby*¹ et *Shavenbaby*² (*svb*¹, *svb*²). Ces mutations affectent spécifiquement une forme somatique du gène *ovo*, nommée *svb* (Mével-Ninio et al., 1995). Ainsi le locus *ovo/shavenbaby* permet la production de trois protéines, OvoA, OvoB et Svb (cf. Figure 1). Les formes OvoA et OvoB, respectivement répressive et activatrice, sont spécifiques de la lignée germinale alors que la forme Svb présente uniquement dans le soma, contrôle la morphogenèse épidermique (Payre et al., 1999) et le comportement des cellules souches digestives (Bohère et al., 2018; al Hayek et al., 2021). Je vous présenterai plus spécifiquement les fonctions de la protéine Svb par la suite.

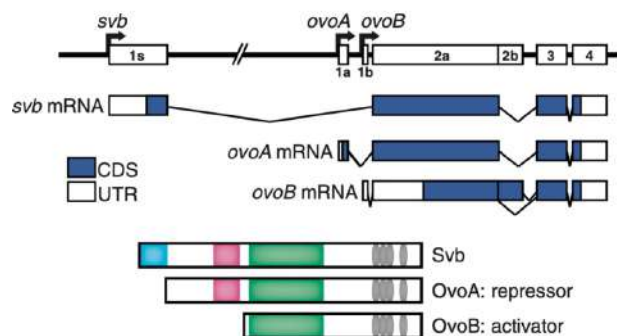


Figure 1 : Le locus de *shavenbaby/ovo* code pour 3 formes différentes
Schéma des 3 ARNs transcrits à partir du gène *ovo/svb* ainsi que leurs protéines associées. Les régions codantes (CDS) sont représentées en bleu. Les 3 protéines présentent 4 doigts de zinc en C-terminal identiques ainsi qu'un domaine activateur (vert). Ovo-A et Svb possèdent un domaine répresseur (magenta). La forme Svb possède spécifiquement l'exon 1S (cyan).
D'après (Kondo et al., 2007 a)

Les vertébrés comptent trois gènes codant pour des protéines OvoL (*ovo1-3*) permettant chacun la production de plusieurs isoformes protéiques. Les gènes *ovoL1* et *ovoL2*, qui sont les plus étudiés, sont exprimés dans différents tissus à divers temps développementaux (Mével-Ninio et al., 1995; Payre et al., 1999; Roca et al., 2013; Li et al., 2014 a; Cross et al., 2015 a; Hayashi et al., 2017; Saxena et al., 2020) (cf. Tableau 1). L'expression de ces facteurs est contrôlée par les voies de signalisation Wnt et BMP (Tsuji et al., 2018). Par diverses approches de mutations dirigées et de phylogénie, il a été montré que la régulation par la voie Wnt des gènes *ovoL* est conservée entre les organismes (Bi et al., 2019 ; Cross et al., 2015b ; Payre et al., 1999 ; Unezaki et al., 2004a). Il a également été observé qu'OvoL1 est capable d'inhiber son expression ainsi que celle d'*ovoL2* (Ito et al., 2017; Tsuji et al., 2018)

Les facteurs OvoL sont impliqués dans le développement de différents lignages épithéliaux tels que la peau, le rein ou l'épithélium mammaire (Mével-Ninio et al., 1991; Oliver et al., 1994; Dai et al., 1998; Payre et al., 1999; Johnson et al., 2001; Mackay et al., 2006; Nair et al., 2007). Ainsi, les souris déficientes pour *ovoL1* présentent des défauts d'implantation des follicules pileux et l'apparition de kystes rénaux (Dai et al., 1998; Li et al., 2005). Les souris dépourvues d'*ovoL2* sont embryonnaires létales (Unzaki et al., 2004 b; Mackay et al., 2006). Par leur fonction dans la différenciation des cellules épithéliales, les facteurs OvoL jouent aussi un rôle dans le processus de plasticité épithéliale. Cette dernière se caractérise par le fait qu'une cellule épithéliale peut devenir mésenchymateuse et inversement. Le premier processus est appelé Transition Epithélio-Mésenchymateuse (EMT) et son inverse, Transition Mésenchymato-Epithéliale (MET). En ce qui concerne la plasticité cellulaire, la fonction des facteurs OvoL est de promouvoir et de maintenir l'état épithélial en contrecarrant l'action des facteurs pro-EMT tels que les facteurs ZEB1&2, Snail et Twist (Li et al., 2014 a; Hong et al., 2015; Chung et al., 2019 a).

Dans ce paragraphe, je vais m'attacher à décrire la fonction des facteurs OvoL dans la formation des gonades, le développement de l'œil et le système pileux.

Tableau 1 : Fonctions et implications des facteurs Ovo à travers les organismes

Organisme	Nom	Fonctions
Homo sapiens	OvoL1-3	EMT ; spermatogenèse ; différenciation épidermique ; développement du rein...
Mus musculus	mOvo	Développement épidermique ; mise en place du système pileux ; mise en place de la lignée germinale ; formation de l'intestin, du cœur...
Bombyx mori	BmOvo	Formation des gonades ; développement des cellules germinales primaires ; segmentation ; développement des ailes...
Schmidtea mediterranea	SmedOvo	Formation de l'œil ; régénération de l'œil
Drosophila melanogaster	Svb OvoA ; OvoB ;	Formation des gonades ; mise en place de la lignée germinale ; différenciation épidermique, remodelage cellulaire ; prolifération, différenciation et protection contre l'apoptose des cellules souches digestives

1.1 La formation des gonades et établissement de la lignée germinale : de la souris à la drosophile en passant par le ver à soie

L'absence totale du gène *ovo* chez les drosophiles femelles altère le développement des cellules germinales conduisant à une atrophie des ovaires et à la stérilité (Busson et al., 1983; Oliver et al., 1987). Trois mutations antimorphes et dominantes, *ovoD1*, *ovoD2* et *ovoD3* ont été identifiées. Si ces mutations permettent le développement des cellules germinales, elles bloquent de façon dominante l'ovogenèse. Lorsqu'elles sont combinées avec des mutations perte de fonction, il en résulte le développement de cystes anormaux contenant un grand nombre de cellules germinales ne se différenciant pas et formant des tumeurs ovariennes (Oliver et al., 1990). Le gène *ovo*, tout comme les gènes *bag of marbles (bam)* et *sex lethal (sxl)*, appartient donc à la classe de gènes dont la dérégulation génère des tumeurs ovariennes (Pauli et al., 1990). Une étude détaillée du locus *ovo* a permis d'identifier un promoteur spécifique pour chaque forme protéique OvoA et OvoB (Andrews et al., 2000). Cela a permis de mettre en évidence qu'OvoB est exprimé pendant toute l'ovogenèse et qu'OvoA est exprimée seulement dans les stades tardifs. Cette même étude par une approche « simple hybride » a mis en évidence l'antagonisme fonctionnel entre les deux formes protéiques. En effet, si OvoA agit en qualité de répresseur transcriptionnel, OvoB se comporte comme un activateur (Andrews et al., 2000). L'expression précoce de la forme OvoA est notamment responsable de la stérilité observée avec les mutations dominantes *ovoD1-3*.

Il a été montré que les cellules primaires germinales (CPG) expriment spécifiquement OvoB au cours de l'embryogénèse (Mével-Ninio et al., 1995; Hayashi et al., 2017) et l'altération de sa fonction conduit à une perte des CPG chez la femelle comme chez le mâle. Au niveau moléculaire, l'altération d'OvoB se traduit par la perte d'expression de gènes spécifiques de la lignée germinale dont *piwi*, *vasa* et *nanos*. La présence d'OvoB dans les CPG semble donc favoriser leur destin germinal et réprimer des gènes impliqués dans le développement somatique (Mével-Ninio et al., 1995; Hayashi et al., 2017).

Chez la souris, les facteurs OvoL1, OvoL2 et OvoL3 sont les orthologues de la protéine Ovo de Drosophile (Li et al., 2002; Unezaki et al., 2004 a) et le gène *OvoL2* est le gène présentant la plus forte et consistante (précoce) expression dans les CPG. Chez les souris mutantes pour ce gène, une diminution significative du nombre de cellules germinales primaires est observée. Du fait de la létalité précoce des embryons mutants (E10,5), une analyse plus fine de la fonction d'OvoL2 fut impossible (Hayashi et al., 2017). Cependant, du fait de la similitude des phénotypes observés chez la drosophile et la souris, les auteurs proposent que les facteurs Ovo jouent un rôle conservé au cours de l'évolution pour la formation des CPG (Hayashi et al., 2017). Une étude récente réalisée chez *Bombyx mori* (ver à soie), montrant une désorganisation des ovarioles et une baisse de la fertilité, vient supporter cette hypothèse (Bi et al., 2019).

L'ensemble de ces études mettent donc en lumière un rôle essentiel des facteurs OvoL dans l'établissement des lignées germinales et des gonades à travers le règne animal.

1.2 La formation de l'œil : l'exemple du ver plat et la dystrophie polycornéenne

Le ver plat ou planaire, *Schmidtea mediterranea*, est un bon modèle pour étudier la formation de l'œil car des nombreux mécanismes et gènes sont conservés comme la présence de cellules souches pluripotentes de l'œil mais aussi de gènes homologues aux gènes exprimés chez les vertébrés dans les cellules photoréceptrices et dans l'épithélium pigmentaire rétinien, comme l'opsine et la bestrophine 1. Lapan et Reddien ont montré, grâce à des expériences de séquençage de l'ARN (RNA-seq) et d'hybridation *in situ*, que le gène *smad-ovo* est enrichi dans les yeux et dans les cellules progénitrices de l'œil chez les planaires. Le ver plat est également connu pour être un bon modèle pour étudier la régénération. Dans la même étude (Lapan et al., 2012), il a été observé que 2 jours après l'amputation de la tête de l'animal les cellules au niveau de la région dorsale antérieure présentent une expression de *smad-ovo* et que cette expression est maintenue jusqu'à 7 jours après amputation dans les cellules progénitrices de l'œil sous le contrôle des gènes *six1-2*, *eya*, *otxA* et *sp6-9*. Des expériences d'interférence à l'ARN (RNAi) de *smad-ovo* sur ces animaux décapités montrent une perte des cellules progénitrices, et donc un arrêt de la régénération de l'œil (Figure 2-A). De plus, ces mêmes expériences RNAi-*smad-ovo* ont été réalisées chez les animaux sains et une perte progressive de l'œil a été observée (Figure 2-B) (Cross et al., 2015 b). Toutes ces observations indiquent un rôle des facteurs OvoL dans la formation et la maintenance de l'œil via la maintenance des cellules progénitrices de l'œil.

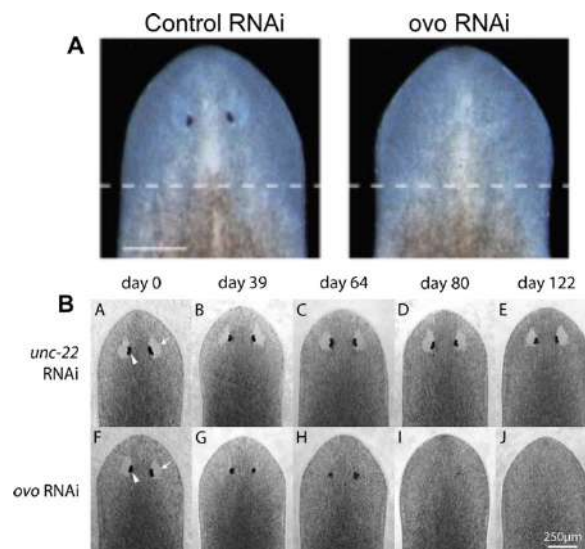


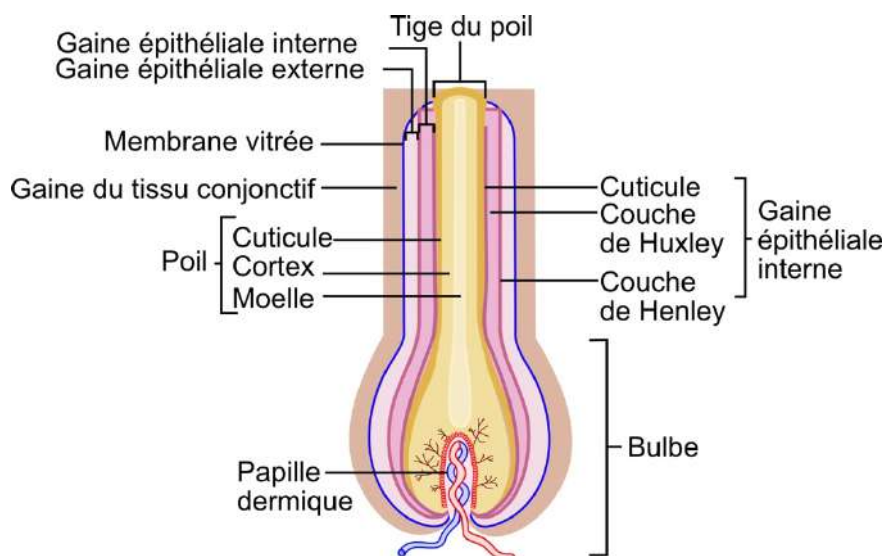
Figure 2 : Le gène *ovo* est nécessaire à la formation et au maintien de l'œil chez *Schmidtea mediterranea* (ver plat)
A) La déplétion du gène *ovo* chez l'animal décapité empêche la régénération de l'œil ; B) La déplétion du gène *ovo* chez l'animal sain sur le long terme empêche le maintien de l'œil. (Adapté de (Lapan et al., 2012; Cross et al., 2015 b))

Chez l'Humain, il a été observé que certaines maladies génétiques impliquent des facteurs OvoL. La dystrophie de Schlichting aussi appelée Dystrophie Cornéenne Postérieure Polymorphe (PPCD) en est un exemple. Cette maladie autosomale dominante, dont l'apparition des symptômes se fait dès l'âge de 1 an, présente quatre formes distinctes impliquant différentes mutations et facteurs. La forme PPCD1 se caractérise par une mutation sur le chromosome 20, corrélant avec une augmentation de l'activité du facteur OvoL2, qui entraîne une augmentation de l'inhibition du facteur ZEB1. Cette augmentation provoque une régulation aberrante de la transition épithélio-mésenchymateuse au sein de l'épithélium

cornéen. Les patients atteints de cette dystrophie présentent ainsi des dégradations de l'endothélium cornéen (stroma) et de la membrane Descemet pouvant aller jusqu'à l'apparition d'œdème et une dégradation de la vision pour les cas les plus graves nécessitant alors des greffes et des traitements lourds (Davidson et al., 2016; Kitazawa et al., 2016; Chung et al., 2017, 2019 a). Les formes PPCD2, PPCD3 et PPCD4 sont associées respectivement à une dérégulation du collagène type VIII, une diminution de ZEB1 et une augmentation de l'activité de Grainyhead-like 2 (GRHL2). Cependant une expression ectopique d'*OvoL2* ou *GRHL2* est suffisante mais pas nécessaire pour provoquer une PPCD, mais une insuffisance en ZEB1 est toujours impliquée dans une PPCD.

1.3 Le système pileux : quand une dérégulation d'OvoL t'hérise le poil

Comme nous le verrons en détail par la suite, une des fonctions de la forme somatique, *Svb*, codée par le locus *ovo/svb* chez la drosophile, est la formation d'extension cytoplasmique apicale (trichome) présent à la surface de certaines cellules épithéliales (Payre et al., 1999; Delon et al., 2004). Chez les vertébrés, l'importance des facteurs OvoL dans la formation et la régulation du système pileux s'est faite à travers l'étude du gène *OvoL1*. En effet, la délétion de ce gène entraîne la formation de poil hirsute et désordonné (Dai et al., 1998). Les études qui ont suivi, ont montré que chez l'Humain, les facteurs OvoL1 et OvoL2 sont exprimés au sein du bulbe pileux à divers endroits du follicule avec des localisations différentes pour chaque protéine. OvoL1 est, par exemple, présent dans le noyau des cellules basales du cortex alors qu'OvoL2 est présent dans le cytoplasme de ces cellules (Figure 2) (Ito et al., 2017). De plus, chez l'Humain, des mesures de la capacité de la formation de poils des cellules basales du follicule pileux montrent que lorsque la concentration en OvoL1 est modifiée, diminuée avec des RNAi-*ovoL1* ou augmentée avec des lentivirus-*ovoL1*, le nombre de follicules formés est, respectivement, augmenté ou diminué (Shin et al., 2014). Il a également été mis en exergue qu'OvoL1 et OvoL2 sont impliqués dans la différenciation et la prolifération des kératinocytes. Ces fonctions passent par une inhibition de la voie Notch par OvoL2 qui permet la prolifération des kératinocytes, une inhibition d'OvoL2 et de c-Myc par OvoL1 pour leur différenciation (Tsuji et al., 2018). Les facteurs OvoL sont donc nécessaires à la formation et la structuration du système pileux et semblent être de bonnes pistes thérapeutiques dans le cadre de traitement de l'alopecie par exemple.



Forme du facteur OvoL	Localisation dans le bulbe pileux	Localisation cellulaire	Concentration
OvoL1	Papille dermique	Cytoplasme / Nucléaire	+ / +
	Bulbe	Nucléaire	+++
	Gaine épithéliale interne	Cytoplasme	+++
	Gaine épithéliale externe	Nucléaire	+
	Moelle	-	-
	Cortex	Nucléaire	+
OvoL2	Papille dermique	Cytoplasme	+++
	Bulbe	Nucléaire	+++
	Gaine épithéliale interne	Cytoplasme/ Nucléaire	+ / +
	Gaine épithéliale externe	Nucléaire	+
	Moelle	Cytoplasme	+++
	Cortex	Cytoplasme	+++

Figure 3 : Localisation des formes OvoL1 et OvoL2 au sein du bulbe pileux.

Les deux formes des facteurs OvoL au sein du système pileux se situent dans des localisations cellulaires complémentaires.

Nous venons de voir différentes implications et caractérisations des fonctions des facteurs de la famille OvoL dans différents tissus et organismes. Bien qu'ayant des rôles différents, tous sont reliés au seul et même processus de la transition Épithélio-Mésenchymateuse. Ce processus est essentiel lors du développement, de la cicatrisation et a également été montré comme étant impliqué dans des pathologies comme les cancers. Je vais m'attacher dans la partie suivante à vous présenter ce mécanisme si particulier et en quoi l'étude de SvB est nécessaire pour une meilleure compréhension de sa régulation.

1.4 Les facteurs OvoL/Shavenbaby dans la Transition Epithélio-Mésenchymateuse.

La transition Epithélio-Mésenchymateuse (EMT) a été définie pour la première fois à la fin des années 1970 par Elizabeth Hay (Yang et al., 2020). Ce processus est essentiel lors du développement des organismes – pour la formation de la crête mais aussi celle de la valve cardiaque par exemple – et également impliqué dans la cicatrisation. Cependant il a été mis en exergue que ce processus était relié au diagnostic de pathologies comme le cancer et en particulier lors à la formation de métastase (Puisieux et al., 2014; Nieto et al., 2016; Jolly et al., 2017). Ce mécanisme se caractérise par la transformation par étapes successives des cellules épithéliales, stationnaires et bien organisées, en cellules mésenchymateuses, circulantes. Ce passage de cellules épithéliales à cellules mésenchymateuses passent par la diminution d'expression d'E-cadhérine aboutissant à une réorganisation du cytosquelette des cellules épithéliales et au détachement de celles-ci de la membrane basale. La lame basale s'en trouve également dégradée. La E-cadhérine est remplacée dans les cellules mésenchymateuses par la N-cadhérine (Francou et al., 2020). Ces transformations réduisent ainsi l'adhérence cellule/cellule et favorisent la migration et l'invasion cellulaire. Cela a été lié au mécanisme de formation de métastases (Nieto et al., 2016; Smith et al., 2016; Jolly et al., 2017). Les cellules épithéliales ne passent pas à un état mésenchymateux de façon directe. Des états hybrides ont été observés pendant lesquels les cellules possèdent des caractéristiques tant épithéliales que mésenchymateuses (Nieto et al., 2016; Yang et al., 2020)

Dans un souci d'homogénéisation, il a été défini que pour être liés à l'EMT les mécanismes étudiés devaient résulter en un phénotype de plasticité cellulaire ainsi que des cellules présentant de multiples étapes hybrides (Yang et al., 2020).

De nombreux facteurs de transcription promeuvent l'EMT. Ces facteurs sont appelés des facteurs pro-EMT et contiennent les facteurs ZEB, Snail mais aussi l'ARN non codant miR200. La présence de facteurs ZEB1,2, de Snail et Twist en de forte concentration promeut le passage EMT.

A contrario, une grande concentration en facteur OvoL mais aussi Grainyhead permet le maintien des cellules épithéliales, en inhibant les facteurs ZEB1&2, et promulgant ainsi le passage de cellules mésenchymateuses à cellules épithéliales. Ce processus est appelé la transition Mésenchyme-Epithéliale (Li et al., 2014 a; Nieto et al., 2016). Les facteurs OvoL sont donc considérés comme des gardiens de l'état épithélial. Dans la revue suivante, publiée dans *Médecine/Sciences*, je vous présente ces mécanismes et leurs impacts dans le cadre pathologique des cancers épithéliaux via le prisme de l'étude de Shavenbaby chez la mouche du vinaigre *Drosophila melanogaster*.

► Des avancées majeures révèlent l'hétérogénéité intra-tumorale des cancers d'origine épithéliale, incluant des cellules initiatrices de tumeurs qui ressemblent aux cellules souches adultes. Les cellules souches normales et tumorales partagent en effet leur plasticité entre phénotypes épithéliaux et mésenchymateux, progressant par une série d'états intermédiaires, réversibles. Si un cœur de régulateurs (Snail, Zeb, ...) est bien connu pour déclencher la transition épithélio-mésenchymateuse (TEM), les facteurs OvoL/Shavenbaby sont récemment apparus comme des stabilisateurs épithéliaux. La balance entre facteurs pro-TEM et OvoL pourrait ainsi réguler la plasticité phénotypique et le potentiel métastatique des tumeurs. Nous abordons cette question chez la drosophile, un modèle pour disséquer *in vivo* la fonction de Shavenbaby. Nos travaux montrent que Shavenbaby est un régulateur clé de l'homéostasie des cellules souches adultes. Shavenbaby est indispensable à leur survie, agissant en interaction directe avec la voie Hippo pour protéger les cellules souches de la mort cellulaire programmée. ◀

La transition épithélio-mésenchymateuse (TEM) intervient à différentes étapes du développement embryonnaire (gastrulation, crêtes neurales, somitogénèse, ...) et aussi chez l'adulte, par exemple pour la cicatrisation [1]. Cette transition phénotypique réduit l'adhérence entre cellules épithéliales et favorise la migration invasive (Figure 1A). La TEM est bien connue pour contribuer à l'agressivité des cancers d'origine épithéliale, ou carcinomes, notamment les cancers du sein, du poumon, de la prostate ou colorectaux, etc. [1, 2]. La TEM déclenche la dissociation des cellules de la tumeur primaire et facilite leur migration, l'intravasation, et leur dissémination vers des sites distants pour former des métastases. Une transition réciproque de l'état

Les facteurs OvoL Des régulateurs clés de la plasticité épithélium-mésenchyme et des cellules souches

Alexandra Mancheno-Ferris^{1,2}, Cédric Polesello^{1,2},
François Payre^{1,2}



¹Centre de Biologie du Développement, Université Paul Sabatier Toulouse III, Bâtiment 4R3, 118 route de Narbonne, 31062 Toulouse, France.
²CNRS, UMR5547, Centre de Biologie du Développement, Toulouse, France.
francois.payre@univ-tlse3.fr

mésenchymateux vers des propriétés épithéliales, ou transition mésenchymo-épithéliale (TME), favorise ensuite la formation de macro-métastases (Figure 1B), largement responsables de la morbidité des cancers [1, 2].

Plasticité phénotypique entre épithélium et mésenchyme

Des travaux récents montrent que les transitions entre phénotypes épithéliaux et mésenchyme (TEM et TME) n'agissent pas comme de simples « interrupteurs » entre deux types cellulaires alternatifs [2, 3]. Au contraire, ces transitions progressent par une série d'états intermédiaires, où les cellules présentent des propriétés mixtes entre propriétés épithéliale (E) et mésenchymateuse (M), appelés états hybrides E/M (Figure 1A). Remarquablement, l'existence de ces états hybrides permet une réversibilité des transitions entre caractères épithéliaux et mésenchymateux. On parle ainsi de plasticité épithélio-mésenchymateuse, un paramètre sans doute fondamental pour la progression tumorale et l'échappement aux thérapies [2, 3]. Élucider comment ces états intermédiaires se forment et se remodelent en fonction du microenvironnement représente aujourd'hui un enjeu majeur en cancérologie.

Les facteurs OvoL dans la plasticité épithélio-mésenchymateuse et les cancers

Les nombreuses études réalisées depuis la découverte de la TEM ont permis l'identification des facteurs déclenchant le passage de l'état épithélial à l'état mésenchymateux [1]. Ces facteurs pro-TEM regroupent des facteurs de transcription (Snail, Slug, Zeb1, Zeb2, Twist, ...) et leurs micro-ARN régulateurs (*miR200*, ...). Ensemble, ces facteurs vont promouvoir l'état mésenchymateux et inhiber les

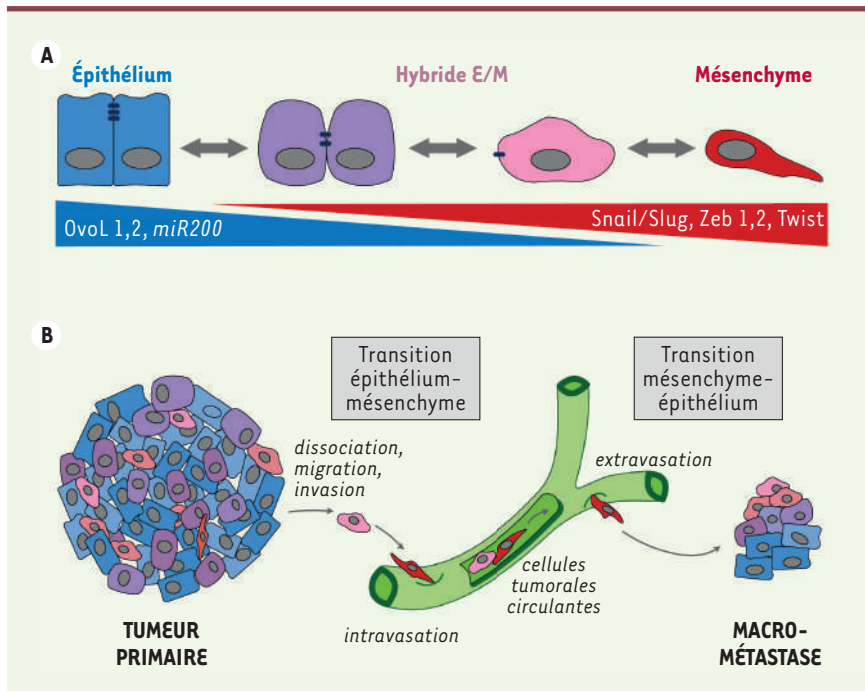


Figure 1. Plasticité épithélium-mésenchyme et progression tumorale. **A.** Les transitions de l'état épithélial à l'état mésenchymateux (de gauche à droite), ou l'inverse (de droite à gauche), passent par une série d'états intermédiaires réversibles caractérisés par des phénotypes hybrides épithélio-mésenchymateux (E/M). Ces transitions résultent des différentes activités de facteurs de transcription : Snail/Slug, Zeb1,2 et Twist qui déclenchent la TEM, et les facteurs OvoLs qui stabilisent l'état épithélial. **B.** La transition épithélio-mésenchymateuse favorise la dissociation, migration, l'invasion, l'intravasation et l'extravasation des cellules tumorales. Quand elles atteignent des organes distants, ces cellules cancéreuses engagent une transition mésenchymo-épithéliale, qui accroît leur potentiel prolifératif, pour former des tumeurs secondaires.

propriétés épithéliales, par exemple en réprimant l'expression de la E-cadhérine, un composant essentiel des jonctions cellulaires [1, 2]. Plus récemment, des facteurs qui, au contraire, stabilisent l'état épithélial, viennent d'être identifiés ; en particulier, la famille des facteurs de transcription Ovo-Like (OvoL1/3) [1, 2] est conservée chez tous les animaux et ses membres vont contrebalancer l'activité des facteurs pro-TEM [4]. Les niveaux relatifs entre facteurs pro-TEM d'une part, et facteurs OvoLs d'autre part, pourraient ainsi réguler et/ou stabiliser les états hybrides E/M (Figure 1A). Le profilage systématique des tumeurs relie en effet l'expression des facteurs OvoL au potentiel métastatique des carcinomes [4]. D'une manière générale, la réduction de l'expression des facteurs OvoL dans les tumeurs primaires est de très mauvais pronostic. La réexpression artificielle d'une isoforme spécifique du facteur OvoL2 permet d'inhiber la formation de métastases dans des souris greffées avec des cellules de tumeurs mammaires dérivées de patients [5]. Cependant, des travaux de profilage de cellules uniques montrent que les tumeurs secondaires ré-expriment les gènes épithéliaux, ainsi que l'influence à la fois du microenvironnement et de l'infiltration des macrophages sur les transitions entre épithélium et mésenchyme [6]. Il est donc important de comprendre la fonction et le mode d'action des facteurs OvoL *in vivo*. Chez les vertébrés, l'existence de trois gènes paralogues (OvoL1-3) aux fonctions souvent redondantes, codant chacun différentes isoformes, complexifie malheureusement leur étude fonctionnelle.

La drosophile pour étudier la fonction des facteurs OvoL

Bien connue pour son apport à la compréhension des mécanismes de l'hérédité et du développement, la drosophile (*Drosophila melanogaster*)

devient désormais un modèle fécond pour l'étude des altérations génétiques conduisant aux tumeurs [7]. Le génome de la drosophile code un seul gène OvoL, *ovo/shavenbaby* (*svb*), le membre fondateur de la famille. Ce gène code trois isoformes protéiques : OvoA et B, qui sont essentielles au développement de la lignée germinale, et la forme somatique Shavenbaby (Svb), qui gouverne la différenciation épithéliale des cellules épidermiques [8, 9].

L'expression de *svb* est régulée par une large région cis-régulatrice qui intègre de nombreuses voies de signalisation et facteurs de transcription [8]. Svb est traduit sous la forme d'un long répresseur de transcription (SvbREP). La protéine Svb est ensuite partiellement dégradée par le protéasome (Figure 2), conduisant à la libération d'une forme courte qui agit comme activateur de transcription (SvbACT) [10, 11]. Cette maturation post-traductionnelle est induite par les peptides Polished-rice (Pri), fondateurs d'une nouvelle famille de micro-peptides régulateurs traduits à partir d'ARN apparemment non-codants [12]. Dans l'épiderme embryonnaire, SvbACT va induire l'expression d'une batterie de gènes cibles, codant des régulateurs du cytosquelette, des jonctions cellulaires, de la matrice extracellulaire, etc. [13], expliquant son rôle dans le remodelage épithélial.

Shavenbaby contrôle l'homéostasie des cellules souches adultes

Chez l'adulte, nous venons de découvrir l'expression spécifique de Svb dans les cellules souches digestives,

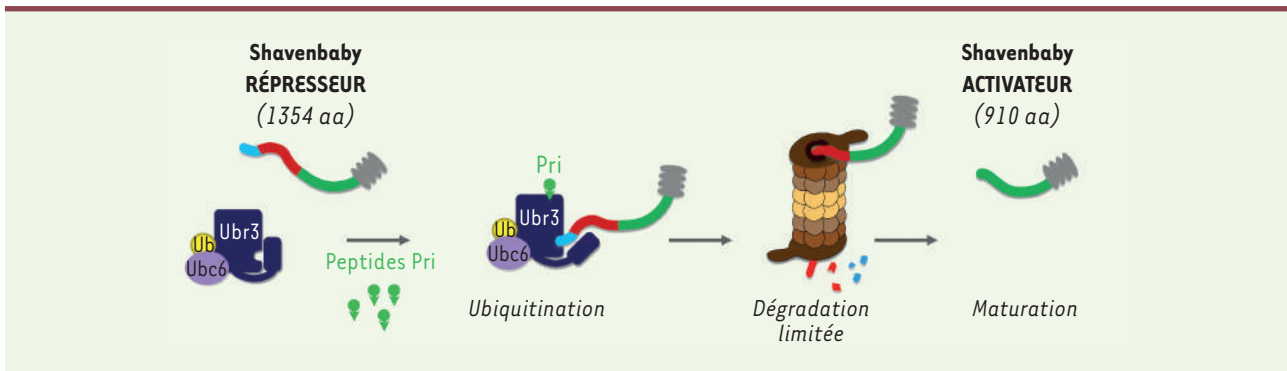


Figure 2. Processus de maturation post-traductionnelle de Shavenbaby. La liaison des micro-peptides Pri à l'ubiquitine ligase Ubr3 permet la fixation du complexe Ubc6/Ubr3 (enzyme de conjugaison E2 et ubiquitine ligase E3) sur la région N-terminale de SvbREP (cyan), conduisant à son ubiquitination, ainsi que la dégradation de la région répresseur (rouge) par le protéasome. Cette maturation produit une protéine tronquée, SvbACT, qui est un activateur de transcription. La région activatrice de la protéine Svb est représentée en vert, le domaine de liaison à l'ADN, en gris.

regroupant cellules souches intestinales et rénales. Les cellules souches digestives de la drosophile ont permis d'identifier de nouveaux facteurs et mécanismes (influence de la nutrition, des infections bactériennes,...), dont l'importance a été par la suite démontrée chez l'homme [7].

Nos travaux montrent que Svb est essentiel à la maintenance des cellules souches adultes [14]. Nous nous sommes d'abord focalisés sur les cellules souches rénales qui sont quiescentes, ce qui permet une analyse quantitative de l'ensemble du compartiment souche (Figure 3A). Les cellules souches digestives sont caractérisées par un phénotype hybride E/M et expriment le cœur des facteurs pro-TEM, *i.e.*, Escargot, Zfh1 (les équivalents respectifs de Snail, Zeb chez la drosophile) dont l'expression est régulée négativement par le micro-ARN *miR8* (l'équivalent de *miR200*) [15]. La surexpression de *miR8* suffit à provoquer la disparition des cellules souches qui se différencient massivement en cellules épithéliales [15], montrant l'importance des régulateurs pro-TEM pour le maintien des cellules souches. Si l'inactivation de Svb aboutit aussi à la disparition des cellules souches (Figure 3B), elle n'induit cependant pas leur différenciation précoce mais déclenche leur élimination par apoptose. On voit ainsi que les facteurs pro-TEM et le facteur épithélial Svb n'agissent pas par simple antagonisme, mais au contraire collaborent pour assurer l'homéostasie des cellules souches. Comme au cours du développement, le contrôle post-traductionnel de Svb, effectué par les peptides Pri, est déterminant pour les cellules souches adultes. Ainsi, le blocage de la maturation de Svb conduit à la disparition des cellules souches rénales, qui meurent par apoptose.

Shavenbaby interagit avec la voie Hippo

Nos travaux démontrent que, dans les cellules couches adultes, Svb interagit fonctionnellement avec la voie de signalisation Hippo, un régulateur clé de la survie cellulaire. Le médiateur nucléaire de la voie Hippo est la protéine Yorkie (YAP/TAZ, chez les mammifères), qui doit se lier à des facteurs de transcription pour réguler l'expression génique [16]. L'analyse bio-informatique des profils de fixation à la chroma-

tine (ChIP-seq) a permis de montrer que Svb et Yorkie partagent environ 30% de leurs sites de liaison génomique [14]. Remarquablement, l'ADN des gènes cibles de Svb responsables de la différenciation de l'épiderme [9, 13] n'est pas liés par Yorkie. Réciproquement, les séquences des gènes régulés par Yorkie pour contrôler la croissance cellulaire [16] ne sont pas fixées par Svb, suggérant que Svb/Yorkie interagissent spécifiquement pour la maintenance des cellules souches. En effet, nous montrons que les protéines Svb et Yorkie s'associent physiquement pour former un complexe nucléaire qui va directement notamment activer dans les cellules souches l'expression du gène *DIAP-1* (Figure 3C), codant l'inhibiteur majeur de l'apoptose.

Ces résultats établissent l'importance *in vivo* des facteurs OvoL/Svb pour le maintien du caractère « souche », et montrent comment chaque isoforme influe positivement ou négativement sur les cellules souches tumorales. La voie Hippo est impliquée dans différents cancers, et YAP/TAZ interagit aussi avec les facteurs pro-EMT [16] soulignant les relations intimes entre les différents régulateurs de la plasticité épithélio-mésenchymateuse. Nos travaux chez la drosophile aident ainsi à mieux comprendre les mécanismes de survie des cellules souches cancéreuses et de leur résistance aux traitements. Ils ouvrent aussi de nouvelles pistes pour le diagnostic et les thérapies anticancéreuses ciblées. Si, aux premiers stades des cancers épithéliaux, l'inhibition de la TEM peut prévenir l'invasion (par exemple en stimulant l'expression des facteurs OvoL), il faut au contraire inhiber la ré-acquisition des propriétés épithéliales des tumeurs secondaires. Nos résultats montrant le rôle clé des facteurs OvoL dans l'organisation épithéliale et la survie cellulaire renforcent l'intérêt de leur ciblage thérapeutique.

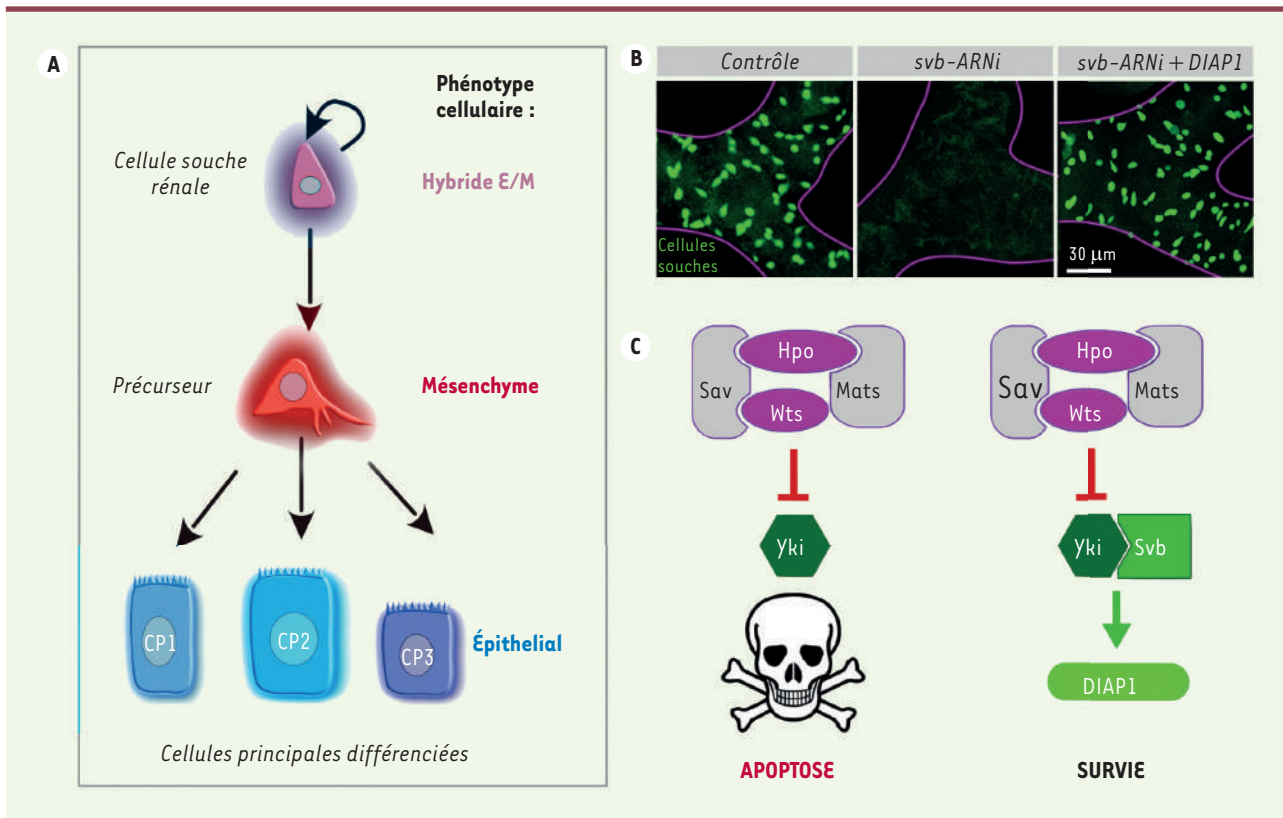


Figure 3. *Svb* protège les cellules souches rénales de l'apoptose. **A.** Les cellules souches adultes rénales présentent un phénotype hybride E/M. Elles assurent leur autorenouveau et la production d'un précurseur transitoire mésenchymateux, qui va migrer et se différencier en différentes cellules épithéliales principales (CP1,2,3). **B.** Images de microscopie montrant la disparition des cellules souches rénales (marquées par la GFP [*green fluorescent protein*], en vert) suite à l'inactivation de Shavenbaby par ARN interférent (ARNi). La restauration concomitante de l'expression de l'inhibiteur d'apoptose DIAP1 suffit à la survie des cellules souches. **C.** Mode d'action du complexe *Svb*/Yorkie (*Yki*) pour activer l'expression de DIAP1 et, ainsi, la maintenance des cellules souches. Hpo, Hippo ; Sav, Salvador ; Wts, Warts ; Mats, Mob as tumor suppressor.

Futures directions

Une caractéristique des facteurs *OvoL* est la production de multiples isoformes, présentant des activités différentes, voire antagonistes, y compris dans la progression tumorale [5]. Éclaircir la fonction des facteurs *OvoL* nécessitera donc de comprendre l'impact respectif de chaque isoforme sur la plasticité épithéliale, le contrôle des cellules souches, saines et cancéreuses. Grâce à l'accumulation des connaissances et l'utilisation d'outils génétiques bien maîtrisés chez cette espèce, la drosophile pourrait aider à percer les mécanismes respectifs des isoformes *OvoL* répresseur et activateur. Nous avons entrepris une combinaison d'approches de génomique fonctionnelle et de bio-informatique pour identifier le mode d'action moléculaire de chaque isoforme sur la régulation de l'expression du génome des cellules souches. Par ailleurs, nous analysons à l'échelle du génome entier les relations réciproques entre *OvoL*/*Shavenbaby* et les facteurs de remodelage de la chromatine, dont on découvre le rôle dans la plasticité épithélio-mésenchymateuse et les tumeurs [2]. Ces approches intégratives devraient fournir un nouvel éclairage sur les mécanismes

reliant plasticité épithélio-mésenchymateuse et la dynamique des cellules souches cancéreuses, et permettre d'identifier les meilleurs régulateurs à cibler, en fonction des différents stades de cancers, pour une médecine personnalisée. ♦

SUMMARY

***OvoL* factors: a family of key regulators of epithelium mesenchyme plasticity and stem cells**

Most prevalent cancers are of epithelial origin and their morbidity often results from secondary tumors. Cancer aggressiveness relates to intratumoral heterogeneity, including rare tumor initiating cells that share many features with adult stem cells. Both normal and cancer stem cells are characterized by their plasticity between epithelial and mesenchymal phenotypes, progressing through a series of reversible intermediates. While a core of regulators (*Snail*, *Zeb1-2*,...) is renowned to promote epithelial to mesenchyme transition (EMT),

OvoL/Shavenbaby factors now emerge as a family of key epithelial stabilizers. Therefore, pro-EMT and OvoL/Shavenbaby transcription factors could provide a molecular rheostat to control stemness and epithelial-mesenchyme plasticity. We address this question in flies, in which the unique OvoL/Shavenbaby factor offers a powerful *in vivo* paradigm for functional analyses. Our results show that Shavenbaby is critical for adult stem cell homeostasis, and directly interacts with the Hippo pathway to protect stem cells from death. \diamond

REMERCIEMENTS

Ces travaux sont soutenus par la Ligue contre le cancer (Allocation Doctorale), la Fondation pour la Recherche Médicale (DEQ20170336739) et l'Agence Nationale de la Recherche (ChronoNet).

LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

1. Nieto MA, Huang RY, Jackson RA, Thiery JP. Emt: 2016. *Cell* 2016 ; 166 : 21-45.
2. Gupta PB, Pastushenko I, Skibinski A, et al. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. *Cell Stem Cell* 2019 ; 24 : 65-78.
3. Lu W, Kang Y. Epithelial-mesenchymal plasticity in cancer progression and metastasis. *Dev Cell* 2019 ; 49 : 361-74.
4. Jia D, Jolly MK, Boareto M, et al. OVOL guides the epithelial-hybrid-mesenchymal transition. *Oncotarget* 2015 ; 6 : 15436-48.
5. Watanabe K, Villarreal-Ponce A, Sun P, et al. Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovo2 transcriptional repressor. *Dev Cell* 2014 ; 29 : 59-74.
6. Pastushenko I, Brisebarre A, Sifrim A, et al. Identification of the tumour transition states occurring during EMT. *Nature* 2018 ; 556 : 463-8.
7. Singh SR, Aggarwal P, Hou SX. Cancer stem cells and stem cell tumors in *Drosophila*. *Adv Exp Med Biol* 2019 ; 1167 : 175-90.
8. Payre F, Vincent A, Carreno S. ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. *Nature* 1999 ; 400 : 271-5.
9. Chanut-Delalande H, Hashimoto Y, Pelissier-Monier A, et al. Pri peptides are mediators of ecdysone for the temporal control of development. *Nat Cell Biol* 2014 ; 16 : 1035-44.
10. Kondo T, Plaza S, Zanet J, et al. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* 2010 ; 329 : 336-9.
11. Zanet J, Benrabah E, Li T, et al. Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* 2015 ; 349 : 1356-8.
12. Plaza S, Menschaert G, Payre F. In search of lost small peptides. *Annu Rev Cell Dev Biol* 2017 ; 33 : 391-416.
13. Menoret D, Santolini M, Fernandes I, et al. Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome Biol* 2013 ; 14 : R86.
14. Bohere J, Mancheno-Ferris A, Al Hayek S, et al. Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis. *Nat Commun* 2018 ; 9 : 5123.
15. Antonello ZA, Reiff T, Ballesta-Illan E, Dominguez M. Robust intestinal homeostasis relies on cellular plasticity in enteroblasts mediated by miR-8-Escargot switch. *EMBO J* 2015 ; 34 : 2025-41.
16. Manning SA, Kroeger B, Harvey KF. The regulation of Yorkie, YAP and TAZ: new insights into the Hippo pathway. *Development* 2020 ; 147: dev179069.

TIRÉS À PART

A. Mancheno-Ferris



Avec m/s, vivez en direct
les progrès et débats
de la biologie et de la médecine

CHAQUE MOIS / AVEC LES ARTICLES DE RÉFÉRENCE DE M/S
CHAQUE JOUR / SUR WWW.MEDECINESCIENCES.ORG

Abonnez-vous sur
www.medecinesciences.org

Chapitre 2 Shavenbaby : Mode de régulation et Fonctions

Comme présenté précédemment, le locus *ovo/shavenbaby* (*ovo/svb*) code pour deux formes protéiques germinales (OvoA et OvoB) et une forme somatique, Svb. L'équipe de François Payre, dans laquelle cette thèse a été réalisée, s'intéresse depuis de nombreuses années aux mécanismes impliqués dans la morphogénèse et le remodelage cellulaire chez la Drosophile. Les travaux réalisés dans l'équipe ont permis d'identifier le facteur de transcription Svb comme acteur essentiel pour le remodelage des cellules épidermiques et la formation d'extensions cytoplasmiques apicales, appelés trichomes. Des études récentes de l'équipe montrent un rôle important de Shavenbaby dans la maintenance, la prolifération mais aussi la différenciation des cellules souches du système digestif adulte.

De plus, les éléments de régulation de *svb* dispersés sur ≈ 100 kb, phénomène atypique chez la Drosophile, ont servi de support pour étudier des processus d'évolution des éléments cis-régulateurs (enhancers) ainsi que pour le développement du concept de « shadow enhancers ».

Enfin, je vous présenterai des travaux novateurs dans l'équipe mettant en évidence que l'activité transcriptionnelle de Svb est contrôlée par des peptides smORF (small Open Reading Frame) nommés peptides Pri (Polished-rice).

2.1 Régulation de Shavenbaby

2.1.1 Contrôle de l'expression de Shavenbaby par une pléiade de régions cis-régulatrices

La compréhension des mécanismes soutenant le développement d'un organisme est essentielle mais complexe à appréhender bien qu'elle soit une des questions fondamentales de la biologie du développement. Les gènes clés du développement doivent être finement régulés du point de vue spatial mais aussi temporel pour permettre un développement correct de l'organisme. Toutes les informations nécessaires à cette régulation fine sont contenues dans le génome de toutes les cellules de l'organisme au sein de régions courtes non codantes appelées « enhancer ». Ces régions localisées en amont et/ou dans les séquences introniques des gènes doivent être activées de façon fine et précise. Il faut noter qu'un même gène peut être régulé par plusieurs enhancers. L'expression de *svb* dans les cellules épidermiques est stéréotypée dans chaque segment de l'embryon et elle préfigure l'apparition des trichomes. Cela en fait un bon modèle pour étudier la régulation spatiale des gènes.

Le patron des denticules est mis en place par la coordination des voies de signalisation Wingless (Wg), EGF (Epidermal Growth Factor), Notch et Hedgehog (Hh) (Payre, 2004). L'interaction génétique séquentielle entre ces voies de signalisation va permettre de segmenter l'embryon et de définir les cellules qui expriment *svb*. Dans le cas de la région ventrale de l'épiderme qui constitue le modèle le plus simple d'intégration des voies, l'équipe a montré que *svb* intègre les voies de signalisation Wg et EGF pour définir les cellules qui vont former un trichome. Ainsi, la voie de signalisation Wg réprime l'expression de *svb* tandis que la signalisation EGF l'active (Payre et al., 1999).

Les efforts convergents de notre équipe et de celle de David Stern ont permis d'identifier les régions cis-régulatrices requises pour l'expression de *svb*. Mais aussi de montrer que l'évolution du patron des trichomes entre les différentes espèces de *Drosophile* est la conséquence de mutations au sein de ces régions (Sucena et al., 2003; McGregor et al., 2007). L'expression épidermique de *svb* est définie par sept régions/enhancers qui ont la particularité de s'étendre sur 90kb en amont du premier exon (Frankel et al., 2012; Kittelmann et al., 2021). En effet, chez la *Drosophile*, les enhancers, à quelques exceptions près dont *svb* fait partie, sont localisés dans le premier intron et/ou juste en amont (moins de 10kb) du gène dont ils contrôlent l'expression (Furlong et al., 2018).

À l'aide de lignées rapportrices couvrant les 50kb en amont du gène *svb*, les trois premières grandes régions ont été identifiées. Ces trois enhancers (7, E et A) sont actifs dans des régions distinctes et récapitulent à eux trois le patron d'expression complet de *svb* dans l'épiderme embryonnaire (McGregor et al., 2007). Un découpage plus précis de la boîte E a permis de mettre en évidence que cette région contient deux modules, E3 et E6 ayant une activité distincte (Nicolas Frankel et al., 2010). Enfin, Alexandre Ahmad Al Sawadi, doctorant au laboratoire, a identifié la séquence minimale de 292pb, appelée E3-14 (ou E3N), récapitulant le profil spatial de la boîte E3 (1kb) (Crocker et al., 2015). Puis, l'identification d'autres régions génomiques fortement conservées au cours de l'évolution ainsi que l'utilisation d'autres lignées rapportrices couvrant 41kb en amont de l'enhancer A, a permis de mettre en évidence deux régions cis-régulatrices additionnelles. De façon étonnante, ces deux régions DG et Z ont des registres d'activité similaire aux enhancers précédemment identifiés (Figure 4-B). La découverte de ces enhancers a participé au développement du concept de « shadow enhancers » ou enhancer redondant. En effet, la vision simpliste qui a longtemps perduré était qu'un enhancer soutient l'expression d'un gène dans un territoire donné, un deuxième enhancer dans un autre territoire et ainsi de suite. Chez la souris, un des exemple type de ce modèle, est l'enhancer nommé ZRS qui soutient l'expression du morphogène Sonic hedgehog (*Shh*) dans une zone particulière dite zone de polarisation active (ZAP) dans le bourgeon de membre. En utilisant la méthode de manipulation des génomes (CRISPR/CAS9), une étude a mis en évidence qu'une séquence de 17 paires de bases contenue dans cet enhancer était suffisante pour contrôler l'expression de *sonic hedgehog (shh)* dans le bourgeon de patte et donc la formation de cette dernière (Kvon et al., 2016). Cependant, le milieu des années 2000 a vu l'émergence d'un nouveau concept venant bousculer le modèle établi. Il a été mis en évidence notamment par des expériences de ChIP (Immunoprécipitation de Chromatine), qu'un facteur de transcription peut se fixer sur deux régions « enhancers » distantes de plusieurs kilobases pour réguler l'expression d'un même gène cible (Zeitlinger et al., 2007; Lagha et al., 2012) Cette découverte a permis l'apparition d'un nouveau concept, celui de « shadow » enhancer (Lagha et al., 2012). Ainsi, deux enhancers, un enhancer « pionnier » et un enhancer « shadow », dirigent l'expression du gène dans un même territoire assurant ainsi, par leur redondance, la forte robustesse de l'expression génétique et donc celle des programmes développementaux qui en découlent (Lagha et al., 2012). L'étude des éléments de régulation du gène *svb* a contribué à l'établissement de ce concept. En effet, les enhancers qui soutiennent l'expression de *svb* dans l'épiderme dorsal de l'embryon, possèdent des fonctions redondantes mises en évidence uniquement dans un contexte de stress (Frankel et al., 2010). Cette étude consistait à observer l'impact de la déplétion d'un des enhancers connus comme actifs dans l'expression de *svb* pour la formation des trichomes de l'épiderme. L'étude a montré que la déplétion de l'enhancer n'avait pas d'impact dans des conditions normales de développement mais que si le milieu devenait stressant pour le

développement, notamment à cause d'une modification de la température, la mutation de l'enhancer entraînait une anomalie d'organisation des trichomes.

En conclusion, la présence de deux enhanceurs distincts permettrait donc d'assurer la robustesse d'expression génique et ainsi de faire face aux variations environnementales (changement de température) ou génétique (mutations) qui peuvent influencer sur l'efficacité de fixation des facteurs de transcription sur leurs enhanceurs.

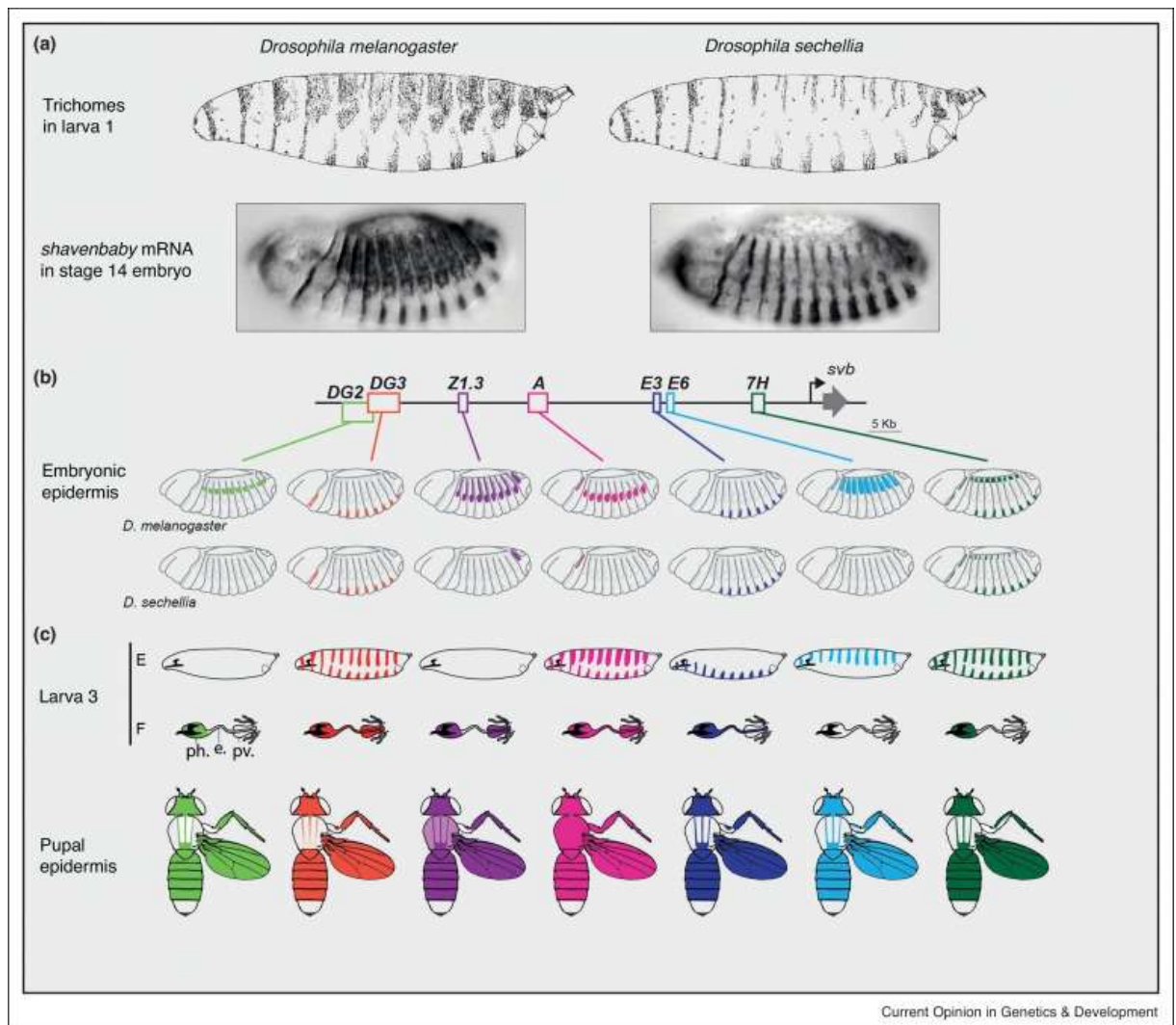


Figure 4 : La régulation de l'expression de *svb* passe par l'utilisation de plusieurs régions cis-régulatrices.

a) Les profils d'expression de *svb* ont évolué avec les espèces. En haut : Dessin des trichomes de vues latérales de larves en stade L1 de *Drosophila melanogaster* (à gauche) et de *Drosophila sechellia* (à droite). En bas : Immunomarquage de l'ARNm de *svb* dans l'embryon au stade 14 de *Drosophila melanogaster* (à gauche) et de *Drosophila sechellia* (à droite). B) Locus de *svb* avec les différents enhanceurs avec leur différents patrons d'expression chez *Drosophila melanogaster* (en haut) et *Drosophila sechellia* en bas. c) Profil d'expression des différents enhanceurs dans différents tissus à différents temps de développement. Dans la Larve L3 : Profils d'expression des enhanceurs dans l'épiderme (E) et dans l'intestin antérieur (F : ph : pharynx, e : oesophage ; pv : proventricule). Profils dans l'épiderme pupal. (Kittelman et al., 2021)

Afin d'approfondir le mode de régulation des enhanceurs de *svb*, Alexandre Ahmad Al Sawadi (doctorant dans l'équipe) a recherché, par une approche à grande échelle, les facteurs de transcription se fixant sur ces régions régulatrices et requis pour leur activité. Il a criblé par une approche RNAi 400 facteurs de transcription exprimés chez l'embryon et testé si leur déplétion influence l'activité des enhanceurs 7, E et A. Pour compléter ce travail, un crible "one-hybrid" a été entrepris pour rechercher les facteurs se fixant sur les séquences régulatrices de

svb. Cela a permis d'identifier les facteurs de transcription Pointed (Pnt, voies de signalisation EGF), Scalloped (Sd, voie de signalisation Hippo), SoxNeuro (SoxN) et Pangolin (Pan, voie de signalisation Wg) comme régulateurs de l'expression de *svb* via la régulation des enhancers E3-14 et 7 (cf. Figure 5). Ces observations ont notamment permis d'identifier des sites de fixation consensus du facteur Pangolin impliqué dans la voie de signalisation Wg (al Hayek et al., 2021). Cette étude ainsi que celle réalisée par Jérôme Bohère (Bohère et al., 2018) à laquelle j'ai participé (voir partie Résultats Chapitre 4) ont mis en évidence que les enhancers E3 et E6 soutiennent l'expression de *svb* dans les cellules souches intestinales et rénales. Cela permet également de mettre en évidence une autre notion importante sur le mode de régulation de l'expression génique et le fonctionnement des enhancers, **la pléiotropie**. En effet, les résultats acquis dans l'équipe et dans l'étude réalisée par (Preger-Ben Noon et al., 2018) mettent en évidence qu'un même enhancer peut réguler l'expression de *svb* dans différents territoires au cours du développement.

Pour résumer, *svb* intègre, via plusieurs enhancers, les voies de signalisation Wg, EGF, Notch et Hedgehog (Hh) pour déterminer le registre spatial, la pléiotropie et la robustesse de son expression au sein de l'épiderme et des cellules souches adultes (Delon et al., 2004; Stern et al., 2013; Arif et al., 2015).

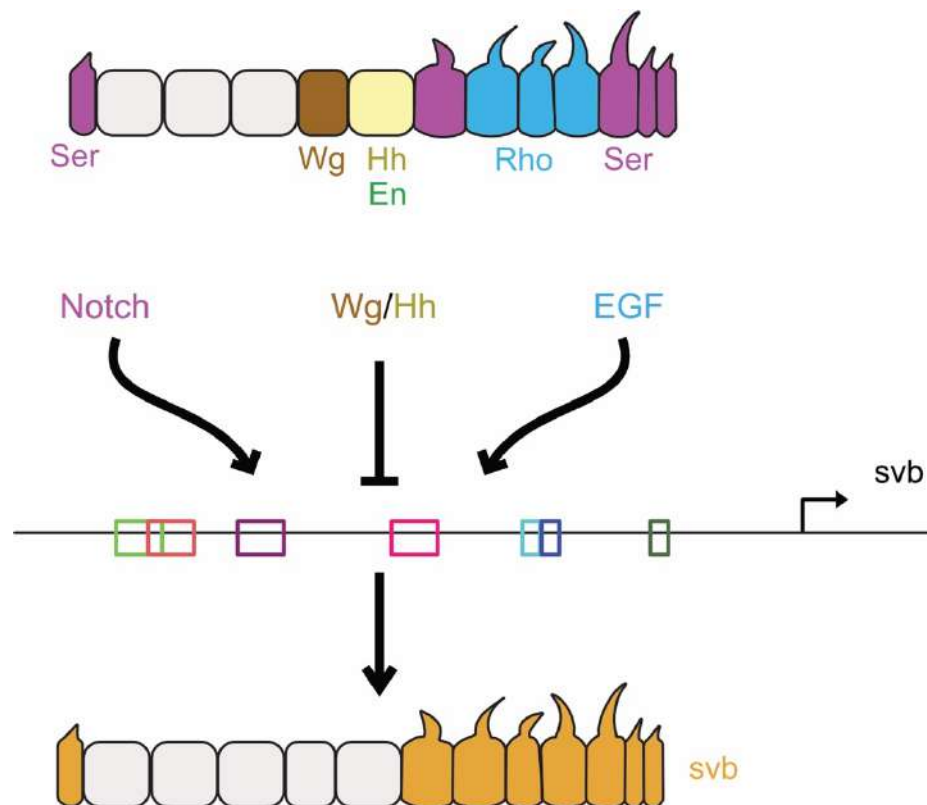


Figure 5 : La régulation de l'expression de *svb* au sein des cellules ventrales épidermiques se fait par l'intégration des voies de signalisations Notch, Wg/Hh et EGF.

Cette régulation se fait par l'action d'enhancers spécifiques de *svb*. Dans les cellules lisses, *Engrailed (En)* est activé par la voie de signalisation Wg et Hh est également présent. Les cellules qui forment les trichomes expriment les facteurs Rhomboid (Rho, voie EGF), Serrate (voie Notch). Les voies Notch et EGF active l'expression de *svb* alors que la voie Wg maintenue par Hh l'inhibe. (Adapté (Payre, 2004))

2.1.2 Régulation post-traductionnelle de l'activité de Svb

Nous avons vu précédemment qu'un promoteur alternatif, actif uniquement dans les cellules somatiques est responsable de la production d'un ARN permettant la synthèse de la protéine Svb. Cette protéine est en tout point similaire à la protéine OvoA mais comprend un exon supplémentaire dans sa partie C-terminale (cf. Figure 1). Les études initiales réalisées *in vitro* et *in vivo* suggéraient que Svb puisse être un activateur transcriptionnel bien que son potentiel soit moindre que la forme OvoB (Andrews et al., 2000; Delon et al., 2003). En effet, si l'expression ectopique de la forme OvoB dans les cellules qui composent le disque d'aile aboutit à la formation de trichomes surnuméraires, celle de Svb permet seulement un simple allongement de ces structures (Delon et al., 2003). Des travaux novateurs réalisés dans l'équipe en collaboration avec le groupe de Y. Kageyama et consécutifs à la découverte du gène *polished rice* (*pri* également connu sous le nom de *tarsal-less*, *tal*, (Galindo et al., 2007 a) sont venus affiner le mode de fonctionnement de la protéine Svb.

2.1.2.1 Identification des peptides smORF Polished-rice

Le gène *pri* présente un patron dynamique au cours de l'embryogenèse et dans l'épiderme selon un profil segmenté recouvrant celui de *svb*. De plus, un des phénotypes des embryons *pri* mutants se caractérise par l'absence de trichomes. La similarité des domaines d'expression et des phénotypes des mutants ont poussé les deux équipes à investir plus en profondeur la relation entre les deux gènes. Cette collaboration a permis de mettre en évidence que les petits peptides Pri appartiennent à la classe des peptides smORF (small open reading frame) et sont synthétisés à partir d'un ARN (Acide Ribonucléique) polycistronique initialement considéré comme non codant (Kondo et al., 2010).

Les ARNs peuvent être divisés en deux catégories, les ARNs codants pour des protéines et les ARNs non codants. Cette distinction en deux classes se fait sur la taille des cadres ouverts de lecture (Open Reading Frame ou ORF). Les ARNs sont considérés comme codant si leur ORF possède une taille supérieure à 100 acides aminés. Les autres (ORF <100 AA) sont appelés smallORF ou smORF. Ils sont considérés comme ayant des ORF trop petites pour être traduites. Pourtant des études de biochimie et plus récemment de séquençage de l'ARN (RNA-seq) mais aussi de profilage de ribosome ont permis de mettre en évidence que certains de ces smORFs pouvaient être traduits en peptides.

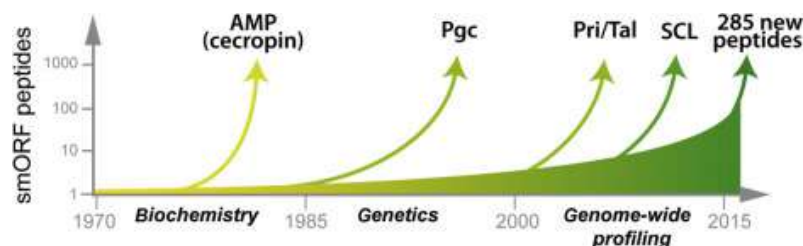


Figure 6 : La recherche sur les smORFs est en pleine expansion.

La recherche concernant les smORFs débute dans les années 80 avec l'identification de l'AMP pour connaître une expansion exponentielle avec le développement de technique d'analyse par approche bio-informatique (Zanet et al., 2016)

Ainsi des ARNs non codant sont en fait capables de produire des smORF peptides à l'instar de l'ARN de *pri*. Depuis, il a été découvert que des peptides smORF peuvent être générés à partir d'ARNm (ARN messager) qui ont été nommés en fonction de la position de leur ORF au sein du locus génique (Zanet et al., 2016; Guerra-Almeida et al., 2021). Ces peptides smORF constituent un réservoir important de molécules bio-actives dont les

fonctions restent à définir (cf. Figure 6 & Figure 7) (Zanet et al., 2016; Plaza et al., 2017; Guerra-Almeida et al., 2021).

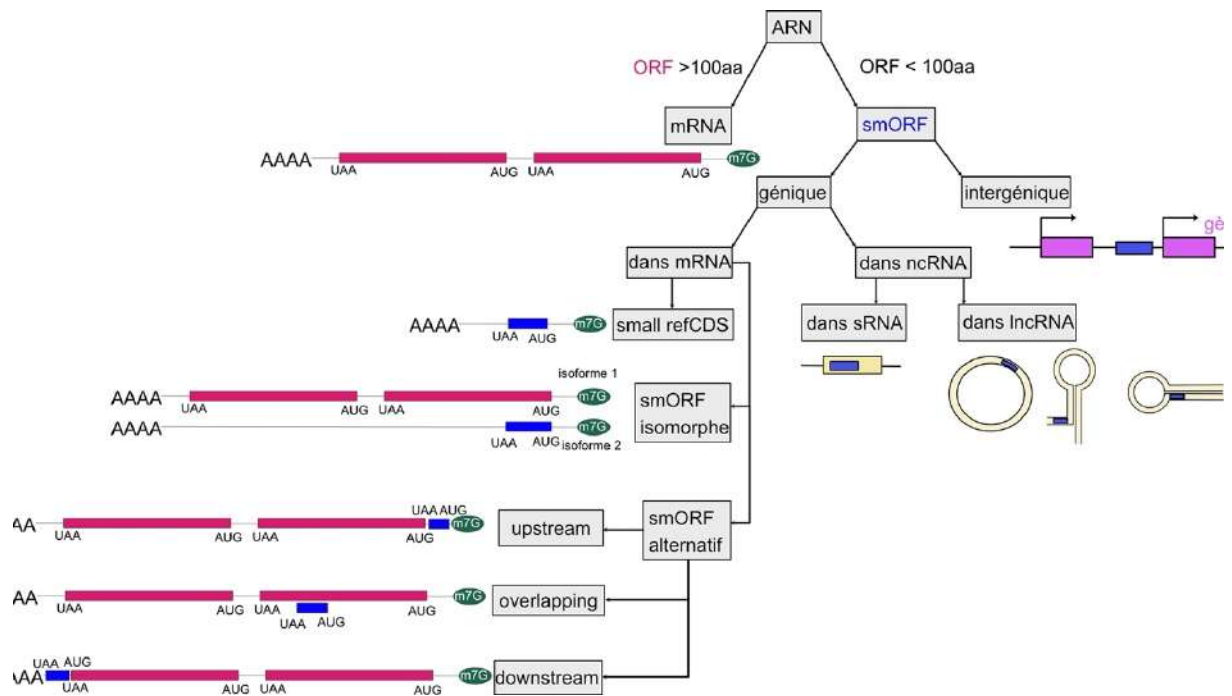


Figure 7 : Classification des smORF

Le gène *pri* est transcrit en un ARN polycistronique ayant 5 cadres de lectures, les quatre premiers sont traduits en petits peptides dits Pri de 11 à 32 acides aminés, le dernier est non-traduit, mais du fait de sa plus grande taille, il a permis de suivre la conservation du locus chez douze espèces de *Drosophile* (Savard et al., 2006; Galindo et al., 2007 b; Zanet et al., 2016).

Ayant mis en évidence l'existence des peptides Pri, il restait à définir leur relation moléculaire avec Svb. Sachant que, la perte de fonction de *pri* n'affecte pas l'expression de *svb* et inversement, une expérience déterminante a été de montrer que la perte de fonction de *pri* pouvait être sauvée par la forme activatrice OvoB mais pas par Svb (Kondo et al., 2010).

2.1.2.2 Shavenbaby : deux protéines pour le prix d'une

La protéine Svb est synthétisée sous une forme longue, qui possède en sa région N-terminale, un exon additionnel dit 1S par rapport à la protéine OvoA (cf. Figure 1 et Figure 8). La présence des peptides Pri induit l'élimination de la partie N-terminale de Svb qui contient l'exon 1S et le domaine répresseur de OvoA. Le produit de cette réaction est une protéine plus courte nommée SvbAct (cf. Figure 8). Des expériences réalisées en culture cellulaire (essai luciférase) ont permis de mettre en évidence que SvbAct agit en qualité d'activateur transcriptionnel tout comme OvoB. La protéine Svb pleine taille, SvbRep semble se comporter comme un répresseur transcriptionnel tout comme OvoA qui est insensible à l'activité des peptides Pri (Kondo et al., 2010).

A partir d'un crible RNAi réalisé sur des cellules S2, Jennifer Zanet et Emilie Benrabah ont mis en évidence le mécanisme moléculaire à l'origine de la conversion de SvbRep en SvbAct dépendante du protéasome (Zanet et al., 2015). La fixation des peptides Pri sur l'ubiquitine E3 ligase Ubr3 semble modifier la conformation d'Ubr3 et permettre sa fixation à

SvbRep. L'ajout d'ubiquitine au niveau de 3 lysines présentes dans l'exon 1S de SvbRep permet son adressage au protéasome et sa dégradation partielle. En effet, seule la partie N-terminale contenant l'exon 1S et le domaine de répression est dégradée, libérant ainsi une protéine plus courte activatrice SvbAct (cf. Figure 8).

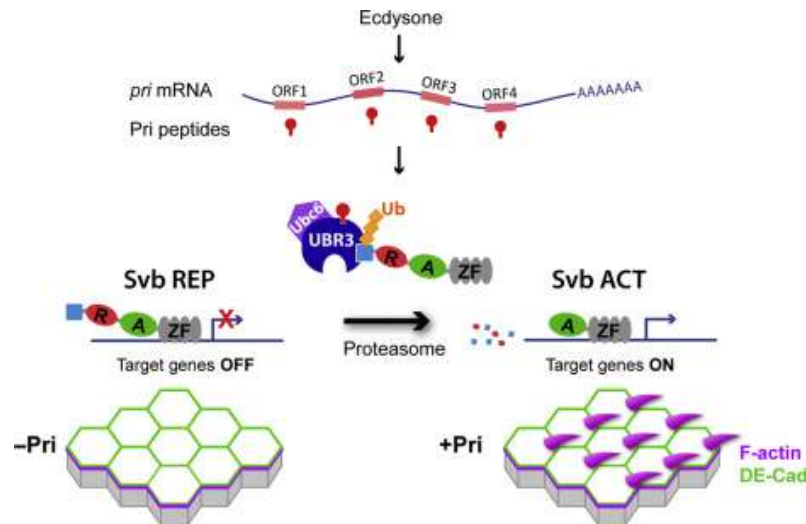


Figure 8 : Processus de maturation de Shavenbaby.

En présence d'ecdysone, l'expression du gène *pri* est activée et il y a synthèse de quatre peptides de petites tailles smORF (11 à 32 acides aminés) appelé Pri. La protéine longue SvbRep en présence de Pri est prise en charge par l'ubiquitine ligase E3 Ubr3. SvbRep est ubiquitinilé et envoyé au protéasome pour subir une dégradation partielle de la partie N-terminale. La forme SvbAct est ainsi produite. Seule la forme SvbAct promeut la formation de trichomes (Zanet et al., 2016).

Afin de décrypter le mécanisme d'activation de Svb, deux lignées cellulaires stables ont été générées à partir de cellules S2. La lignée 1B permet l'induction de SvbAct et la lignée FS, celle de SvbRep. Dans les deux lignées cellulaires, l'induction de SvbRep et de Pri s'effectue sous le contrôle d'un promoteur inducible (pMT) aux métaux lourds (cuivre). Dans la lignée FS (Frame Shift), l'ARN correspondant à *pri* est muté pour induire un décalage de cadre de lecture et donc la non production des peptides Pri. Ces lignées ont constitué la base de mon travail de thèse pour caractériser la fonction et le mode d'action de SvbRep et ACT (cf. Figure 9). (Kondo et al., 2010; Zanet et al., 2016)

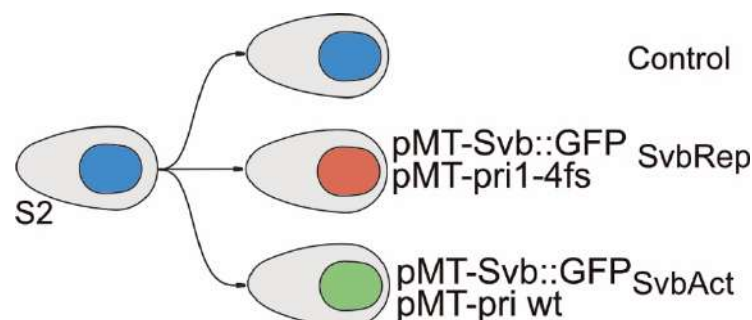


Figure 9 : Lignées stables exprimant les deux formes de Svb issues de la modification de cellules S2.

Svb est absent des cellules Contrôles (Ctrl : bleu). La lignée SvbRep (rouge) expriment Svb couplé à la GFP et une forme mutée de *pri* sous le contrôle d'un promoteur inducible. Dans cette lignée, les peptides Pri sont absents donc Svb reste sous sa forme Répresseur. Dans la lignée SvbAct, SvbGFP et *pri* sont tous deux exprimés, il y a donc présence de la forme Activatrice de Svb.

De façon intéressante, au sein de ces lignées cellulaires, il a été observé que SvbRep et SvbAct se localisent différemment au sein du noyau. Si SvbAct diffuse dans tout le noyau,

SvbRep se concentre sous forme de foci (Kondo et al., 2010) suggérant que l'organisation de la chromatine puisse être différente dans les deux contextes cellulaires.

2.2 Rôle de Svb dans l'épiderme

Les travaux initiaux réalisés par François montrent que l'expression de *svb* est à la fois suffisante et nécessaire à la formation des trichomes (appelés aussi denticules chez les larves) présente à l'apex des cellules épidermiques (Payre et al., 1999). Ces trichomes sont la résultante d'extensions cytoplasmiques générés par un spectaculaire réarrangement du cytosquelette d'actine (Chanut-Delalande et al., 2006; Fernandes et al., 2010). Afin de mieux appréhender le rôle de Svb dans l'architecture tridimensionnelle des cellules épidermiques embryonnaires, l'équipe a entrepris de rechercher les gènes cibles de ce facteur de transcription par des approches candidates ou globales (Chanut-Delalande et al., 2006; Menoret et al., 2013).

2.2.1 Régulation centrale des gènes effecteurs du remodelage épidermique

L'approche candidate a consisté à rechercher sur les bases de données disponibles, les gènes dont le profil d'expression correspondaient à celui du gène *svb*, c'est-à-dire un profil segmentaire dans l'épiderme embryonnaire. Les gènes dont les mutations affectent la formation des trichomes tels que *singed* (*sn*) ont été également considérés. Cette approche candidate a permis de mettre en avant que *svb* régule l'expression d'une quinzaine de gènes codant tous pour des effecteurs cellulaires et impliqués dans l'organisation du cytosquelette d'actine, le dépôt de la cuticule et la matrice extracellulaire (Figure 10) (Chanut-Delalande et al., 2006; Fernandes et al., 2010).

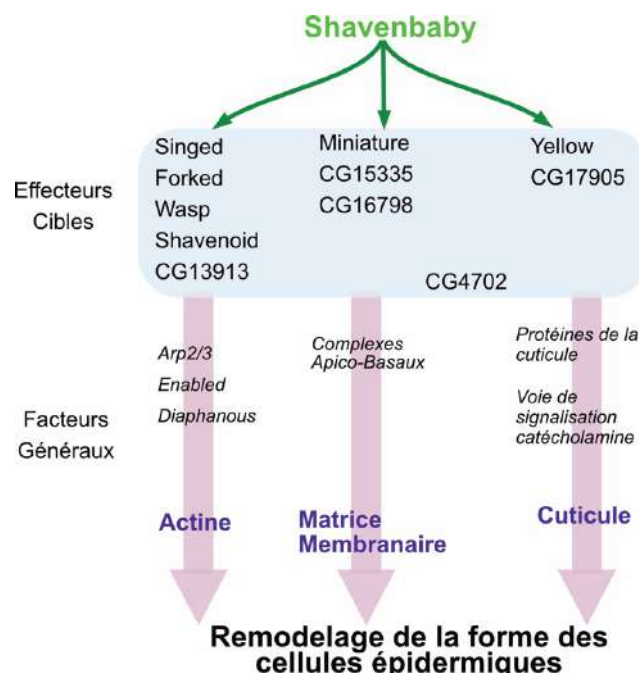


Figure 10 : Shavenbaby contrôle le remodelage des cellules épidermiques embryonnaires. Cette régulation se fait via la régulation des gènes cibles impliqués dans la régulation de l'actine, la formation de la matrice membranaire et le développement de la cuticule. (Adapté de Chanut-Delalande et al., 2006)

Les données acquises avec l'approche candidate ont servi de guide pour l'identification d'une cohorte plus large de gènes cibles de *svb*. Cette deuxième approche a consisté à identifier les gènes cibles directs de *svb* en combinant une analyse transcriptomique (puces Affymetrix) avec des expériences de fixation à l'ADN (ChIP-seq) (Menoret et al., 2013). Cette approche a permis de mettre en évidence que Svb au sein des cellules épidermiques contrôle une cohorte de ± 150 gènes cibles dont une majorité n'a pas encore été caractérisée fonctionnellement (Menoret et al., 2013). Elle confirme également que Svb contrôle l'expression d'effecteurs cellulaires impliqués pour l'essentiel dans le remodelage du cytosquelette, la cuticule et la matrice extracellulaire.

Cette analyse nous a aussi éclairés sur le mode de fixation de Svb sur les enhancers. En effet, Svb se fixe sur un motif plus sélectif que le motif Ovo dit OvoQ6 (CnGTTa) et défini par Selex (Lee et al., 2000). Ce motif, nommé SvbF7 (ACHGTTAK), chevauchant le motif OvoQ6 fut identifié par un algorithme permettant de discriminer des enhancers de gènes cibles et des régions négatives. Deux autres motifs, appelés motifs « blue » ou « yellow » ont été trouvés pour être surreprésentés dans les éléments cis-régulateurs fixés par Svb (cf. Figure 11). Dans de nombreux cas, l'introduction d'une mutation ponctuelle dans le site SvbF7 est suffisante pour affecter de manière significative l'activité des enhancers régulés par Svb (Menoret et al., 2013). De manière générale, si la fixation de Svb nécessite la présence d'au moins un site de liaison à Svb, il n'y a pas de signe évident de regroupement de ces sites (clustering) sur les régions fixées par Svb et situées à proximité de ses gènes cibles comme cela peut être le cas pour les facteurs Dorsal ou Bicoïd (Markstein et al., 2002; Chen et al., 2012).

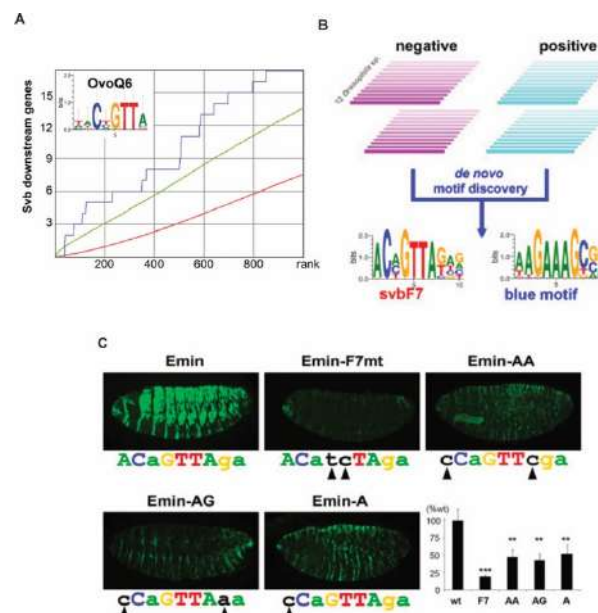


Figure 11 : Détermination des sites de fixation de Shavenbaby dans les cellules épidermiques embryonnaires. A) Courbe caractéristique d'exploitation du récepteur montrant un enrichissement significatif des sites de liaison putatifs de Svb (matrice de poids de position OvoQ6) parmi les 39 gènes en aval de Svb (axe des ordonnées) par rapport à un ensemble aléatoire de 1 000 gènes de drosophile (axe des abscisses) en utilisant cisTargetX. La courbe bleue montre la détection des gènes en aval de Svb, la courbe rouge une distribution aléatoire, et la courbe verte un intervalle de 2 sigma par rapport à l'aléatoire. B) Une analyse statistique des enhancers positifs par rapport aux régions négatives (séquences génomiques intergéniques utilisées comme contrôle) a été réalisée pour la découverte de novo de motifs, montrant la conservation évolutive entre les espèces de drosophiles et les caractéristiques des enhancers actifs. Les motifs svbF7 et bleus sont les plus performants. C) Alors que la perturbation du motif central CnGTT OvoQ6 abolit l'activité Emin, les mutations ponctuelles qui affectent les nucléotides flanquants 5' et 3' réduisent fortement l'expression épidermique, comme le montrent l'immunocoloration anti-lacZ et la quantification des signaux de fluorescence. wt, type sauvage. Les barres d'erreur représentent l'écart-type. *** Valeur $P < 0,001$; ** Valeur $P < 0,01$. (Menoret et al., 2013)

2.3 Shavenbaby dans le système digestif

2.3.1.1 Le système digestif

Le système intestinal est un des plus gros organes de l'organisme. Il possède la plus grande surface en contact avec le milieu extérieur. Il remplit non seulement les fonctions d'absorption mais il participe aussi à l'immunité (Colombani et al., 2020) et a des effets sur le comportement. Les mâles n'ayant plus de microbiome intestinal sont moins agressifs envers les autres mâles et moins compétitifs lors des accouplements. Ces mâles ne présentent aucun défaut de locomotion. Ces défauts de comportements sont contrecarrés en restaurant le microbiome intestinal de ces mâles (Y. Jia et al., 2021). La Drosophile adulte se nourrit essentiellement de fruits en décomposition ce qui l'expose à de nombreux pathogènes pouvant altérer l'intégrité des cellules intestinales. Afin de pallier la mort des cellules endommagées et de maintenir une homéostasie correcte de l'organe, il est important de pouvoir assurer leur renouvellement. Les cellules souches intestinales identifiées en 2006 (Micchelli et al., 2006 a; Ohlstein et al., 2006 a) soutiennent ce renouvellement actif et constant. Pour cela, elles répondent à des stimulations locales, tissulaires et systémiques. Par sa relative simplicité et la forte conservation des mécanismes qui soutiennent l'activité des cellules souches, le système intestinal de la Drosophile est un modèle d'étude adapté pour étudier le renouvellement cellulaire.

2.3.1.1.1 Caractéristiques structurelles et fonctionnelles du système digestif

Chez le jeune adulte, l'intestin est encore immature et sa maturation nécessite trois jours pour qu'il atteigne sa taille adulte ($\approx 6\text{mm}$) (Micchelli et al., 2006 b; Jiang et al., 2012) . Il sera alors constitué d'un épithélium monostratifié permettant l'absorption des nutriments et de barrières contre les pathogènes externes. Chez la femelle, l'activité de reproduction viendra parfaire son fonctionnement et adapter sa taille et son fonctionnement avec le grand besoin énergétique requis pour l'ovogenèse (Hudry et al., 2016).

La taille de l'intestin est maintenue tout au long de la vie de l'individu, et les pertes cellulaires sont compensées pour assurer l'homéostasie de l'organe au cours du temps.

Le système digestif chez la Drosophile est composé de quatre parties :

- **L'intestin antérieur**, d'origine ectodermique, incluant la bouche, le pharynx, l'œsophage et le jabot. Il permet le mixage, le stockage de la nourriture et la détoxification.
- **L'intestin moyen**, d'origine endodermique, spécialisé dans la digestion et l'absorption des nutriments.
- **L'intestin postérieur**, d'origine ectodermique, ayant pour fonction la fin de l'absorption des nutriments, l'ajustement de la quantité d'eau et d'électrolytes.
- **Les tubules de Malpighi** sont reliés à l'intestin au niveau de la jonction entre l'intestin moyen et postérieur. Ils assurent la grande partie des fonctions rénales et sont considérés comme un organe à part entière.

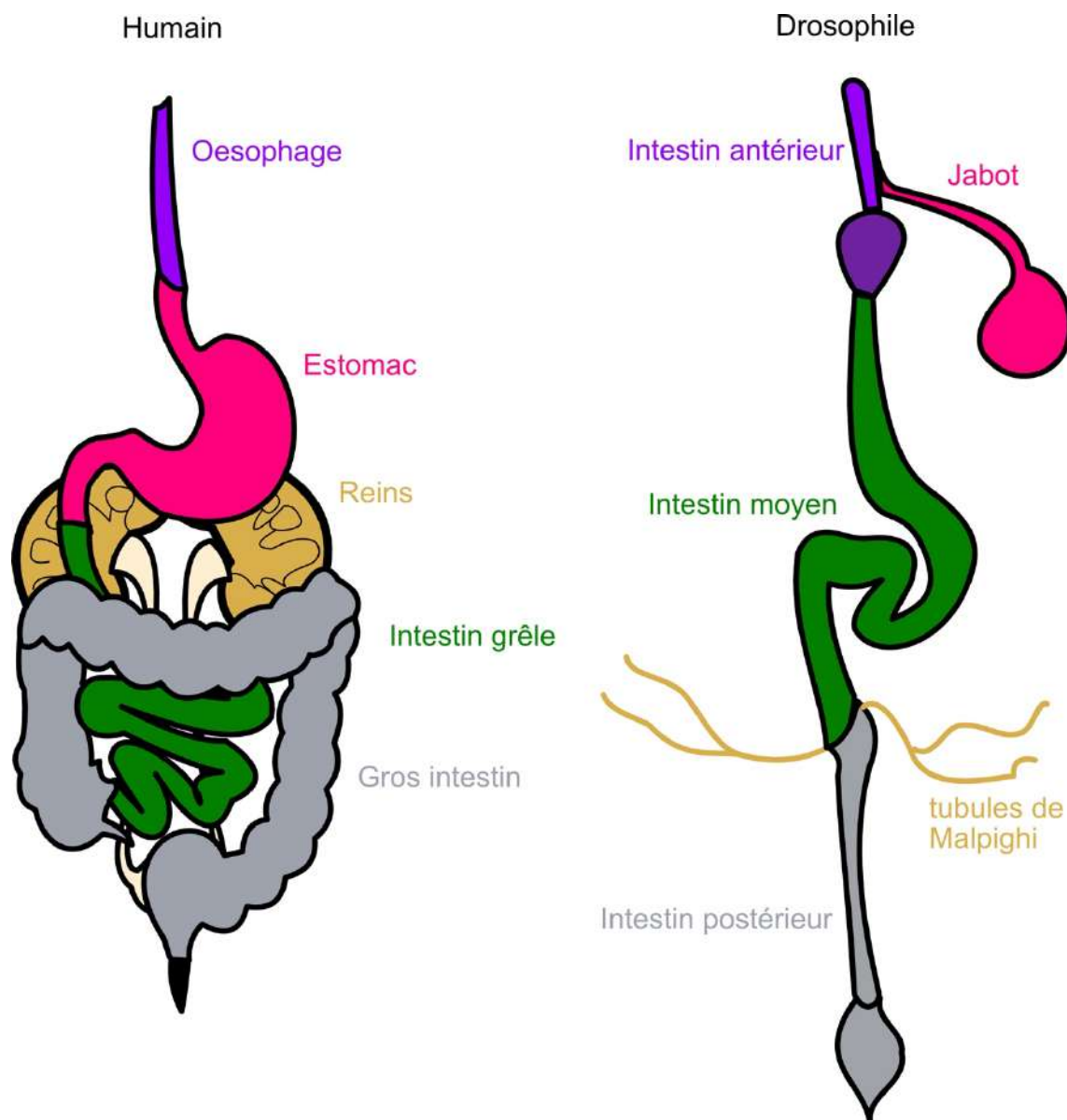


Figure 12 : Composition et comparaison des systèmes digestifs humain (gauche) et de drosophile (à droite). Les structures principales et fonctions sont conservées. L'œsophage humain correspond à l'intestin antérieur de la drosophile (en violet) ; l'estomac correspond au jabot de la mouche (rose) ; l'intestin moyen de drosophile correspond à l'intestin grêle (vert) ; le gros intestin correspond à l'intestin postérieur (gris) et les reins correspondent aux tubules de Malpighi.

La structure du système digestif est maintenue entre la Drosophile et l'Humain. L'intestin antérieur correspond à l'œsophage, le jabot à l'estomac, l'intestin moyen à l'intestin grêle, l'intestin postérieur au gros intestin et les tubules de Malpighi aux reins (cf. Figure 12). Cette conservation est tant structurelle que fonctionnelle faisant ainsi de la Drosophile un modèle essentiel dans l'étude des pathologies touchant le système digestif comme les cancers colorectaux, les néphropathies ou les coliques néphrétiques (Rodan, 2019; Singh et al., 2019; Cohen et al., 2020)

2.3.1.1.1.1 *L'intestin moyen*

L'intestin moyen qui partage de fortes homologues fonctionnelles avec l'intestin grêle des mammifères est la partie la plus étudiée pour comprendre les mécanismes du renouvellement soutenus par les cellules souches. Il est composé de cinq régions différenciées les unes des autres en fonction de leurs structures, leurs compositions enzymatiques mais aussi en peptides antimicrobiens et leurs pH (Buchon et al., 2015). Il est admis de façon un peu simpliste (voir chapitre suivant) que l'intestin moyen est constitué de deux types de cellules différenciées : les entérocytes (EC) et les entéro-endocrines (EE). Les entérocytes, polyploïdes qui représentent la majorité des cellules de l'intestin moyen, sont spécialisées dans l'absorption et la sécrétion des enzymes digestives. Les entéro-endocrines assurent, quant à elles, la sécrétion des hormones et sont impliquées dans l'immunité. Les cellules souches intestinales (CSI) sont le troisième type cellulaire présent dans cette région de l'intestin. Elles ont la capacité de proliférer et de se différencier pour renouveler en fonction des besoins les EC et les EE. Ce renouvellement se fait via des stades progéniteurs appelés respectivement entéroblastes (EB) pour les EC et pré-EE pour les EE (Micchelli et al., 2006 b, 2011; Ohlstein et al., 2006 b; Guo et al., 2016 a).

2.3.1.1.1.2 *Le lignage cellulaire*

La découverte des CSI par les groupes de Norbert Perrimon et celui d'Allan Spradling en 2006 a permis d'établir les bases de la compréhension des mécanismes de renouvellement de l'intestin (Micchelli et al., 2006 b; Ohlstein et al., 2006 b). Les CSI expriment notamment les facteurs Escargot (Esg) et Delta (Dl, ligand de la voie Notch). Sur la base de ces études, le modèle admis jusqu'à récemment est qu'une cellule souche intestinale donne naissance à une cellule au statut hybride : l'entéroblaste (EB) qui en fonction de la quantité de signal Notch reçu se différencie soit en EC soit EE (Bardin et al., 2010; Boumard et al., 2021). Ce modèle simple a été toutefois complexifié par des études qui mettent en avant l'existence d'un autre type cellulaire intermédiaire permettant la différenciation des EE, les cellules pré-étéroendocrine (pré-EE) (Biteau et al., 2014; Guo et al., 2015 d, 2016 b; Hung et al., 2020; Li et al., 2021). Ainsi, la division des CSI peut donner naissance à des EC en passant par le stade intermédiaire EB ou à des EE par l'intermédiaire des cellules pré-EE. Les données récentes issues d'analyse de transcriptomes à partir de cellules uniques (single-cell RNA-seq ou scRNA-seq) viennent appuyer et compléter cette version du lignage intestinal (cf. Figure 13). En effet, si la genèse des types cellulaires est conforme au modèle, ces études nous révèlent la diversité et la complexité des EC et des EE présentent dans l'intestin moyen¹⁴ (Hung et al., 2020). L'ensemble de ces données sont également disponibles dans la base de données de transcriptomique en cellule unique dans différents tissus mise en place par un consortium de différents groupes de chercheurs : le Fly Cell Atlas¹⁵ (Li et al., 2021).

¹⁴ <https://www.flyrnai.org/scRNA/>

¹⁵ <https://flycellatlas.org/>

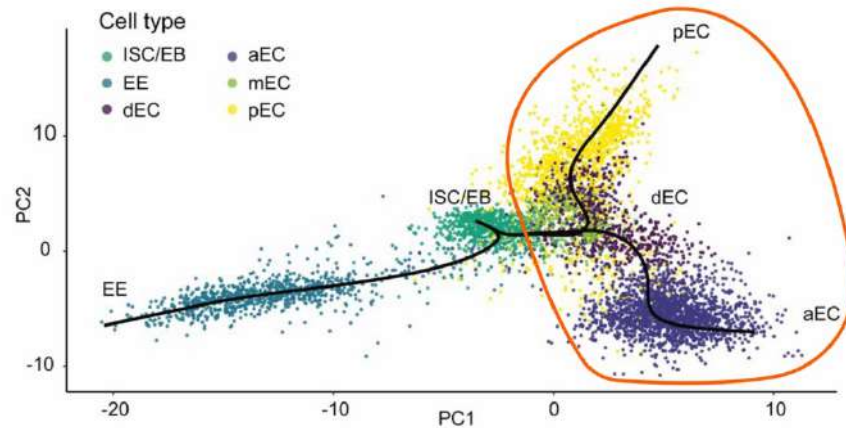


Figure 13 : 0 Lignage en cellule unique des cellules intestinales.

Pseudotime issu de l'analyse scRNA-seq, on peut distinguer 3 grands groupes les ISC/EB ; les EE et les EC (entouré en orange). Parmi les EC, il y a des sous-clusters les dEC qui sont les EC en cours de différenciation qui présentent des caractéristiques d'ISC/EB et de EC, les aEC correspondant aux EC de l'intestin antérieurs, les pEC pour l'intestin postérieur et les mEC pour les EC de l'intestin moyen. On peut que les ISC donnent les EE et les EC. (Hung et al., 2020)

Si le gène *esg* fut le premier marqueur des cellules souches, il fallut des années pour comprendre que sa fonction est de réguler négativement l'expression des gènes de différenciation en EC tel que le gène *Pdm1/nubbin* (Korzelius et al., 2014). La régulation de l'expression et donc de l'activité d'*esg* fait intervenir des facteurs participant au processus de transition épithélio-mésenchymateuse (EMT) dont le micro-ARN *mir8* (homologue de mir200) et *Zfh1* (homologue ZEB1&2). La fonction majeure de *mir8* est de déstabiliser l'ARNm d'*esg* et donc de promouvoir la différenciation des EC (Antonello et al., 2015). Cette étude a permis de proposer un modèle de plasticité épithéliale dans lequel les cellules souches intestinales seraient dans un état hybride Epithélial-Mésenchyme donnant naissance à un EB de type mésenchyme, capable de se déplacer au sein de l'épithélium pour assurer le renouvellement des EC de type épithélial.

L'étude des cellules souches intestinales et de leur devenir est donc bien plus complexe qu'initialement pensé et devrait encore permettre d'adresser des questions sur le lignage cellulaire et la plasticité cellulaire.

2.3.1.2 Fonction antagoniste de Shavenbaby pour le maintien de l'homéostasie de l'intestin

Lors de la caractérisation des régions de l'intestin moyen de Drosophile, Dani Osman a pu observer que les gènes cibles du facteur de transcription de Shavenbaby dont *miniature* sont exprimés dans le jabot. Grâce à des lignées rapportrices générées et disponibles dans l'équipe il a été observé que *svb* est exprimé dans toute les cellules souches digestives : intestinales et rénales. Cette observation a donné lieu à l'émergence dans l'équipe de deux projets visant à caractériser le rôle de *svb* dans les :

- les Cellules souches intestinales, projet initié par Sandy Al-Hayek sous la supervision de François Payre.
- les Cellules souches rénales, travail porté par Jérôme Bohère et Cédric Polesello.

Les résultats obtenus par Jérôme lors de sa thèse, ayant donné lieu à une publication à laquelle j'ai participé, seront présentés dans la partie Résultats Chapitre 4 de ce manuscrit. Les résultats et observations effectués par Sandy au sein de l'intestin constituent un élément de départ à une partie de mes travaux.

Les données d'expression soutiennent que *svb* est présent dans les CSI ainsi que dans les EC (Buchon et al., 2015; Hung et al., 2020; al Hayek et al., 2021; Li et al., 2021). L'expression dans les cellules souches est sous le contrôle positif des voies de signalisation Wnt et EGF qui agissent sur les deux enhanceurs de la boîte E, E3 et E6 pour réguler l'expression de *svb*. Le fait que deux enhanceurs différents soutiennent l'expression dans les CSI suggère qu'ils puissent agir de manière redondante pour assurer l'expression de *Svb* dans ces cellules. Des expériences préliminaires réalisées par Cédric Polesello soutiennent cette hypothèse puisqu'en condition de stress (15 jours à 29°C), la perte de l'enhancer E3 semble réduire le nombre de CSI. En effet, la perte de fonction de *svb* dans les CSI a pour conséquence de réduire progressivement leur nombre pour arriver vers 30 jours à leur complète disparition. La perte des CSI est notamment le fruit d'une augmentation significative de l'apoptose, phénotype qui peut partiellement être sauvé par l'expression de la protéine anti-apoptotique DIAP1. A l'inverse, l'expression de *SvbAct* dans les CSI induit une forte prolifération de ces cellules et celle de *SvbRep* pousse les CSI vers la différenciation en EC, révélant l'antagonisme fonctionnel entre les deux formes de *Svb*. Les études réalisées par Sandy ont permis de mettre en évidence que la régulation post-traductionnelle de *Svb Pri* dépendante se produit dans les CSI par un processus similaire à ce qui se produit dans les cellules épidermiques. Ainsi, la perte de fonction de *pri* et d'*ubr3* conduit également à la disparition progressive des CSI. Cette perte peut être sauvée par la réintroduction de *SvbAct*. La régulation de *pri* et la maturation de *Svb* est assurée par la production d'Ecdysone par les chambres ovariennes en fonction des conditions nutritives. La production d'Ecdysone régule ainsi de manière systémique le comportement des CSI (Ahmed et al., 2020; Zipper et al., 2020). Il est aussi important de noter que la perte de fonction de *svb* est épistatique sur la prolifération consécutive à la dérégulation des voies de signalisation Notch, Jak/STAT et Wg. A l'opposé, la présence de *SvbAct* est capable d'inhiber le processus de différenciation induit par la surexpression du micro-ARN *mir8* révélant si besoin l'antagonisme entre *Svb* et les facteurs pro-EMT.

Svb est aussi présent dans les EC où son expression médiée par l'enhancer proximal 9CJ2 est dépendante du facteur de différenciation *Pdm1* (*nubbin*). Dans ces cellules dépourvues de *pri*, *Svb* s'accumule en foci suggérant que la forme *SvbRep* est dominante dans ces cellules et participe au maintien de leur caractère épithélial. En effet, l'expression précoce de *SvbRep* dans les CSI favorise leur différenciation en EC.

En conclusion, *SvbAct* protège les CSI de l'apoptose et promeut leur prolifération alors que *SvbRep* induit leur différenciation en EC pour lesquels elle garantit l'intégrité épithéliale. L'intestin offre donc un modèle intéressant pour adresser la dualité fonctionnelle de *SvbRep* et *SvbAct* et caractériser plus en profondeur les spécificités de *SvbRep* seul présent dans les EC. Au cours de ma thèse, en collaboration avec Clément Immarigeon (postdoc-génération de données RNA-seq) et Alexia Rivero (M2 Bio-informatique et Biologie des Systèmes), nous avons identifié les gènes cibles de *Svb* dans les CSI. Ce travail sera présenté dans la section Résultats Chapitre 6.

Chapitre 3 Au croisement entre l'informatique et la biologie : la bio-informatique pour la génomique chez la drosophile

L'équipe de François Payre est spécialiste dans l'étude du facteur de transcription Shavenbaby en utilisant à la fois des approches de biologie moléculaire et cellulaires. Bien que déjà utilisées lors de la détermination des gènes cibles embryonnaires de Shavenbaby (Menoret et al., 2013) les analyses bio-informatiques étaient réalisées en collaboration. Avec mon arrivée pour mon stage de fin d'études en Master 2, ces approches ont été internalisées au sein de l'équipe. J'ai ainsi intégré une équipe de biologistes expérimentateurs avec mon bagage particulier de bio-informaticienne pour cette aventure du doctorat.

La Bio-informatique, en mon sens, est un domaine de la Biologie tout autant que la Biologie Cellulaire, la Biologie Moléculaire ou autre. La considérer uniquement comme des techniques ou de l'ingénierie serait réducteur et contre-productif.

Dans ce chapitre, je vais tenter de vous expliciter ce point de vue sur la bio-informatique et de vous présenter les concepts et approches que j'ai mis en place tout au long de mon doctorat. Dans une première partie, je vous présenterai le développement de la Bio-informatique et son histoire, qui contrairement à ce que l'on pourrait penser est bien antérieure aux ordinateurs. Puis, je me concentrerai sur la Drosophile et vous présenterai les concepts d'enhancer, de régulation de la transcription via ces régions cis-régulatrices et les approches de génomique (RNA-seq ; CHIP-seq) que j'ai utilisées. Pour finir, la troisième partie de ce chapitre traitera de la conformation spatiale de l'ADN et en particulier de l'épigénomique en me concentrant sur l'étude des modifications des histones.

3.1 Le séquençage : De Sanger au Séquençage Nouvelle Génération

Le dictionnaire Larousse donne comme définition au mot bio-informatique : « nom féminin et adjectif. Application de la recherche en informatique au progrès des connaissances dans les sciences de la vie ». Cette définition, en mon sens, est peu précise et même confondante, réduisant ainsi la bio-informatique à juste de l'application. Je vais tenter dans cette partie de vous transmettre mon point de vue sur cette discipline complexe mais essentielle à l'interface entre deux mondes, dont les façons de penser sont parfois antagonistes. Pour cela, il me faut vous expliciter le développement de la bio-informatique.

Le terme de « bio-informatique » a été utilisé pour la première fois par Paulien Hogeweg lors d'une conférence à Berlin en 1970, où elle et son collègue Ben Harpen introduisirent « Bioinformatica een werkconcept » (Hesper et al., 1970) qui pourrait se traduire par « *La Bio-informatique, un concept de travail* ». Elle explicitera cette définition quelques années plus tard dans sa publication *The Roots of Bioinformatics in Theoretical Biology* (Hogeweg, 2011) en disant qu'à cette époque, elle considérait déjà qu'elle voulait procéder à « *l'étude des processus informationnels dans les systèmes biotiques* ». De plus, elle mettait en avant que l'une des propriétés de la vie était le traitement et la transmission de ces informations présentes sous diverses formes : accumulation au cours de l'évolution, transmission via l'ADN, signalisation intra et extracellulaire et interprétation de ces informations. Les systèmes biologiques sont donc une succession d'informations sous plusieurs formes qu'il faut décortiquer et analyser. Paulien Hogeweg pensait que ces études

au même titre que la Biophysique ou la Biochimie devaient être considérée comme une discipline à part entière. Pour Paulien Hogeweg, le cœur de la recherche en bio-informatique est l'intégration des analyses des variations d'information dans divers organismes, la relation entre génotype, comportement, environnement et phénotype, ainsi que la détection de phénomène émergeant dans la modélisation mais aussi la comparaison du modèle avec les observations *in vivo* (Hogeweg, 2011). Point de vue que je partage.

Bien que le terme de Bio-informatique soit apparu en 1970, il est admis que son origine est bien plus ancienne avec le développement du séquençage par Frederick Sanger entre 1945 et 1951 (Sanger et al., 1951; Sanger, 1988). Sanger et ses collaborateurs, en s'appuyant sur les caractéristiques des acides aminés, ont déterminé la composition par chromatographie de l'insuline et ont ainsi pour la première fois séquencé une protéine. Les acides aminés ont tous des propriétés physico-chimiques différentes, ils migrent ainsi différemment sur le papier de chromatographie. C'est sur cette observation que s'appuie la méthode de Sanger pour le séquençage protéique (Sanger et al., 1951).

Pour anecdote, dans sa publication biographique, Frederick Sanger avoue que le choix d'utiliser l'insuline s'est fait car à cette époque il s'agissait de la seule protéine que l'on pouvait se procurer facilement sous sa forme purifiée, et non comme la légende le veut pour ses propriétés biomédicales (Sanger, 1988). Frederick Sanger reçoit le Prix Nobel de Chimie en 1958 pour le développement du séquençage protéique.

Durant cette même période, Rosalind Franklin observe par diffractométrie aux rayons X la structure en double-hélice de l'Acide Désoxyribonucléique (ADN). L'ADN est composé de quatre bases nucléotidiques : Adénine, Thymine, Cytosine, Guanine (WATSON et al., 1953). Ces travaux seront publiés par James Watson et Francis Crick en 1953. James Watson et Francis Crick reçoivent le Prix Nobel de Médecine en 1962, Rosalind Franklin ne le reçoit pas puisqu'elle décède quatre ans plus tôt d'un cancer, probablement lié à l'exposition aux radiations lors de ses travaux de recherche.

À cette époque, les chercheurs pensaient que la transmission d'information de génération en génération se faisait grâce aux protéines. En 1957, Francis Crick vient perturber cette hypothèse en proposant la théorie du Dogme Central de la Biologie Moléculaire. L'ADN est la molécule portant l'information héréditaire, celle-ci se réplique pour donner à nouveau une molécule d'ADN, l'ADN est également transcrit en ARN et cet ARN est traduit en protéine (cf Figure 14) (Crick, 1970).

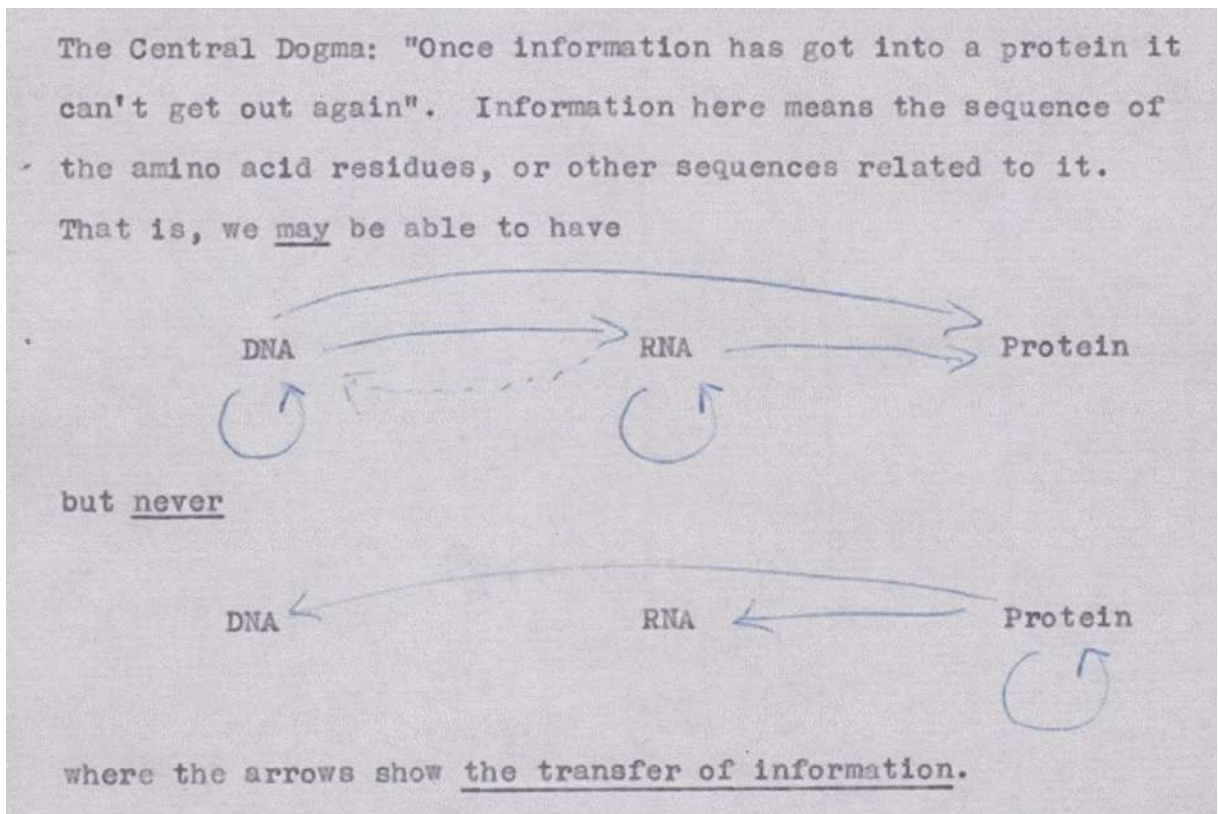


Figure 14 : Prise de notes du carnet de Francis Crick sur le Dogme Central de la Biologie Moléculaire (Cobb, 2017)

Il faudra attendre 1977 pour que la molécule d'ADN soit séquencée. Cette fois encore, cette technique de séquençage a été développée par Frederick Sanger, qui reçoit son second prix Nobel pour cette découverte (Sanger et al., 1977 a)

Entre temps, les recherches sur les protéines se poursuivent.

C'est dans ce contexte que le premier programme informatique en FORTRAN est développé par Margaret Dayhoff et Robert S. Ledley (Hagen, 2000; Gauthier et al., 2019). Ce programme permet de reconstituer une séquence protéique à partir de petits fragments de séquençage chevauchants. Margaret Dayhoff, Robert S Ledley et Richard V. Eck publient le premier *Atlas of Protein Sequence and Structure* en 1965. Cette première édition de 100 pages contenait les séquences de 70 protéines (des cytochromes C, des hémoglobines et des fibrinopeptides) de diverses espèces. Chaque page donnait le nom de la protéine, de l'organisme, suivi de la séquence d'acides aminés annotée avec le système à trois lettres, mais aussi avec un nouveau système à une lettre. Cette page contenait également les méthodes de séquençage et la référence bibliographique. Cet Atlas comprenait aussi des pages de comparaison et d'alignement d'une même protéine dans divers organismes (Dayhoff et al., 1965; Hersh et al., 1967; Strasser, 2010). On peut donc considérer *l'Atlas of Protein Sequence and Structure* comme étant la première base de données protéiques.

En 1967, Pehr Edman développe le premier séquenceur protéique automatisé, en s'appuyant sur sa méthode de dégradation du résidu en N-terminale couplé à un automate. Celui-ci permet d'obtenir la séquence protéique en quelques minutes. (Hagen, 2000; Gauthier et al., 2019).

Ainsi, la bio-informatique a vu le jour non pas à l'étude de l'ADN, mais bien grâce à l'étude des protéines.

En 1977 sont mis sur le marché les premiers ordinateurs, il s'agit des Commodore PET ; Apple II ; Tanduy TRS 80. Au même moment Frederick Sanger séquence pour la première fois l'ADN (Sanger et al., 1977 a; b).

3.1.1 Le principe du séquençage de l'ADN par Sanger

L'ADN à séquencer est purifié, fragmenté et est amplifié par réactions en chaîne par polymérase (PCR) en présence d'une amorce, de désoxyribonucléotides (dNTP) - dATP, dCTP, dTTP, dGTP - en grande quantité mais aussi des didésoxyribonucléotides (ddNTP) - ddATP, ddCTP, ddTTP et ddGTP - en plus faible quantité. Les différents ddNTP sont mis dans des tubes différents. L'incorporation aléatoire de ddNTP arrête la réaction de PCR du fragment en cours de synthèse. Les produits d'amplification sont mis à migrer sur un gel d'électrophorèse et révélés par autoradiographie. Le gel est ensuite lu de bas en haut pour déterminer la séquence du brin complémentaire synthétisé par PCR, et pour conclure on en déduit par complémentarité la composition du brin à séquencer.

S'ensuit en 1979 le développement de la première suite de logiciels (package) pour le traitement et l'analyse des séquences d'ADN : le package STADEN, encore maintenu de nos jours. Il s'agit du premier package avec interface graphique permettant de reconstruire des séquences à partir des gels de séquençage et introduisant la notion de base incertaine lors d'un séquençage. Lors d'un séquençage, les bases incertaines sont annotées N. Ce package introduit les bases des programmes de traitement de séquences utilisées aujourd'hui :

- 1) Recherche de chevauchements entre les fragments lus (reads) sur le gel ;
- 2) Vérification, édition et combinaison des reads en séquences composite appelées contig ;
- 3) Annotation et manipulation des séquences (Staden, 1996).

À partir de 1984 tout s'accélère, les premiers ordinateurs de bureaux sont mis en vente, les premiers logiciels en ligne de commande sont développés, le concept de « logiciel libre » est amorcé ("Le manifeste GNU - Projet GNU - Free Software Foundation," n.d.). Le premier journal spécialisé dans le domaine de la Bio-informatique est créé : le *Journal Computer Application in BioSciences*, qui deviendra *Bioinformatics*. Ces avancées dans le monde informatique apportent des outils et concepts encore essentiels aux bio-informaticiens de nos jours. Il est ainsi courant d'utiliser des outils en ligne de commande pour le traitement des fichiers de séquençage comme les logiciels de la suite Bedtools (Quinlan et al., 2010), Deeptools (Ramírez et al., 2016) mais encore MACS (Feng et al., 2011) pour la détection de régions de fixation d'une protéine à l'ADN. De plus, la communauté bio-informatique est très attachée au concept de logiciel libre et de la mise à disposition des codes.

En parallèle, de 1986-1987, l'European Molecular Biology Laboratory (EMBL), GeneBank et la DNA DataBank of Japan fusionnent pour créer la International Nucleotide Sequence Data, regroupant ainsi les connaissances sur les séquences au niveau international et non plus juste continental, facilitant ainsi le partage et la diffusion des données génomiques. Cette fusion donnera quelques années plus tard l'émergence du Genomic Standard Consortium (Field et al., 2011). Ce Consortium regroupe des bio-informaticiens et des biologistes expérimentateurs. Il a pour mission de créer des standards pour les données génomiques pour leur intégration dans les bases de données mais aussi pour leur échange. Un standard, appelé MIGS pour « Minimum Information about Genome Sequence », a été mis en place en 2008. Les données de séquençage, pour être intégrées dans une base de données,

doivent obligatoirement être accompagnées des métadonnées concernant l'organisme, la géolocalisation et de la méthode de séquençage utilisée (Field et al., 2008).

En 1987, 1989 et 1993 sont développés trois des langages informatiques les plus usuels en bio-informatiques, respectivement, le PERL, le PYTHON (Lutz, 1996) et le R (Hagen, 2000; Gauthier et al., 2019).

Ces langages sont dits de haut niveau, interprétés et multi paradigmes, c'est-à-dire qu'ils peuvent être fonctionnels, orientés objet ou impératifs en fonction de leur utilisation (cf. Tableau 2).

Tableau 2 : Définitions des différents paradigmes informatiques

Paradigme	Fonctionnel	Orienté Objet	Procédural
Définition	Les opérations se passent via l'utilisation de fonctions que l'on peut combiner entre elles. Il faut déclarer des entrées et on obtient des valeurs de sorties.	Les opérations se déroulent en déclarant des objets. Ces objets contiennent les données sous formes d'attributs ou propriétés et le code sous forme de procédures	Les opérations se font séquentiellement les unes après les autres. Il s'agit d'une suite d'instructions.
Exemples	JavaScript ; XQuery ; etc.	Java ; Scala ; etc.	Fortran ; C ; etc.

Des packages spécifiques à leur utilisation en biologie sont disponibles : le BioPerl à partir de 1996, le Biopython depuis 2000 et Bioconductor (R) depuis 2002. Il est à noter que le Perl perd de l'importance avec l'arrivée du Python et du R car celui-ci est difficile à lire, la maintenance des logiciels en devient vite complexe.

3.1.2 Séquençage et obtention des séquences de référence

Le développement du séquençage de la molécule d'ADN de Sanger, son adaptation pour le séquençage global d'un génome (« whole genome shotgun sequencing », WGS), et le développement des capacités de mémoire des ordinateurs ont marqué le début d'une nouvelle ère. C'est à cette période que commence le séquençage des organismes modèles. Le premier génome à avoir été séquencé est celui de la bactérie *Haemophilus influenzae* (Fleischmann et al., 1995). Cette bactérie est responsable de maladies comme les otites, les méningites mais aussi les pneumonies, les vaginites et les arthrites. Son génome est de 1,8Mbp.

Le génome de la Drosophile a été séquencé en 2000 et a été le premier génome d'eucaryote à avoir été séquencé. Le séquençage de ce génome a permis d'optimiser la technique du WGS aux eucaryotes (Adams et al., 2000; Myers et al., 2000; Celniker et al., 2003).

Le projet de séquençage du génome humain a débuté en 1991 avec la création du Human Genome Project, avec un séquençage effectif de 1999 à 2001. Une première version du génome a été publiée en 2001.

Le génome humain fait 3,272Mbp pour \approx 26000 gènes.

Ce séquençage s'est effectué à partir de prélèvements venant de 21 personnes d'origine ethnique différente et a coûté 2.7 milliards de dollars. En 2018, le même travail coûte 1000\$ pour une semaine de travail (Craig Venter et al., 2001).

3.1.2.1 Le principe du séquençage global

Dans cette technique, l'ADN à séquencer est fragmenté aléatoirement puis inséré par clonage, en fonction de la taille des fragments obtenus, dans des vecteurs de type **BAC** (« Bacterial Artificial Chromosome ») correspondants. Les clones obtenus sont séquencés en utilisant comme amorces les extrémités des BAC. Les fragments obtenus sont appelés « read ». Ces reads font entre 500 et 1000 paires de base. Du fait de cette taille, les reads appartenant au même clone sont rarement chevauchant. Les reads appartenant au même clone sont appelés « paire d'appariement ». Le logiciel d'assemblage récupère les reads, les rassemble grâce à leur partie chevauchante en séquences composites appelées contigs. Les contigs peuvent ensuite être assemblés entre eux grâce aux paires d'appariements pour créer des « scaffolds ». La distance entre les contigs dans le scaffold peut être inférée grâce aux connaissances *a priori* sur les BACs utilisés lors de l'étape de clonage. Les trous (gap) entre les scaffolds peuvent être comblés par PCR si ceux-ci sont courts, soit par une nouvelle étape de clonage puis séquençage, puis à nouveau assemblage.

3.1.2.1.1 Assemblage du génome

L'algorithme d'assemblage Celera est utilisé pour les premiers séquençages (124, 126–128). Cet algorithme comprenant six étapes majeures, prend en entrée les séquences auxquelles sont associées des scores de qualité en fonction de la confiance dans le fait que la base séquencée soit correcte ou que cela soit une base N, les informations des distances entre les vecteurs de séquençages et la taille des fragments (cf. Figure 15) (Myers et al., 2000).

3.1.2.1.1.1 Nettoyage

Le read est ensuite redimensionné pour obtenir un score de 98% de confiance. Les bouts de séquence correspondants au vecteur de clonage ou à l'adaptateur de séquençage sont supprimés. Puis pour finir, pour vérifier qu'il n'y ait pas eu contamination par de l'ADN étranger, les reads sont alignés au génome d'organismes étrangers, le vecteur de clonage ou de séquençage. Tout fragment s'alignant avec de l'ADN contaminant est supprimé.

On obtient ainsi des fragments de haute qualité.

3.1.2.1.1.2 Criblage

Chaque fragment est ensuite aligné aux régions connues comme étant répétitives. S'il y a correspondance, les reads sont soit annotés, soit mis de côté pour un criblage ultérieur plus strict. Les régions répétitives sont soit des régions microsatellites, de l'ADN ribosomal, des transposons, des histones ou encore de l'hétérochromatine en fonction de l'organisme séquencé.

Lors des premiers séquençages, les régions correspondants à des régions répétées n'étaient pas ou très peu séquencées. Les différentes versions des génomes de référence tentent de version en version d'améliorer cela.

3.1.2.1.1.3 *Chevauchement*

Tous les fragments sont comparés les uns aux autres, en empêchant la comparaison du même fragment sur lui-même, un chevauchement d'au minimum 6% impliquant au moins 40bp est recherché. Avec ces critères, soit le chevauchement est un « vrai » chevauchement, soit il est induit par la présence de régions répétées. Les fragments ayant un grand chevauchement sont des fragments induits par la présence de région répétée. La recherche de chevauchement s'appuie sur celle du BLAST sauf pour l'implémentation qui est adaptée aux calculs sur de gros volumes de données (comparaison de 32 Millions de paires de reads/seconde) et permet le traitement en parallèle. Cette configuration permet d'optimiser le calcul et de répartir le traitement sur plusieurs processeurs.

Les fragments chevauchants sont assemblés en un plus grand fragment appelé unitigs. Ceux qui ne contiennent pas de régions répétées sont notés « U-unitig » et les autres sont annotés « potentielles frontières ». Un U-unitig est considéré étant correctement assemblé et représentant une région unique d'ADN dès lors que le log-ratio de sa probabilité d'apparition sur celle d'apparition de unitig de régions répétées est supérieure à 10. Les « potentielles frontières » permettent d'étendre les unitigs. En effet, si un unitig chevauche deux autres unitigs, qui eux-mêmes se chevauchent jusqu'à un certain point, ces trois unitigs forment un U-unitig. Un U-unitig qui chevauche un unitig entièrement du côté répétitif peut être éliminé si aucun autre chevauchement n'existe. Cette étape permet de réduire efficacement le nombre de fragment en créant des U-unitigs de taille plus grande.

3.1.2.1.1.4 *Échafaudage*

À l'étape suivante, nous obtenons une population d'U-unitigs. Tous les U-unitigs contenant des paires d'appariements ou des extrémités de BACS communes sont assemblés en contig. De la même façon, les contigs sont assemblés en « scaffold » dont la distance entre les contigs est connue.

3.1.2.1.1.5 *Résolution récurrente : roches, pierres, galets*

Les différents gaps intra ou inter scaffolds sont divisés en trois catégories en fonction de la difficulté à les combler : les gaps de type **roche**, les gaps de type **Pierre** et les gaps de type **galet**. Le comblement des gaps de type roche se fait par la recherche de unitigs qui possèdent deux liens ou plus avec les contigs de part et d'autre du gap, où un seul lien différent est autorisé. Les gaps de type pierre sont comblés par des unitig qui présentent un seul lien avec un seul contig d'un seul côté du gap. Le comblement des gaps de type galet se fait par pavage avec des unitigs qui présentent la plus grande probabilité de chevauchement avec l'un ou l'autre des extrémités du gap.

3.1.2.1.1.6 *Consensus*

Ces étapes sont répétées et les reads sont alignées plusieurs fois jusqu'à obtenir la meilleure qualité de reconstruction de séquences en tenant compte de bien distinguer un polymorphisme ou une erreur de séquençage.

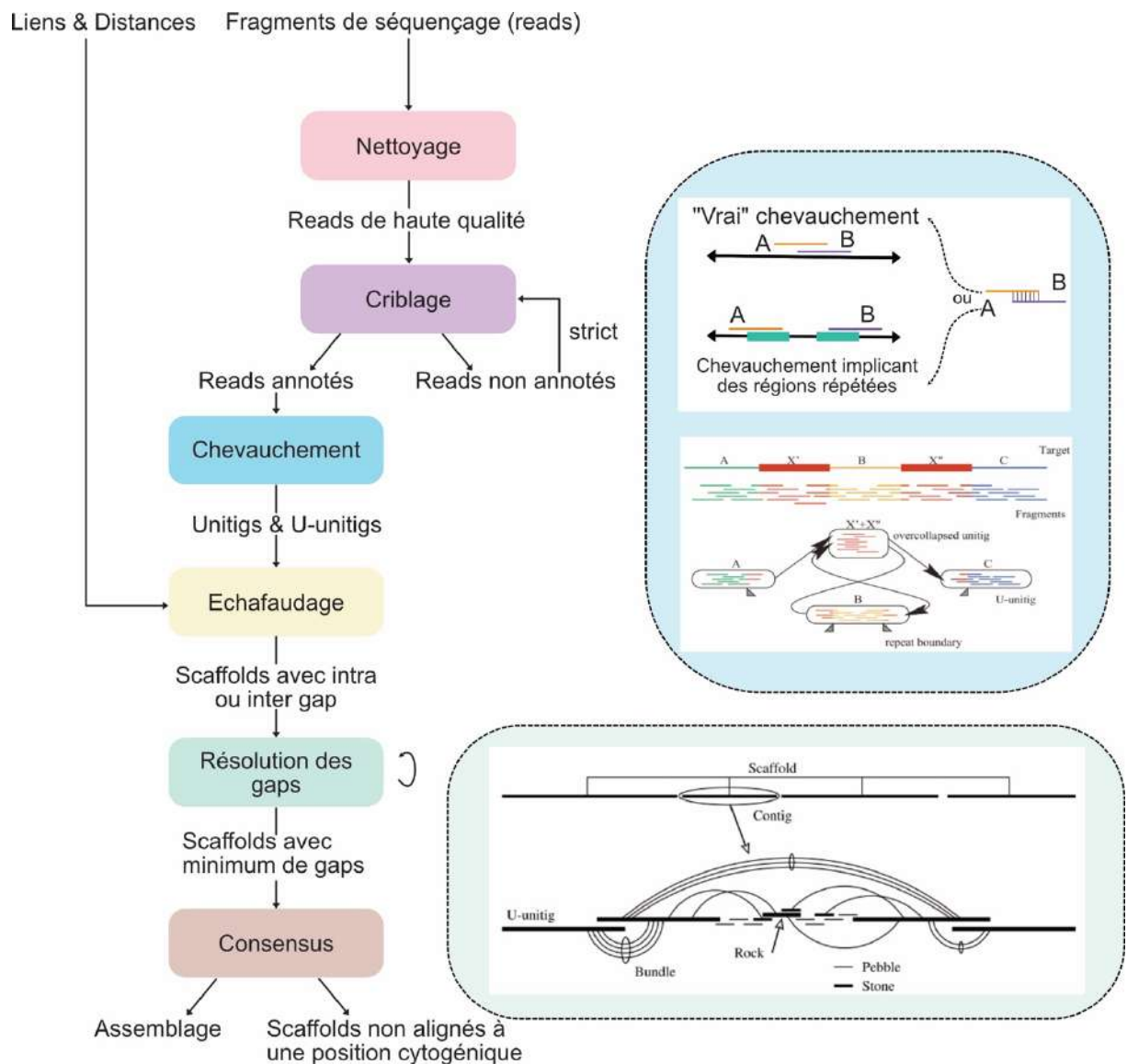


Figure 15 : Algorithme de Celera Adaptée de (Myers et al., 2000)

Toutes ces étapes sont encore utilisées de nos jours dans les séquenceurs de Nouvelle Génération avec quelques améliorations concernant le temps de calcul, la productivité, la réduction des coûts et de l'espace nécessaire pour stocker l'appareil de séquençage. Certaines des étapes sont utilisées lors de l'alignement des reads au génome de référence lors d'analyses génomiques pour la détermination de la fixation de protéines à l'ADN par exemple.

3.1.3 Séquenceurs Nouvelle Génération

Le Séquençage Nouvelle Génération (NGS), ou encore à haut-débit, se développe à partir de 2002. Ces séquenceurs ont remplacé les séquenceurs de Sanger dû au fait de leur faible coût et de leur rapidité. En 2021, on compte deux générations de séquenceurs NGS, dont je vais rapidement vous présenter les caractéristiques. La Seconde génération de séquenceur est une adaptation de la méthode de Sanger, les molécules d'ADN sont amplifiées en présence de ddNTP couplés à des fluorochromes de couleurs différentes en fonction de la base remplacée. Le décodage de la séquence nécessite une étape d'amplification par PCR qui peut engendrer des erreurs d'insertion. De plus, lors de l'utilisation de long reads, le

séquenceur perd en précision sur la longueur. La Troisième Génération de séquenceur décode la molécule d'ADN individuellement soit par la synthèse du brin complémentaire par une ADN polymérase, soit par détection des bases une à une en faisant passer la molécule d'ADN dans un nanopore ou bien en s'appuyant sur des techniques de microscopies. Cette génération a pour but de réduire les coûts et d'augmenter la vitesse d'obtention des résultats en s'affranchissant de l'étape d'amplification du signal (GLENN, 2011; Slatko et al., 2018). De ce fait la bio-informatique est fortement utilisée lors des études en biologie avec un bon dans les années 2000 avec l'essor du NGS. Une rapide recherche du terme « bioinformatics » dans PUBMED en témoigne avec plus de 40k publication en 2019 (cf. Figure 16).

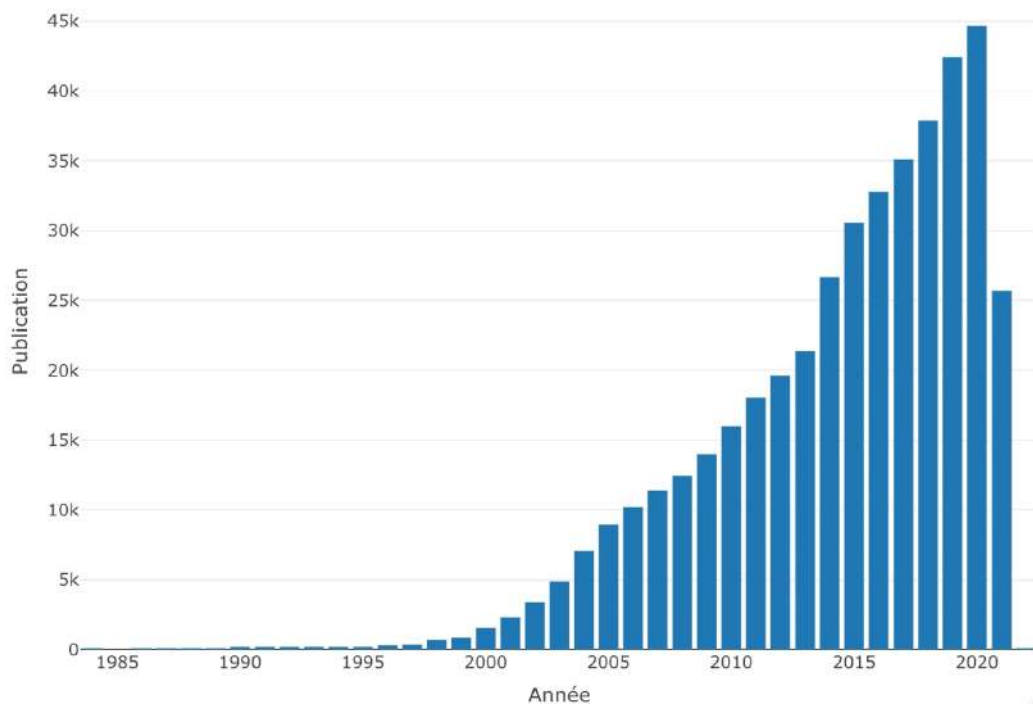


Figure 16 : Nombre de publications contenant le terme "bioinformatics"
source PUBMED

Je vous ai présenté ici le développement de la bio-informatique sous le prisme de l'analyse de séquence. Ce point de vue est utilisé autant en génomique qu'en épigénomique. Mais en réalité, cette discipline regroupe bien d'autres domaines dont la modélisation moléculaire, la biologie des systèmes, la métabolomique entre autres. Ce prisme est celui du travail que j'ai entrepris au sein du groupe de François Payre. Mon travail a consisté durant ces quatre ans à caractériser les formes de Shavenbaby en utilisant des approches de génomique pour déterminer son comportement de fixation et son effet sur la transcription. Je me suis aussi intéressée à l'environnement chromatinien autour de Shavenbaby. Ce travail a été réalisé chez de la Drosophile et je vais m'attacher maintenant à vous en présenter certaines de ses caractéristiques génomiques.

Faits marquants de la Bioinformatique

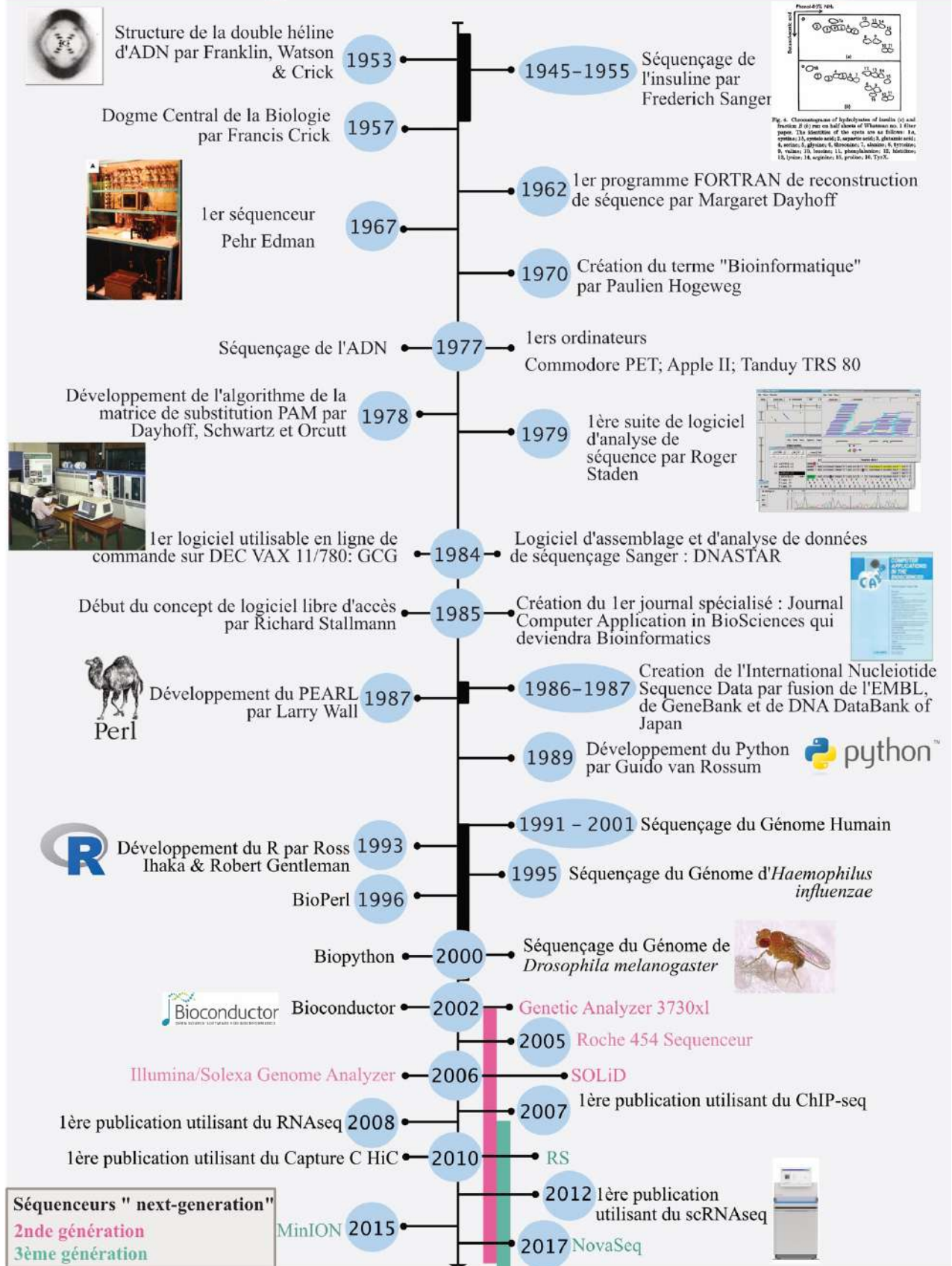


Figure 17 : Frise chronologique du développement de la bio-informatique

3.1.4 Le génome de *Drosophila melanogaster*

Comme je vous l'ai présenté précédemment, le séquençage du génome de *Drosophila melanogaster* a pris un an. Une première version a été publiée en 2000 et a permis de valider la méthode du WGS pour le séquençage des eucaryotes supérieurs. Lors de la première version du séquençage de ce génome, il est apparu que la majorité de l'hétérochromatine n'était pas insérée dans les BACs de séquençage. Il est important de savoir qu'un tiers du génome de la mouche du vinaigre est représenté par de l'hétérochromatine. Les versions suivantes du séquençage de ce génome s'attacheront donc à améliorer ce biais. De nos jours, la version 6 disponible depuis 2015 (131) est utilisée. Dans cette version, Hoskins et al ont pour objectif d'utiliser les avancées en termes de séquençage et de microscopie pour augmenter la précision de l'annotation du génome et en particulier des régions dites d'hétérochromatines. Des techniques d'hybridation *in situ* par fluorescence (FISH) dirigées contre les scaffolds mal annotés lors de la version 5 permettent ainsi de leur réattribuer une position cytologique (132). Trois types d'hétérochromatine sont alors définis : l'hétérochromatine centrale contenant des régions fortement répétées située au niveau des centrosomes, l'hétérochromatine périphérique désignant les régions autour des centromères riches en régions répétées comme les éléments transposables, et l'hétérochromatine télomérique correspondant aux régions subtélomériques composées de répétitions en tandem (132, 133).

Dans la version 6, le génome séquencé atteint la taille de 143 Mbp (cf. Tableau 18), correspondant à 5% de la taille du génome humain.

Table 1. Summary of the Release 6 sequence assembly

Chr. arm	Size (bp)	Size w/o N's	N50 ^a	Sized gaps	Estimated sum of gap sizes (bp)	Unsize gaps ^b
X	23,542,271	23,476,151	21,907,215	4	65,520	6
2L	23,513,712	23,513,512	21,485,538	0	0	2
2R	25,286,936	25,280,236	17,021,040	1	6000	7
3L	28,110,227	27,992,067	19,478,218	4	117,660	5
3R	32,079,331	32,054,759	27,905,053	9	22,772	18
4	1,348,131	1,331,131	1,200,662	1	17,000	0
Y	3,667,352	3,409,719	107,634	61	242,633	150
Subtotal	137,547,960	137,057,575	21,485,538	80	455,185	182
Auxiliary ^c	6,362,942	5,515,449	2927	256	657,793	1897
Total	143,910,902	142,573,024	21,485,538	336	1,112,978	2079

^aN50 is the contig length for which 50% of called base pairs in a chromosome arm assembly are contained in contigs this length or larger.

^bUnsize gaps are represented in the sequence files by 100N's.

^cDefined in Table 2.

Tableau 18 : Caractéristiques de l'assemblage du génome de *Drosophila melanogaster* issu de (Hoskins et al., 2015)

La prédiction des gènes lors du premier séquençage par Adams et al annote un peu plus de 17 000 gènes (Adams et al., 2000). Grâce à des études phylogénétiques entre 12 espèces différentes de drosophiles, il est apparu que \pm 13 000 gènes sont considérés comme étant codants, les autres sont encore à définir ou considérés comme non codants (Adams et al., 2000 ; Clark et al., 2007 ; Myers et al., 2000). Ces gènes ont une longueur d'en moyenne 1kb et sont divisés en deux catégories :

- **les gènes développementaux** nécessaires à la différenciation terminale
- **les gènes de « ménage »** codant plutôt pour des mécanismes ubiquitaires comme les ribosomes

Il est intéressant de noter que le génome de la drosophile est dense en gènes (cf. Figure 19) et que certains gènes sont emboîtés les uns dans les autres. Ces constatations sont

importantes lors de la réattribution d'un site de fixation d'une protéine à un gène potentiellement cible de cette protéine.

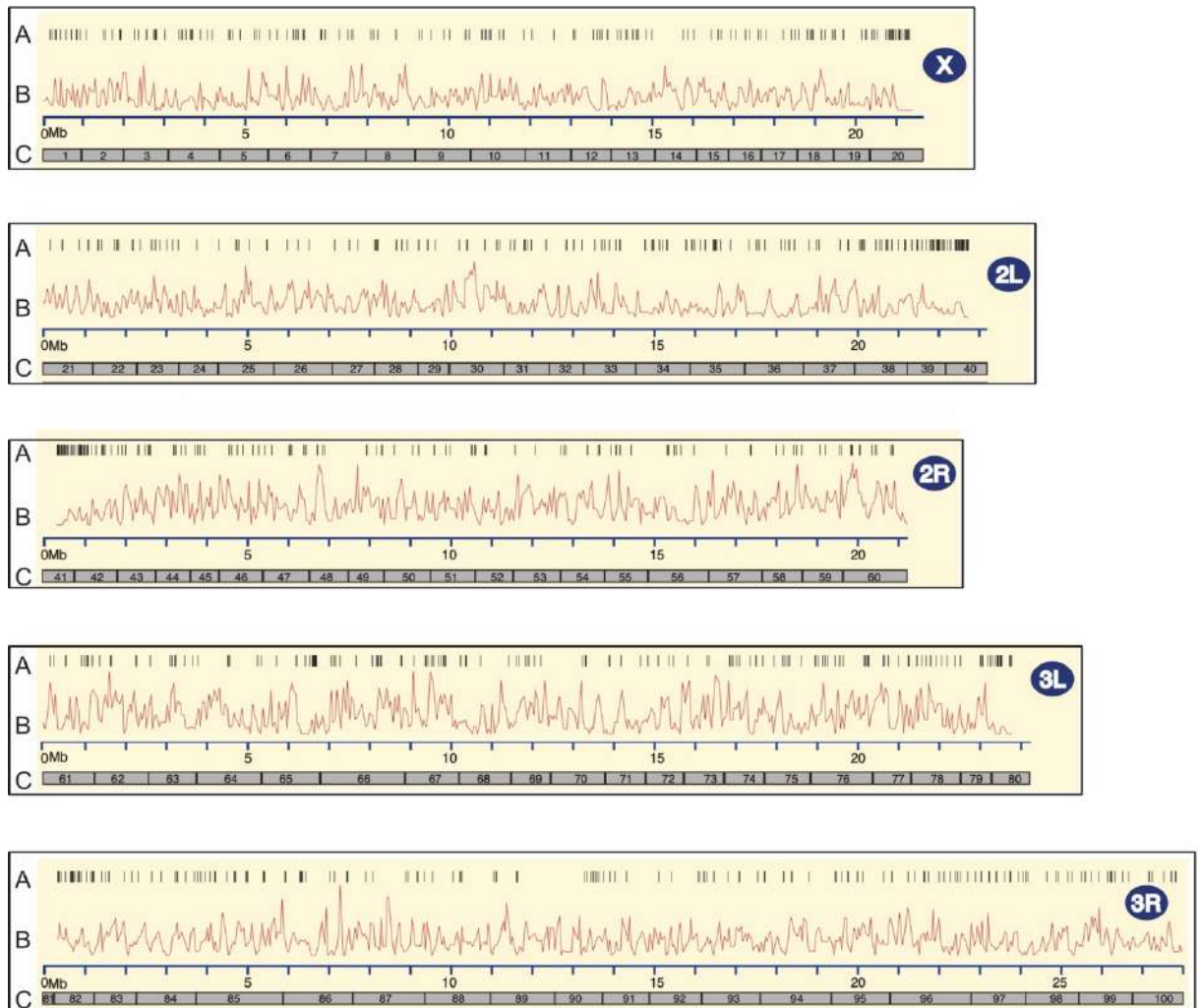


Figure 19 : Assemblage du génome de la *Drosophile*.
Chaque bras chromosomique est représenté avec (A) les éléments transposables (B) la densité en gènes (C) les sections du chromosomes polytènes. Adapté d' (Adams et al., 2000)

Grâce à des études de comparaison entre l'humain et la drosophile, il a été établi que 60% des gènes de la drosophile sont homologues aux gènes humains et que parmi ces gènes, 70% sont impliqués dans des pathologies humaines tels que les cancers ou les maladies mendéliennes (135, 136). Il a même été montré que l'insertion du gène homologue humain peut sauver le phénotype de la mouche mutante pour ce gène (Yamamoto et al., 2014).

Ces observations appuient donc le bien-fondé de l'utilisation de *Drosophila melanogaster* comme organisme modèle pour étudier ce type de pathologie via des approches génomiques.

Pour conclure cette partie, nous soulignerons que la mouche du vinaigre possède un génome compact de 142Mbp comprenant 17 000 gènes. Ces gènes sont fortement conservés entre les espèces et notamment chez l'Homme, qui présente des processus et gènes homologues, faisant ainsi de la Drosophile un bon organisme modèle.

3.2 Éléments de régulation de la transcription : les enhancers

Pour qu'un organisme se développe et se mette en place de façon correcte, l'expression des gènes qui contrôlent son développement (appelés gènes développementaux) doit être finement régulée autant temporellement que spatialement. L'expression spécifique de ces gènes dépend de l'action concertée des grandes voies de signalisation (Notch, Wnt, Hedgehog, Hippo...) ainsi que de régions cis-régulatrices appelées enhancers fixées par des facteurs de transcription et présentes généralement à proximité du gène régulé (Levine et al., 2005). Un enhancer peut être défini comme une courte séquence d'ADN permettant l'activation de l'expression génique en réponse à des signaux intrinsèques ou extrinsèques et indépendamment de son orientation. Nous verrons dans cette partie, comment ces enhancers ont été mis en évidence, quel est leur mode de fonctionnement et quelles sont les techniques pour les identifier.

3.2.1 Mise en évidence des régions cis-régulatrices

Pour donner suite aux travaux pionniers chez la levure *Saccharomyces cerevisiae* sur la régulation de la transcription, Moreau *et al* insèrent, dans des cellules simiesques CV1, une région répétée de 72pb du virus SV40 en amont du gène de la conalbumine du poulet (Moreau et al., 1981). En parallèle, Banerji et al. la même année, insèrent quant à eux cette même région en amont du gène de la β -globuline dans des cellules HeLa. Les deux groupes observent une augmentation de l'expression des gènes lorsque la séquence est insérée à une distance > 1kpb, aussi bien en 5' qu'en 3', du promoteur du gène cible. C'est ainsi que la notion de régions cis-régulatrices a vu le jour.

Chez la drosophile, la mise en évidence de régions cis-régulatrices s'est initialement effectuée à travers l'étude de la régulation des gènes de segmentation antéro- postérieur (*hairy, eve, ...*) (Nüsslein-Volhard et al., 1980; Akam, 1987; Ingham, 1988) par la technique dite d'« enhancer bashing » qui consiste à isoler, à muter et à tester l'activité d' enhancers potentiels grâce à des lignées rapportrices (voir ci-dessous). Ainsi pour les gènes *hairy* et *eve* il a été montré que différents enhancers sont nécessaires pour reproduire l'intégralité de leur patron d'expression (Riddihough et al., 1991; Fujioka et al., 1999) et ont permis de définir le registre spatial de ces régions cis-régulatrices.

3.2.2 Mécanismes de fonctionnement des enhancers

Chez les mammifères, la régulation de l'expression des gènes via des régions cis-régulatrices peut se faire jusqu'à 1Mbp de distance. Chez la Drosophile, bien qu'il existe quelques gènes pour lesquels les enhancers sont compris entre 70 à 100kb, *svb* en étant l'exemple canonique, la majorité des enhancers sont bien plus proches et sont situés au maximum à 10kb du gène cible (Furlong et al., 2018), voire situés au sein de la région génique, en particulier dans le 1^{er} intron ou dans les régions 3' ou 5'UTR (Arnold et al., 2013; Kvon et al., 2014). Différents modèles d'interaction promoteurs-enhancers ont été décrits au cours du temps afin de mieux appréhender comment la fixation des facteurs de transcription sur les

enhancers affecte l'activité de la machinerie transcriptionnelle et notamment de l'ARN Polymérase II (Furlong et al., 2018).

3.2.2.1 *Le pistage ou tracking*

Dans ce modèle de communication, l'ARN Polymérase II est fixée à l'enhancer en amont du promoteur du gène cible et commence son activité de synthèse d'ARN. Son activité d'élongation induit alors un rapprochement de la région cis-régulatrice fixée jusqu'au promoteur basal, créant ainsi une boucle (cf. Figure 20-A). Ce modèle est étayé par deux observables :

- La synthèse d'ARN spécifiques des enhancers, appelés eRNA. Plusieurs études montrent que lorsque la transcription de ces eRNAs est diminuée ou altérée par RNAi ou diminution de l'activité polymérase, le rapprochement enhancer-promoteur est également altéré (Arnold et al., 2020)
- L'ajout de régions insultrices entre l'enhancer et le promoteur bloque la formation de la boucle et donc l'activation du promoteur, comme si la progression de la Polymérase II le long de l'ADN avait été bloquée par ces insulateurs (Furlong et al., 2018).

3.2.2.2 *Le chaînage*

Dans ce cas, les facteurs de transcription se fixent à l'enhancer, s'oligodimérisent pour former un maillon de plusieurs facteurs de transcription et autres cofacteurs protéiques, allant de l'enhancer jusqu'au promoteur basal, permettant ainsi la régulation du gène cible (cf. Figure 20-B). Ce modèle s'appuie sur des observations concernant le facteur de transcription Chip interagissant avec Suppressor of Hairly wing, Su(Hw), chez la Drosophile pour réguler les gènes de la segmentation (149). Chez la souris, il a été observé que la protéine Ldb1, homologue murin de Chip, se fixe à la région cis-régulatrice, se dimérise et recrute d'autres facteurs de transcription de type GATA, GATA1 en particulier, jusqu'à former une chaîne pour réguler le promoteur de la β -globuline (Bulger et al., 1999).

3.2.2.3 *La formation de boucle ou looping*

Le modèle suivant, appelé « looping », s'appuie sur la formation de boucle d'ADN et comprend deux sous-catégories se différenciant en fonction de la distance comprise entre la région cis-régulatrice et le promoteur (cf. Figure 20-C ; D). La différence avec la formation de la boucle dans le modèle du tracking et le modèle du looping est la suivante : dans le tracking, la boucle est la résultante de synthèse d'ARN, alors que dans la looping, il faut qu'il y ait formation de la boucle pour qu'il y ait synthèse d'ARN (Bulger et al., 1999; Deng et al., 2012; Furlong et al., 2018). Dans la looping, la formation de boucle est donc un processus dynamique. Cependant, le tracking et le looping ont en commun d'impliquer des moteurs protéiques de type ATP tels que respectivement la Polymérase II et les complexes cohésine et/ou la condensine.

La première mise en évidence que la formation d'une boucle peut agir sur l'expression génique a été réalisée chez la souris en ciblant la protéine homodimère Ldb1 couplée à des

protéines de liaison LIM1 sur le promoteur de la β -globuline (P-ZF) et sur la région enhancer LCR (L-ZF). Cela crée une boucle rapprochant le promoteur à la région LCR et induit l'activation du promoteur de la β -globuline et cela en absence de facteur GATA1 (Deng et al., 2012; Schoenfelder et al., 2019).

La formation de boucle à courte distance est plutôt observée au sein de bactérie telle que *Escherichia coli* et également chez les phages *lambda*. A courte distance la formation de la boucle implique la fixation de facteur de transcription (TF) sur le promoteur et l'enhancer sur le même côté de la double-hélice d'ADN (cf. Figure 20-C) (Bulger et al., 1999; Furlong et al., 2018).

De nombreuses études montrent que le phénomène d'extrusion de boucle à longue distance semble être largement médié par les complexes cohésine via des motifs convergents CTCF, la condensine et YinYang (YY1). Outre les motifs CTCF, des cofacteurs peuvent être impliqués, comme le complexe Médiator ou encore des lncRNAs ((Schoenfelder et al., 2019).

Ghavi-Helm et al., en étudiant la conformation de l'ADN grâce à des techniques de 4C (Chromosome Conformation Capture Sequencing) dans des embryons à deux temps développementaux différents (3-4h après éclosion et 6-8h après éclosion) et dans deux types de tissus différents (mésoderme et dans tout l'embryon), a pu démontrer que, malgré le fait que le génome de la drosophile soit compact, les enhancers distants présentent une interaction avec en moyenne 10 régions génomiques distinctes et les promoteurs proximaux présentent une interaction plus forte avec les enhancers distants, dont 98% sont à plus de 10kb. Ces phénomènes de boucles impliquent des éléments spécifiques des régions cis-régulatrices actives comme de l'ARN Polymérase II. De plus, il semble que ces phénomènes de looping soient maintenus et évoluent au cours du développement (Ghavi-Helm et al., 2014).

Ce phénomène de looping à longue distance n'est pas uniquement médié par la cohésine et la présence de motif CTCF. Par exemple, chez la drosophile, il a été montré qu'en présence d'ecdysone, le complexe Polycomb en associant avec la forme courte du facteur de transcription Pipsqueak, forme une boucle rapprochant les enhancers au promoteur. Cette boucle permet ainsi l'activation des gènes cibles et enhancers jusqu'alors en pause (Lagha et al., 2012; Gutierrez-Perez et al., 2019). Ce résultat est surprenant car Polycomb est plus connu pour être associé avec des événements de répression se fixant sur des sites Polycomb Responsive Element (PRE) (Schuettengruber et al., 2007; Bantignies et al., 2011).

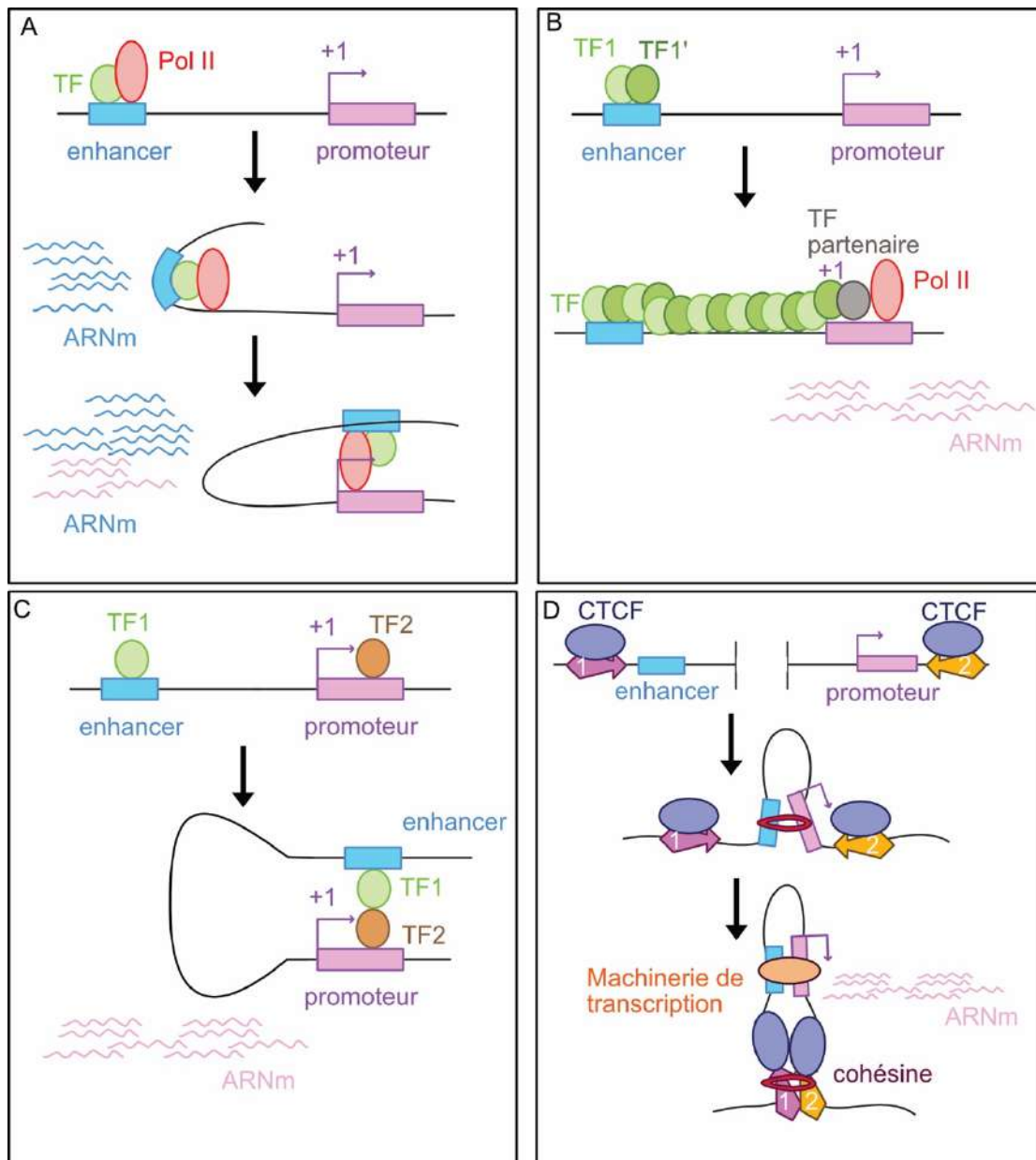


Figure 20 : Modèles de communication enhancer-promoteur

A) Le modèle du « tracking » la fixation du facteur de transcription (TF : vert) recrute l'ARN Polymérase II (Pol II : rouge) sur l'enhancer (bleu). La transcription s'initie et l'avancée de la Polymérase II (Pol II) le long du brin d'ADN provoque la formation d'une boucle et le rapprochement de l'enhancer au promoteur (violet) du gène cible et l'activation de ce gène. B) Modèle du chaînage : Le facteur de transcription se fixe à l'enhancer et s'oligodimérise pour former une chaîne jusqu'au promoteur du gène cible. Ensuite un TF partenaire est recruté sur le promoteur et la transcription est enclenchée par le recrutement de la Pol II. C et D) Le modèle de looping : C) Looping à courte-distance : deux facteurs de transcription distincts (vert et marron) sont recrutés l'un sur l'enhancer et l'autre sur le promoteur sur le même côté de l'hélice d'ADN. Les deux TFs se rapprochent créant une boucle et provoquant ainsi l'activation de la transcription. D) Looping à longue distance avec CTCF : recrutement d'un anneau cohésine (rouge) qui rapprochent les motifs CTCF fixés en créant une boucle. La cohésine arrête son effet de rapprochement dès qu'elle rencontre les motifs CTCF. L'enhancer et le promoteur, ainsi rapprochés, sont fixés par la machinerie de transcription et la transcription a lieu.

3.2.3 Les techniques d'identification de régions cis-régulatrices

L'identification des régions cis-régulatrices est essentielle pour la caractérisation et la compréhension de la régulation d'un gène. Pour cela, diverses techniques ont été mises au point au cours du temps. Certaines techniques sont plutôt basées sur une approche « gène candidat » alors que les autres sont plus globales.

3.2.3.1 Principe de gènes rapporteurs - promoter bashing

Historiquement l'identification des éléments de cis-régulation requis pour l'expression d'un gène a été réalisée en découpant des fragments génomiques de plusieurs kilobases situé en amont du gène d'intérêt. Ces fragments placés en amont d'un gène rapporteur constituent des lignées rapportrices qui peuvent reproduire tout ou partie du profil d'expression du gène endogène. Initialement, le gène LacZ codant pour la β -Galactosidase a été utilisé comme rapporteur. Depuis, des outils basés sur l'utilisation de marqueurs fluorescent type GFP se sont développés. Ils ont l'avantage de permettre de suivre le profil d'expression génique sur des individus vivants. L'introduction de mutations (délétions ou mutations ponctuelles) au sein de l'enhancer permet de tester les séquences d'importance au sein de ce dernier. Des délétions successives permettent - pas toujours - l'identification de fragment minimum soutenant l'activité du gène étudié. Cette méthodologie fréquemment utilisée *in vivo*, peut aussi servir dans des systèmes cellulaires. Dans ce cas, avec des plasmides contenant l'enhancer à tester intégrés ou non au génome est couplé à la GFP ou autre Luciférase. Lorsque l'enhancer est actif, la GFP est synthétisée et la cellule émet un signal de fluorescence qui peut être quantifié.

Si cette approche a permis l'identification de nombreuses régions cis-régulatrices, comme toute technique, elle a ses limitations. Due à la stabilité des rapporteurs utilisées (β -Gal ou molécules fluorescentes), il est notamment difficile de suivre la dynamique transcriptionnelle. Pour répondre à cela, des molécules fluorescentes plus instables ont été développées, les avancées les plus probantes sont venues du développement de la technique dite « MS2 tagging » (Pichon et al., 2018). Cette technique consiste à insérer 24 séquences reconnues par la protéine MS2 (MBS) du bactériophage entre le promoteur et l'unité de transcription du gène *yellow*. Lors de l'activation de la transcription, ces séquences MBS seront transcrites en formant des structures en tête d'épingle qui sont ensuite reconnues par la protéine d'enveloppe MCP (MCP : Ms2 Coat Protein) couplée à la GFP. L'ARN en cours de transcription est alors détectable en microscopie par visualisation de la GFP (cf. Figure 21).

La technique de SunTag permet elle, de suivre la dynamique de la traduction. Elle consiste à insérer des étiquettes (épitopes) « suntag » de façon répétée au sein de l'ARN à traduire. Lors de la traduction par les ribosomes, celles-ci seront insérées dans la protéine et reconnues par l'anticorps ScFv couplé à la GFP. La GFP indique donc les sites où la traduction est en cours de synthèse (Dufourt et al., 2018; Pichon et al., 2018) (cf. Figure 21).

La technique de lignées rapportrices est bien adaptée à l'étude de gènes d'intérêts mais quand est difficilement transposable à l'échelle génomique et des méthodes s'appuyant sur le séquençage ont été développées permettant d'identifier au global les régions cis-régulatrices présentes dans le génome.

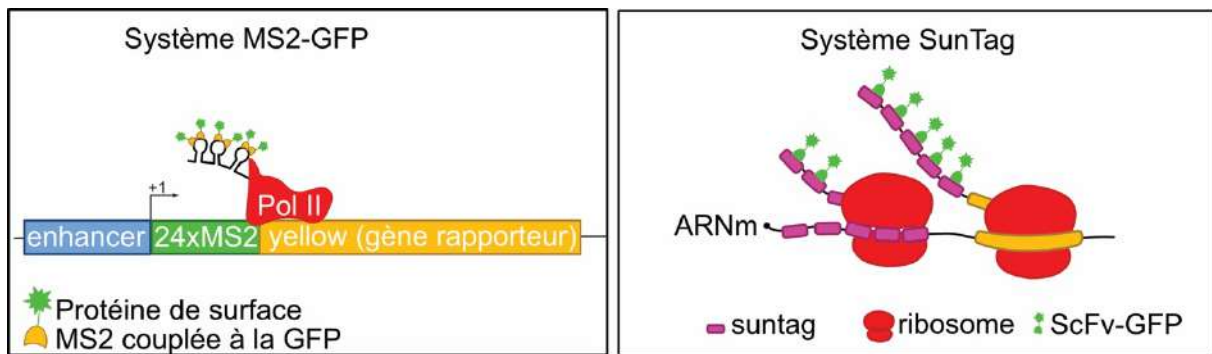


Figure 21 : Principes de caractérisation de régions cis-régulatrices in vivo par approche de microscopie. Le MS2-GFP permet de visualiser les ARNs en cours de synthèse et le SunTag permet de visualiser les ARNs en cours de traduction (Adapté du site <http://www.laqualab.com/research/>)

3.2.3.2 Approche globale : Principe du STARR-seq « self-transcribing active regulatory region sequencing »

Cette technique, mise au point dans le groupe d'Alexander Stark, se fonde sur le criblage systématique de régions aléatoirement fragmentées du génome et s'appuie sur la capacité des enhancers à activer localement leur propre transcription (Arnold et al., 2013; Muerdter et al., 2015).

L'ADN est récupéré, fragmenté, inséré dans un vecteur de clonage entre un promoteur basal et un site de polyadénylation. Il est ensuite transfecté au sein des cellules. Les ARNs messagers (ARNm) sont isolés grâce à leur queue polyA. Les ARNm sont « reverse » transcrits en ADN complémentaire pour être amplifiés par PCR et séquencés, puis des analyses bioinformatiques sont effectuées pour déterminer les régions pouvant être le plus probablement des régions enhancers (cf. Figure 22). Si la région insérée entre le promoteur minimal et la queue polyA est une région cis-régulatrice, celle-ci active elle-même le promoteur en amont et cette région sera retrouvée lors du séquençage. Si elle ne l'est pas, cette région ne sera pas transcrite et donc absente lors du séquençage (Muerdter et al., 2015).

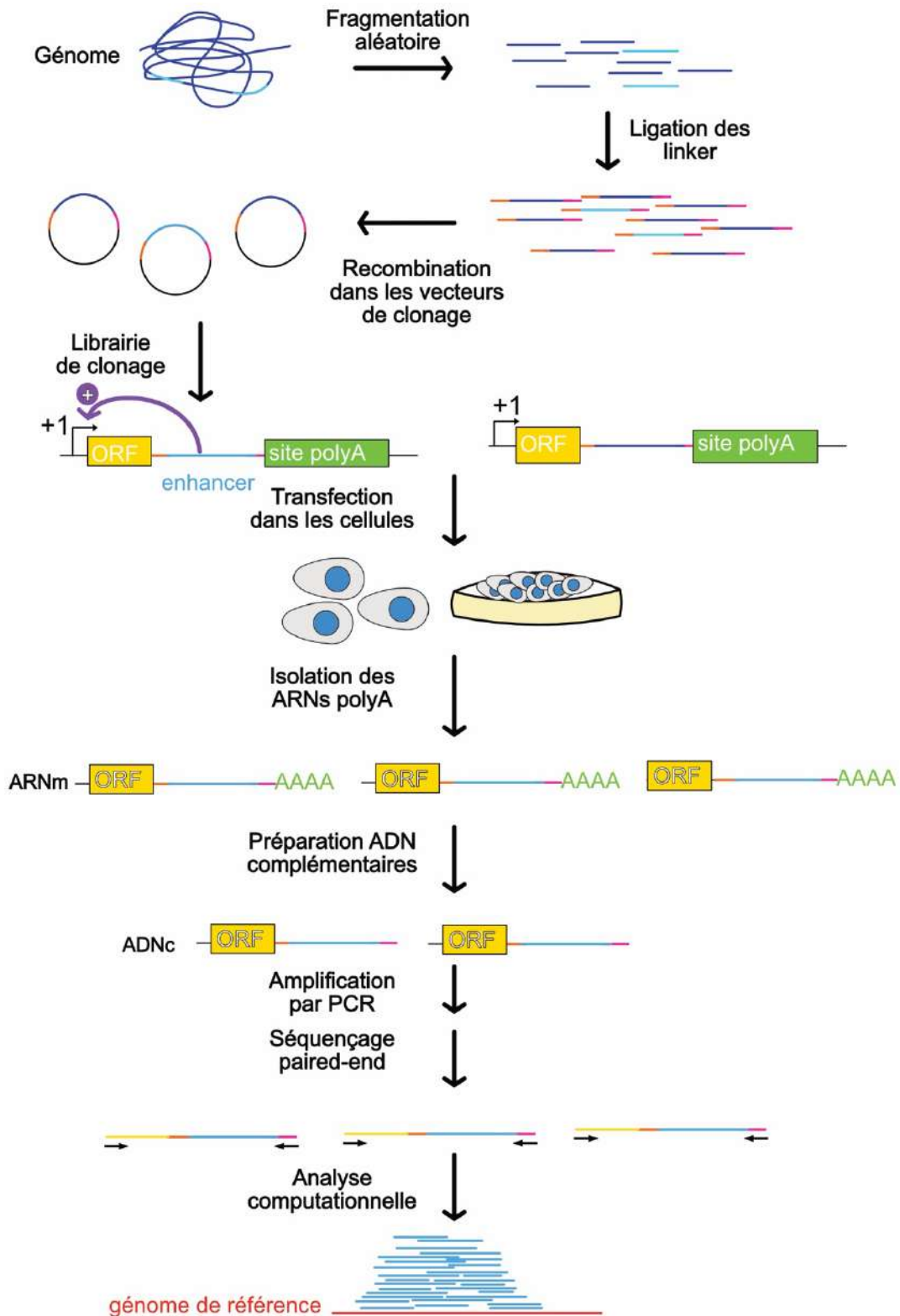


Figure 22 : Principe de la technique du STARR-seq

Grâce à cette technique, *Arnold et al.* (146) ont identifié 5 499 régions cis-régulatrices au sein de la drosophile dont la majorité (55.6%) est située au sein du premier intron et dans les régions intergéniques, mais également au niveau du TSS (Transcription Start Site) (cf. Figure 23-A). Outre la localisation des enhancers au sein du génome global, cette technique permet d'appréhender leur activité. En effet, plus celui-ci sera actif, plus le signal de

séquençage sera fort. Avec cette propriété, les auteurs ont pu voir que les enhancers les plus actifs sont ceux proches des gènes de ménage codant pour des enzymes ou des constituants du cytosquelette, mais aussi ceux proches de régulateurs développementaux, en particulier les facteurs de transcription (cf. Figure 23-B).

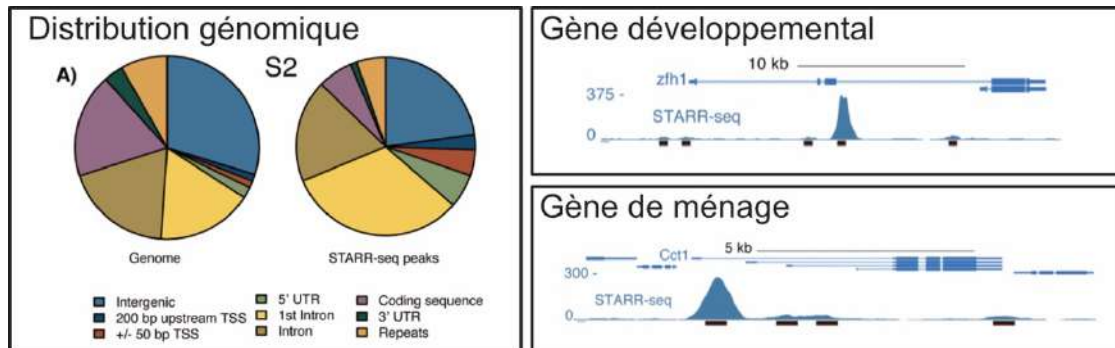


Figure 23 : Localisation des enhancers au sein de la Drosophile.

La majorité des régions cis-régulatrices chez la Drosophile sont situées dans le promoteur proximal et le premier intron A) Distribution des régions enhancers détectées en STARR-seq B) Exemples de pics de STARR-seq pour le gène développemental *zfh1* (en haut) et le gène de ménage *Cct1* (en bas). (Adapté d'Arnold et al., 2013)

Les régions cis-régulatrices, bien que pouvant activer elles-mêmes leur promoteur minimal associé – c'est sur cela que s'appuie la STARR-seq – nécessitent la coopération de plusieurs protéines et facteurs pour réguler finement l'activité de leur gène cible. Les approches que je viens de vous présenter, sauf le MS2 et le SunTag, ne tiennent pas compte de cet environnement protéique et chromatinien et peuvent donc amener à de fausses interprétations. Des techniques comme le CRISPR-Cas9 ou ciseaux génétiques tentent donc de modifier les séquences de cis-régulation de façon endogène pour pouvoir étudier l'activité cis-régulatrices dans son environnement.

3.2.3.3 Approche globale : Identification des sites de fixation des facteurs de transcription Principe du CHIP-seq

Les facteurs de transcription (TF) sont des protéines constituées d'un domaine de fixation à l'ADN et de domaine d'activation et/ou de répression d'un gène. Ces protéines voient leur séquence fortement conservée entre les espèces et servent de relais entre les signaux intra et extracellulaires (Levine et al., 2005). Chez la drosophile, il a été établi que celle-ci présente 700 facteurs de transcriptions dont la majorité sont des facteurs à doigts de zincs qui régulent l'expression des 13 000 gènes (Hammonds et al., 2013). Grâce à des approches *in vitro* (protection à la DNAase I, de retard sur gel) et le développement des matrices poids position (PMW), il a été montré que les facteurs de transcription, pour se fixer à l'ADN, utilisent des suites de bases nucléotidiques d'environ 6bp, appelées motif ou site de fixation, spécifiques à chaque TF (Small et al., 2020; Zeitlinger, 2020) Du fait de la présence de sites de fixation multiples (Zeitlinger, 2020) et de formation de complexes protéiques, il est possible que la région de fixation atteigne jusqu'à 20bp (Noyes et al., 2008; Crocker et al., 2015).

L'étude des facteurs de transcription a elle aussi bénéficié de l'essor de la bio-informatique et des méthodes de séquençage. En effet, il est maintenant usuel d'utiliser des méthodes impliquant du séquençage pour caractériser les régions de fixation dont les enhancers, d'un TF particulier et par extension, les gènes potentiellement régulés. Cette avancée a été permise grâce au développement des techniques de CHIP-seq (Chromatine ImmunoPrécipitation) et

de RNA-seq. La méthode de ChIP-seq s'appuie sur l'interaction physique entre les facteurs de transcription et l'ADN. Les protéines sont fixées grâce à du formaldéhyde à la chromatine, puis l'ADN est soniqué pour obtenir des fragments de maximum 250pb. Les protéines d'intérêt sont ciblées grâce à un anticorps couplé à des billes d'agarose. On procède ensuite à une immunoprécipitation par centrifugation. De ce fait l'ADN lié à la protéine d'intérêt, elle-même reconnue par l'anticorps, sédimente entraînée par le poids de la bille auquel est couplé l'anticorps. Les fragments d'ADN sont libérés de leur fixation aux protéines par un traitement à la chaleur et ils sont ensuite séquencés. Les régions de fixation de la protéine d'intérêt sont ainsi déterminées par analyse bio-informatique (cf. Figure 24). On peut ensuite, grâce à des approches d'intégration de données hétérogènes, caractériser l'environnement génique de fixation du facteur de transcription. Puis déterminer les gènes cibles directes en intégrant ces données de fixation avec les données de modification d'expression des gènes (RNA-seq), comme j'ai pu le faire pour Svb (se référer à la partie Résultats Chapitre 6). Il est également possible de comparer deux facteurs ou cofacteurs et savoir si ceux-ci présentent des régions de fixation communes. Ce sont des approches que j'ai mises en place pour comparer les sites de fixation de Yki et Svb lors de la caractérisation du rôle de Svb au sein des cellules souches rénales (cf. partie Résultat Chapitre 4).

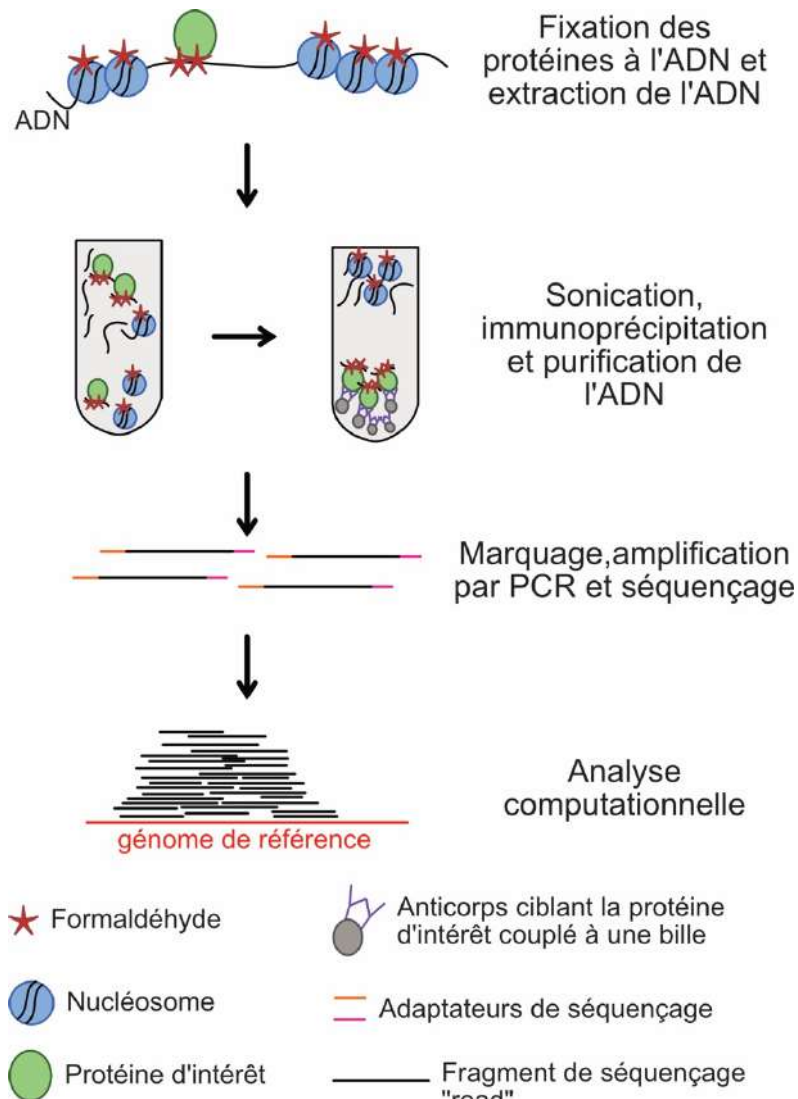


Figure 24 : Principe de la technique de ChIP-seq

En conclusion, les régions cis-régulatrices sont essentielles à la régulation précise de l'expression des gènes lors du développement. De plus, leur interaction avec leur promoteur cible semble impliquer des modifications locales de la chromatine. Précédemment nous avons vu que la caractérisation des régions cis-régulatrices s'est faite grâce à des approches « gènes candidats » mais aussi en utilisant des approches plus globales comme la ChIP-seq des facteurs de transcription, j'ai volontairement omis de vous parler d'une autre utilisation de la ChIP-seq pour caractériser les enhanceurs qui est de cibler les histones. Les histones sont des protéines impliquées dans le repliement de la molécule d'ADN pour acquérir sa structure en 3D. Avant de vous parler des histones et de leur modification post-traductionnelle, je vais vous présenter rapidement l'organisation spatiale de l'ADN passant d'une double-hélice seule à une forme compactée appelée chromosome.

3.3 Organisation de la chromatine et accessibilité des enhanceurs

Nous venons de voir le rôle essentiel du couple enhanceur-facteur de transcription dans la régulation de l'expression génique. Un des facteurs limitant dans le processus de fixation des facteurs de transcription (TF) est l'accessibilité des enhanceurs au sein de l'ADN qui doit être compacté afin de se loger dans le noyau des cellules. La compaction de l'ADN est possible grâce à l'action des histones qui en s'associant à l'ADN forment une structure, appelée **nucléosome**. Le nucléosome est un octamère de deux copies de chaque type d'histone (H2A, H2B, H3 et H4) sur lequel viennent s'enrouler 146 bases nucléotidiques (Rosa et al., 2013; Schwartz et al., 2017). En 1973, grâce à des approches de microscopie électronique, Woodcock *et al.* ainsi qu'*Olins et Olins* observent pour la première fois la structure en « collier de perle » aussi appelé structure 10nm. A ce niveau de compaction, les nucléosomes sont situés à environ 200pb les uns des autres (Olins et al., 1974; Woodcock et al., 1976). Vient ensuite une autre étape de compaction pour atteindre une fibre de 30nm appelée **chromatine**.

Traditionnellement la chromatine est divisée en deux catégories reflétant le niveau de compaction : l'**hétérochromatine** et l'**euchromatine**. L'hétérochromatine, constituée en majorité de séquences répétées, est fortement compactée alors que l'euchromatine est plus relaxée et semble être plus riche en gène (Rosa et al., 2013). Cependant, cette division manichéenne de la chromatine a été améliorée et remaniée en cinq catégories en fonction de la composition en protéines associées (complexe Trithorax, Polycomb, *etc.*) et en types de modification post-traductionnelle présentes sur la queue des histones (Filion et al., 2010). La localisation des enhanceurs dans une catégorie de chromatine ainsi que son association avec des modifications post-traductionnelles portées par les histones permet de prédire l'accessibilité de ces enhanceurs aux TF.

3.3.1 Les marques des histones :

Les histones du nucléosome sont modifiées post-traductionnellement par différentes enzymes (acétylase-désacétylase, méthylase-déméthylase) sur leur chaîne latérale (Kouzarides, 2007; Bannister et al., 2011; Boros, 2012; Jones et al., 2016 a). Ces enzymes sont appelées « writer » pour l'ajout d'un groupement fonctionnel ou « eraser » pour le retrait de ce même groupement. En 1964, *Alfrey et al.* observent pour la première fois des phénomènes

d'acétylation et de méthylation post-traductionnels des histones et mettent également en évidence que ces modifications qui ont lieu au sein du noyau, semblent être corrélées avec la synthèse de l'ARN (ALLFREY et al., 1964). Depuis, il a été mis en évidence que les histones peuvent être aussi modifiées par phosphorylation, sumoylation ou encore ubiquitinylation (Kouzarides, 2007; Bannister et al., 2011).

Les différentes modifications post-traductionnelles des histones sont souvent appelées **marques des histones** et ont été associées en fonction de leur enrichissement à des éléments géniques et des environnements chromatiniques spécifiques (168, 172, 174, 175). Par exemple l'acétylation de l'Histone 3 sur la lysine 27 (H3K27ac) est associée à des enhanceurs actifs présents dans l'euchromatine (Cubebñas-Potts et al., 2017).

3.3.1.1 L'acétylation et la désacétylation des histones

L'acétylation et son pendant inverse la désacétylation sont régulées par deux familles d'enzymes : les histone acétyl-transférases (HAT) et les histones désacétylases (HDAC). Ces modifications post-traductionnelles ont lieu au niveau de la chaîne latérale des lysines (K) des histones H3 et H4 (cf. Figure 25-A). Les HATs catalysent le transfert d'un groupe acétyle (Ac) grâce au recrutement du cofacteur CoA. Ainsi ajouté, ce groupe Ac modifie la charge de l'histone H3 ou H4 et diminue ainsi son interaction avec la molécule d'ADN. Il existe deux familles de HATs, les types A et les types B. Les HATs de type B sont plutôt cytoplasmiques et acétylent les histones libres. Les HATs de type A sont, quant à elles, en majorité situées dans le noyau et modifient les histones des nucléosomes et de par leur action, elles sont souvent considérées comme des coactivateurs de la transcription. Les HATs de type A sont séparées en trois sous-groupes dépendant de leur homologie de séquence et leur structure conformationnelle : les familles GNAT, MYST et CBP/p300. Chaque HAT modifie la conformation de résidu différent, par exemple CBP acétyle les lysines 14 et 18 de l'histone H3 alors que HAT1, de la famille des GNAT, acétyle les lysines 5 et 12 de l'histone H4 (Sterner et al., 2000; Kouzarides, 2007).

Les HDAC, séparées en quatre classes (de I à IV), en opposition aux HATs retirent le groupement Ac, restaurant ainsi la charge positive des histones. De ce fait, elles sont souvent considérées comme des répresseurs de la transcription (Bannister et al., 2011; Verza et al., 2020).

3.3.1.2 La méthylation et la déméthylation des histones

La méthylation a lieu sur des résidus arginines et lysines des histones H3. Les lysines (K) pouvant être mono- (me1), di- (me2) ou tri-méthylés (me3) et les arginines (R) pouvant être mono- et diméthylés. L'ajout d'un groupement méthyle ne modifie pas la charge des histones et est médié par des lysine méthyltransférases : **HKMT** (quand il s'agit de modifier une lysine) ou par des arginine méthyltransférases : **PRMT** (quand il s'agit de modifier une arginine) (cf. Figure 25-B). La première HKMT dont la caractérisation a été facilitée par la conservation de son cœur catalytique est SUV39H1. Elle modifie l'Histone 3 sur la Lysine 9 (H3K9) (Rea et al., 2000; Kouzarides, 2007). Les HKMT catalysent le transfert du groupement méthyle à partir de S-adosylméthionine (SAM) sur la lysine de l'histone (Kouzarides, 2007). Comme pour les HATs les HKMTs sont extrêmement spécifiques des résidus à modifier et niveau de méthylation à obtenir. Par exemple la protéine DIM5 ne peut triméthyliser que la

lysine 9 de l'histone H3 alors que SET7/9 ne peut que mono-méthyle la lysine 4 de l'histone H3 (Kouzarides, 2007; Hyun et al., 2017; Husmann et al., 2019)

Les **PRMT**, sont séparées en quatre types. Toutes les PRMT mono-méthylent l'arginine sur le nitrogène guanidinium terminal (N ω). Cette modification est annotée Rme1. Les différents types de PRMT se distinguent en fonction de la symétrie du groupement méthyle ajouté sur le N ω ou sur le composé interne de l'arginine Rme1 δ . Les type-I ajoutent des groupements asymétriques Rme2as sur le même composé N ω en plus du premier groupement méthyle Rme1. Les protéines modifiées sont annotées Rme1/Rme2as. Les PRMT de type-II génèrent des groupements symétriques Rme2s sur le composé terminal N ω . Les protéines sont cette fois-ci annotées Rme1/Rme2s. Les PRMT de type-III n'ajoutent que le groupement Rme1. Pour finir le type-IV méthyle un composé interne de l'arginine Rme1 δ (Lorton et al., 2019). Tout comme les HKMT, les PRMT transfèrent le groupement méthyle de SAM à l'arginine (Kouzarides, 2007).

Pendant longtemps il a été considéré que la méthylation des histones était irréversible. La découverte de la protéine LSD1, en 2004, la première lysine déméthylase, a permis de casser ce dogme. En effet, il a été montré que LSD1 est impliquée dans la déméthylation de la lysine 4 et de la lysine 9 de l'histone H3 (Shi et al., 2004). Il faudra cependant attendre 2006 pour que les premières arginine déméthylases soient identifiées. Cette identification s'est faite par l'observation de la déméthylation des arginine R2 de l'histone H3 et l'arginine R3 de l'histone H4 par la protéine JMJD6 en culture cellulaire et au sein de souris déficiente pour le gène JMJD6 (Chang et al., 2007; Kouzarides, 2007; Bannister et al., 2011; Lorton et al., 2019). Cette protéine appartient à la famille protéique Jumonji. Les protéines appartenant à la famille Jumonji sont des métaloenzymes possédant des domaines JmjC¹⁶ et JmjN¹⁷ déméthylant les histones par des mécanismes d'oxydation grâce à l'aide des cofacteurs Fe(II) et l'alpha-ketoglutarate (Klose et al., 2006).

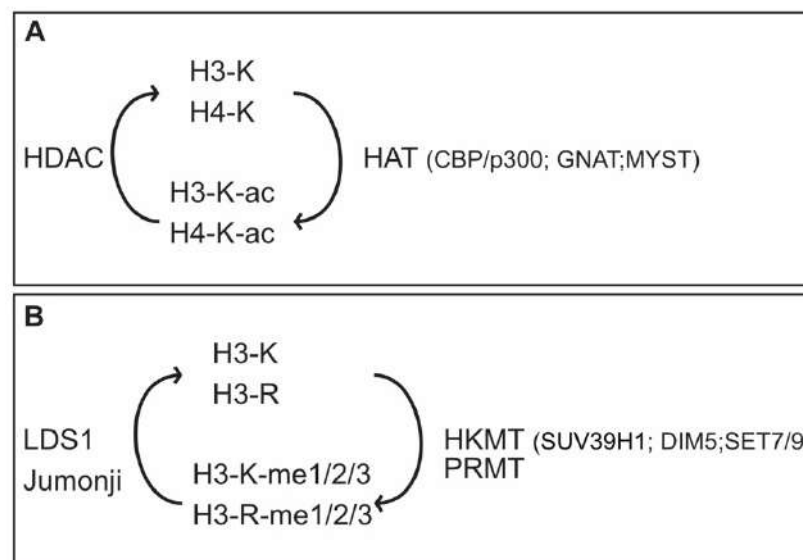


Figure 25 : Modifications des histones

- A) Acétylation / Désacétylation des histones H3 et H4 par les HAT et les HDAC sur la lysine K
 B) Méthylation des histone H3 sur la lysine (K) ou l'arginine (R) par respectivement les HKMT et le PRMT. Déméthylation par les protéines Jumonji et la LSD1.

¹⁶ <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR003347/>

¹⁷ <https://www.ebi.ac.uk/interpro/entry/InterPro/IPR003349/>

Les diverses modifications post-traductionnelles (ou marques d'histones) qui peuvent être perçues comme un code, sont souvent associées à une fonction au sein de la cellule : une région transcriptionnellement active, un enhancer plutôt actif ou réprimé *etc.* L'analyse de ces marques d'histones et leurs enzymes associées est appelée **épigénétique**.

3.3.2 L'organisation de la chromatine chez la *Drosophile*

La *Drosophile*, et en particulier ses chromosomes polythènes, est un bon modèle pour étudier le code des histones et appréhender la fonction d'une marque spécifique. La première observation reliant des marques d'histones à une possible fonction s'est faite pour l'histone H4 acétylé sur la lysine 16 (H4K16ac). Par exemple, *Turner et al.* ont observé un enrichissement de la marque H4K16ac sur le chromosome X et cela spécifiquement dans les noyaux des mouches mâles. Ils ont établi un lien entre la présence de cette marque et la compensation de dose du chromosome X qui se produit chez les mâles de la *Drosophile* (Turner et al., 1992). De part son statut d'organisme modèle et de la relative facilité pour générer des données de CHIP-seq, la *Drosophile* a été ajoutée comme source d'information au sein du projet ModENCODE dont l'objectif est d'identifier dans le génome de *Caenorhabditis elegans* et *Drosophila melanogaster* les éléments fonctionnels en se basant sur les séquences.

"The modENCODE Project will try to identify all of the sequence-based functional elements in the Caenorhabditis elegans and Drosophila melanogaster genomes."

De plus, il faut savoir que la *Drosophile* présente un seul locus codant pour les différentes histones, le complexe HisC, situé sur le chromosome 2 et composé d'environ 100 régions répétées en tandem dépendantes de la réplication. Cela a permis au groupe de Mathias Mannervik de développer une technique de remplacement des histones *in vivo*. Cette technique consiste à muter spécifiquement une des régions de l'histone d'intérêt grâce au système de recombinaison médié par la Flipase couramment utilisé en génétique chez la *Drosophile* (Regadas et al., 2021). Cette technique permet de modifier facilement les différentes marques d'histones au sein de la *Drosophile*.

Grâce à la facilité d'accessibilité des données, les chercheurs intègrent et comparent les diverses données issues de CHIP-seq pour différentes marques d'histones mais aussi de protéines associées à la chromatine pour définir des signatures permettant de prédire les éléments géniques comme les promoteurs, les enhancers, les corps du gène mais aussi l'état de compaction de la chromatine (Zang et al., 2009; Kharchenko et al., 2011 a; Cubeñas-Potts et al., 2017; Jiang et al., 2018 a).

3.3.2.1 Les couleurs de la chromatine

En 2010, *Filion et al.* entreprennent de caractériser les différents états de la chromatine. Leur approche s'appuie sur l'idée que la distinction historique et dichotomique de la chromatine en hétérochromatine et euchromatine n'explique pas les observables obtenus par les approches de séquençage (Filion et al., 2010; Filion, 2011). Leur stratégie est donc d'utiliser les profils de fixation de 56 protéines associées à la chromatine ainsi que ceux de quatre modifications post-traductionnelles des histones différentes : la lysine 4 de l'histone

H4 diméthylée (H3K4me2), la lysine 9 de l'histone H3 diméthylée (H3K9me2), la triméthylation de la lysine 27 de l'histone H3 (H3K27me3) ainsi que la triméthylation de la lysine 79 de l'histone H3 (H3K79me3) dans les cellules KC167 de drosophile en culture. Pour obtenir ces profils de fixation, des expériences de DamID (« DNA adenine methyltransferase identification ») ont été réalisées.

Le DamID consiste à fusionner la protéine d'intérêt à l'enzyme ADN adénine méthyltransférase, appelée Dam, qui méthyle l'adénine. Cette méthylation est absente des génomes eucaryotes et permet donc d'identifier les sites de fixation de la protéine d'intérêt (cf. Figure 26). Le DamID bien que moins résolutive que le ChIP-seq permet d'obtenir plus rapidement les sites de fixation de la protéine cible car cela ne nécessite pas de générer l'anticorps spécifique de cette protéine (Filion et al., 2010; Delandre et al., 2019).

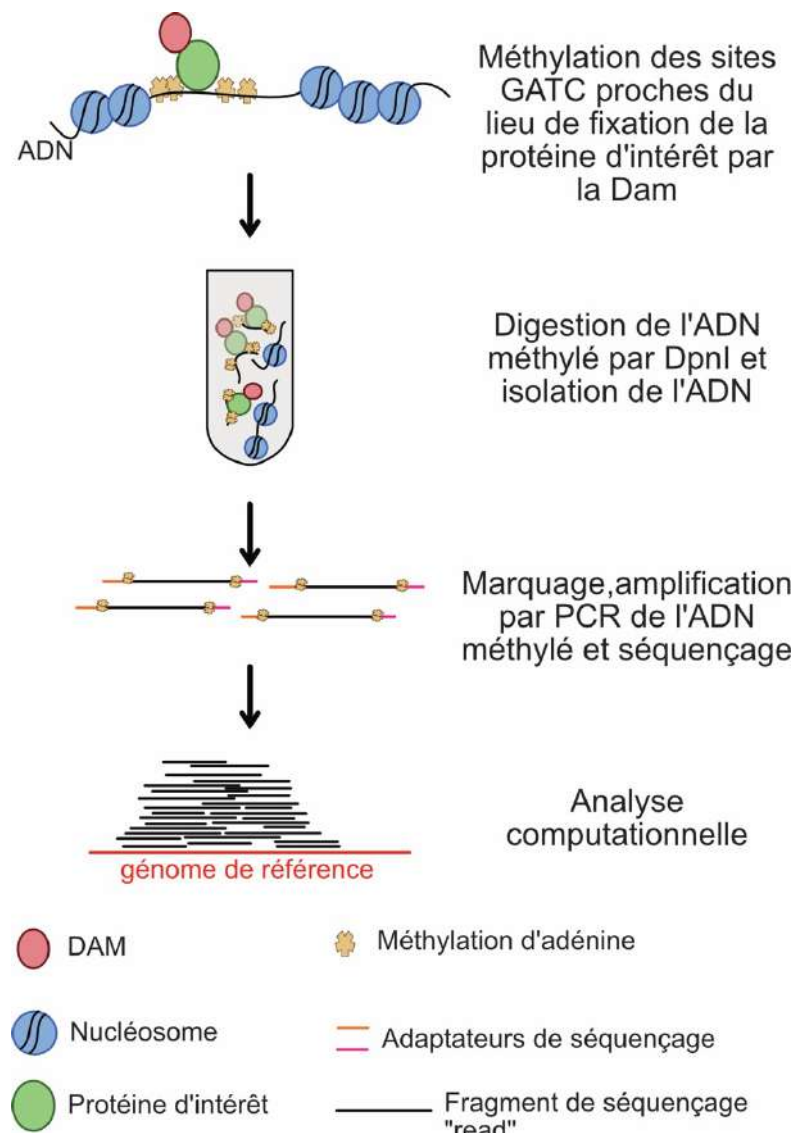


Figure 26: Principe du DamID

Avec les profils de fixations des 56 protéines associées à la chromatine ainsi obtenus par DamID, des analyses en composante principale (ACP) (cf. Figure 29-A) et d'inférence par modélisation de Markov caché (HMM) ont été réalisées (Filion et al., 2010). Étonnamment, malgré la foison de données (8428 domaines définis) et donc les possibilités théoriques de regroupement en découplant, seuls cinq groupes sont nécessaires pour expliquer le jeu de

données. En d'autres termes, pour un site donné, celui-ci peut appartenir à un des cinq groupes définis par une combinaison de différentes protéines. Ces cinq groupes constituent les cinq couleurs de la chromatine : **bleue**, **verte**, **rouge**, **jaune** et **noire** (Filion et al., 2010). Pour anecdote, Guillaume Filion explique le choix d'identifier les groupes par un code couleur comme trouvant son origine dans l'étymologie grecque du mot « chromatine » : khrōma signifiant couleur (Filion, 2011). Ces cinq groupes ont été obtenus pour 8428 domaines faisant en moyenne environ 1-52kb mais pouvant atteindre jusqu'à 737kb pour les plus larges, indiquant ainsi que le génome de drosophile est généralement organisé en de larges régions (Filion et al., 2010). Il est important de retenir que c'est la **combinaison de plusieurs protéines** qui définit chaque groupe et non pas la présence d'une seule protéine.

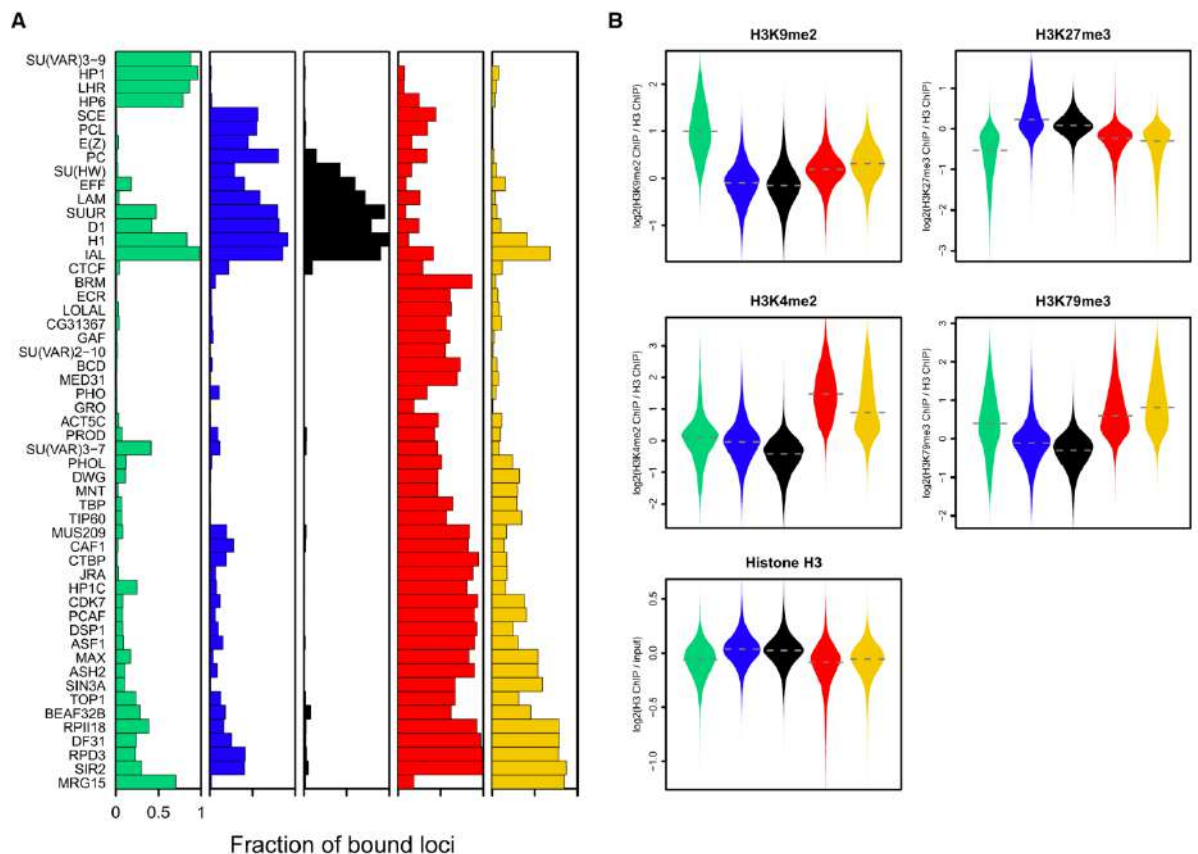


Figure 27 : Composition des cinq types de chromatine en protéines associées à la chromatine et en histones. A) Enrichissement (fraction de sites fixés) des différentes couleurs de chromatine en protéines associées à la chromatine. B) Enrichissement en H3K9me2 ; H3K27me3 ; H3K4me2 ; H3K79me3 ($\log_2(\text{signal}/\text{input})$) dans les différentes chromatines (Filion et al., 2010).

3.3.2.1.1 La chromatine dite « non permissive »

Parmi les cinq groupes identifiés, trois (**vert**, **bleu** et **noir**) appartiennent à la chromatine « non permissive » et deux (**rouge** et **jaune**) à la chromatine « permissive ».

3.3.2.1.1.1 La chromatine noire

Ce type de chromatine couvre 48% du génome et contient 4165 gènes soit 32% des gènes (cf. Figure 29-B). La chromatine noire présente toujours une forte signature en quatre protéines spécifiques : H1, D1, Suppressor of Underreplication (SUUR) et IAL/auroraB. Elle présente de façon moins systématique un enrichissement en Lamine, en insulateur Suppresseur de Hairy wing Su(Hw) et en E2 ubiquitine ligase Effete (cf. Figure 27-A) L'histone

H1 permet de stabiliser les nucléosomes (Hergeth et al., 2015), la protéine D1 est essentielle pour le silencing des gènes lors de la différenciation des gamètes (Aulner et al., 2002), SUUR est impliquée dans l'assemblage de l'hétérochromatine¹⁸, IAL/auroraB est impliquée dans la condensation chromosomique en particulier lors de la mitose¹⁹. Cependant ce type de chromatine ne présente pas d'enrichissement particulier en une marque d'histone (cf. Figure 27-B).

Grâce à des expériences de RNA-seq, il a été montré que la majorité des gènes présents dans cette région n'est pas exprimée ou alors très faiblement. Ces gènes qui sont dits **silencieux** (Filion et al., 2010) contiennent des gènes clés du développement neuronal. Grâce à des expériences de DamID dans des cellules souches neuronales larvaires, il a été montré que ces gènes passent de la chromatine noire à un état permissif lors de la différenciation neurale (194). Cette étude démontre si besoin le caractère réversible de l'état chromatinié et confirme le rôle actif de la chromatine noire dans la répression/« silencing » de la transcription lors du développement (cf. Figure 31).

3.3.2.1.1.2 *La chromatine verte*

La chromatine verte est la chromatine correspondant à la chromatine associée aux protéines en Su(Var)3-9, HP1 et ses partenaires LHR, HP6 (cf. Figure 27-A). Su(Var)3-9 est une HKMT dont l'activité est de méthyliser la lysine 9 de l'histone H3 (Bannister et al., 2001). HP1 et ses partenaires LHR et HP6 sont impliqués dans le « silencing » des régions centromériques (196). Cette chromatine recouvre très peu du génome et ne contient que très peu de gènes et leur expression est faible (cf. Figure 29). Il a été observé que la chromatine verte est essentielle pour le maintien de la différenciation des cellules neurales via la mise sous silence des gènes de dédifférenciation (191, 194).

Les histones H3 dont les lysines sont di ou triméthylées (H3K9me2/me3) sont fortement enrichies dans cette chromatine. Ces marques d'histones sont fortement associées aux régions péri-centriques ainsi qu'au chromosome 4 (cf. Figure 27-B). Ces modifications d'histones sont associées au phénomène de « silencing » des gènes (Bannister et al., 2001; Ebert et al., 2006).

La chromatine verte est donc spécialisée dans la mise sous silence des gènes et est spécifique des régions péri-centriques (cf. Figure 31).

3.3.2.1.1.3 *La chromatine bleue*

La chromatine bleue est enrichie en protéines du complexe protéique Polycomb (PcG) (cf. Figure 27-A). Ce complexe protéique est fortement connu pour être impliqué dans la régulation des gènes via des modifications post-traductionnelles des histones dans des processus majeurs comme l'inactivation du chromosome X chez le mâle, la plasticité et le renouvellement des cellules souches (Schuettengruber et al., 2007; Ogiyama et al., 2018; Tauc et al., 2021). Chez la Drosophile, ce complexe protéique est composé de cinq sous-complexes :

- **Complexe Répressif Polycomb 1 (PRC1)** composé de Polycomb (Pc), Polyhomeotic (Ph), Posterior Sex Comb (Psc), SexComb Extra (Sce/dRing).
- **Complexe Répressif Polycomb 2 (PRC2)** constitué de Enhancer de zeste (E(z)), ExtrasexCombs (Esc), suppresseur de zeste 12 (Su(z)12), de Nurf 55 et Polycomb-like (Pcl).

¹⁸ (<https://flybase.org/reports/FBgn0025355>)

¹⁹ (<https://flybase.org/reports/FBgn0024227>)

- **Complexe Répressif Pho (PhoRC)** composé de Pleiohomeotic (Pho) et dSfmbt (gène lié à Sexcomb of midleg contenant quatre domaines mbt).
- **Complexe de facteurs associés à dRing (dRAF)** comprenant Psc, dRing ainsi que la méthylase dKmd2.
- **Deubiquitinase Repressive Polycomb (PR-DUB)** formé de Calypso deubiquitinase et Additional sex combs (Asx) (cf. Figure 28)

La fixation du complexe PcG est séquentielle. Tout d'abord il y a recrutement du complexe PRC2 et de PhoRC sur des motifs PRE (« Polycomb Responsive Element »). Une fois fixée la sous-unité E(z) tri-méthyle la lysine 27 de l'histone H3. La modification post-traductionnelle H3K27me3 est formée. Cette méthylation permet le recrutement du complexe PRC1 via Pc. La sous-unité Sce/dRing va à son tour mono-ubiquitinyler la lysine 119 de l'histone H2A (H2AK119). La dKmd2 vient déméthyle la lysine 36 de l'histone H3 (H3K36me3). Les trois modifications post-traductionnelles médiées par le PcG sont considérées comme des modifications répressives. Calypso, suite au recrutement de PR-DUB sur les sites PRE déubiquitine la H2AK119, ce qui entraîne la répression génique. Il existe donc une balance entre l'action de dRing et de Calypso au sein du PcG (Schuettengruber et al., 2007; Bantignies et al., 2011).

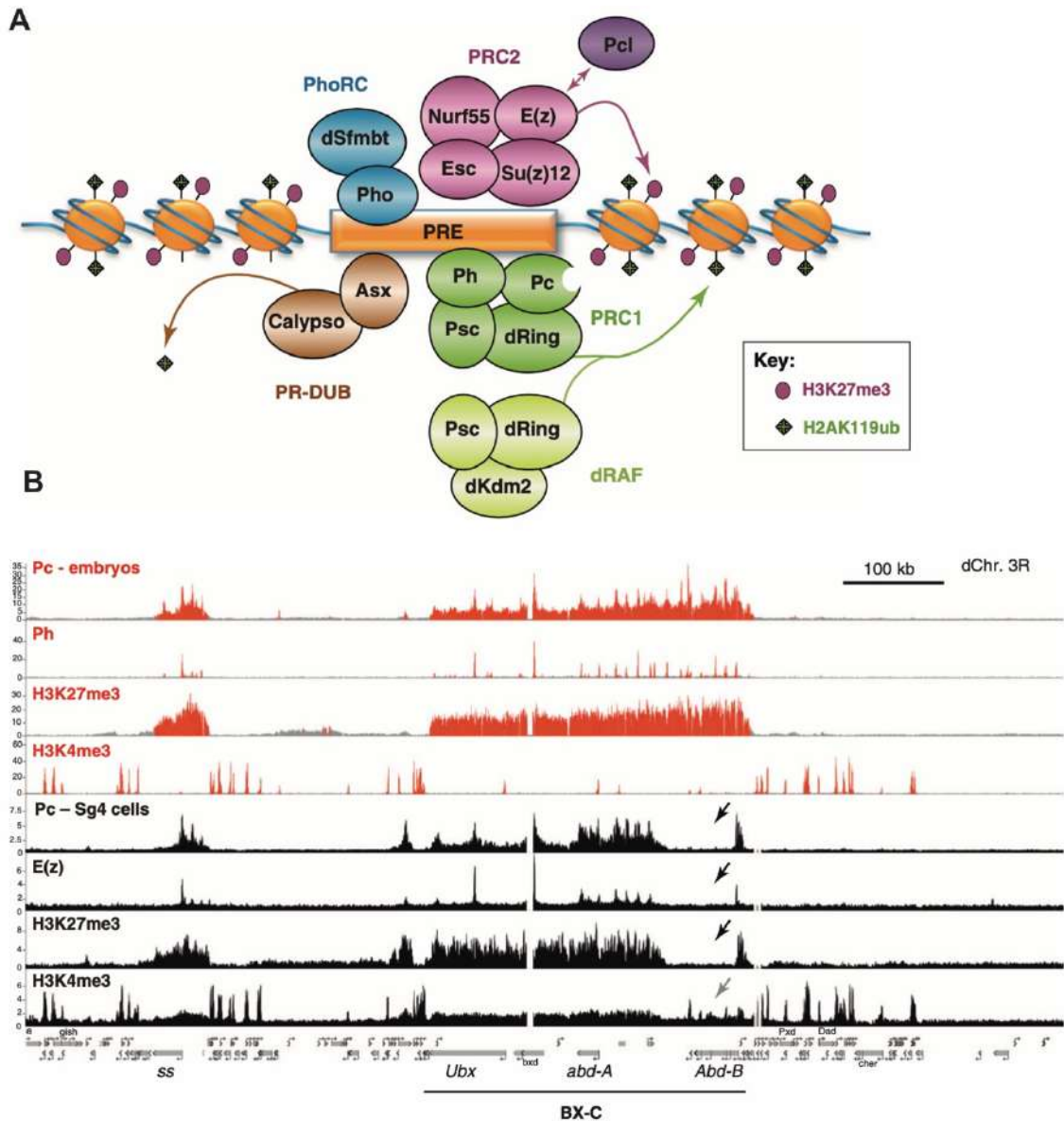


Figure 28 : Les protéines du complexe protéique Polycomb.

A) Le complexe Polycomb est constitué de cinq groupements protéiques (PRC1 ; PRC2 ; PR-DUB ; dRAF et PhoRC), sa liaison est séquentielle et se fait sur des séquences d'ADN spécifiques : les Polycomb Responsive Elements (PRE). B) Profils de fixation de Polycomb (Pc), de Polyhoméotic (Ph), E(z) et des histones H3K27me3 et H3K4me3 dans les embryons et les cellules Sg4 sur le locus du complexe bithorax. On peut observer une fixation de PRC1 et PRC2 qui correspond à la fixation de H3K27me3 (Bantignies et al., 2011).

Le complexe PcG est fortement décrit comme étant associé à la répression des gènes développementaux. Cette répression peut se faire par différents moyens comme la compaction de la chromatine par formation de boucle (Entrevan et al., 2016; Ogiyama et al., 2018) mais aussi par la répression de l'initiation et de l'élongation de transcription (Schuettengruber et al., 2007; Bantignies et al., 2011).

La chromatine bleue enrichie en H3K27me3, contient un grand nombre de gènes développementaux réprimés (cf. Figure 27-B ; C), en accord avec l'action connue du complexe Polycomb (cf. Figure 31).

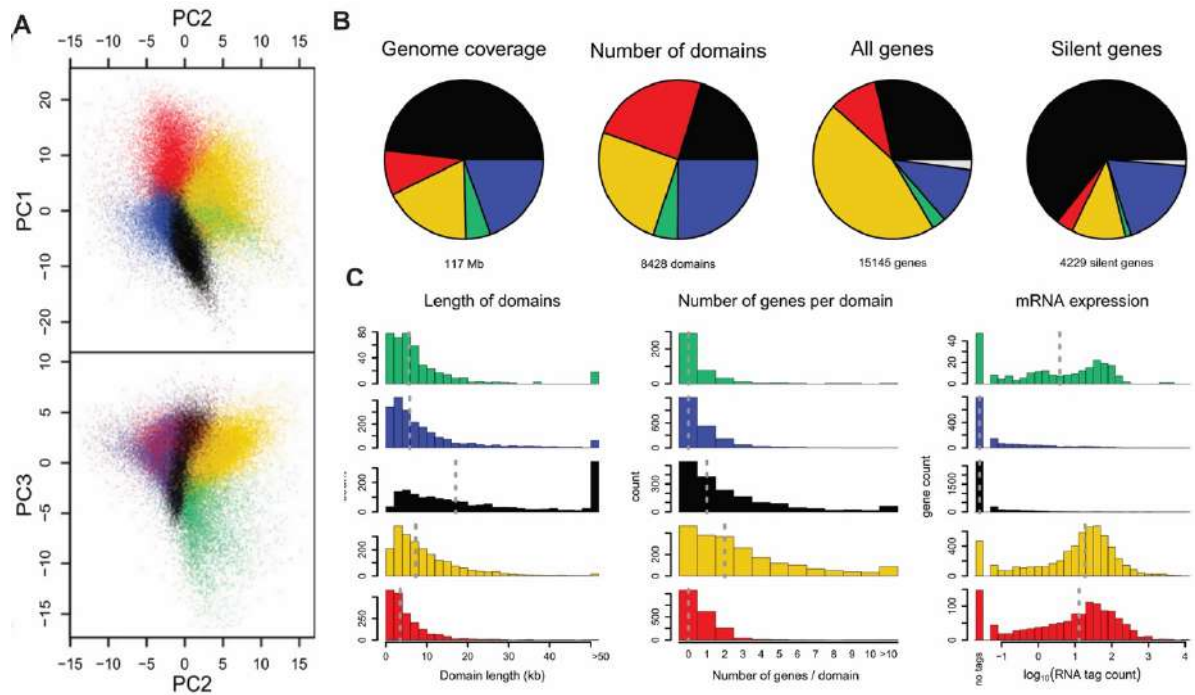


Figure 29 : Les différentes couleurs de la chromatine et sa composition.

A) Analyse en Composante Principale des différents sites de fixation des 56 protéines associées à la chromatine. Cinq groupes sont suffisant pour expliquer les données B) Composition en régions recouvertes, domaines, gènes et gènes réprimés pour les différentes chromatines C) Longueur, nombre de gènes et expression génique des différents types de chromatines. (Filion et al., 2010)

3.3.2.1.2 La chromatine « permissive »

La chromatine « permissive » est divisée en deux catégories de chromatine : la **rouge** et la **jaune**. Ces deux catégories contiennent les gènes exprimés et sont enrichies en ARN Polymérase II, H3K4me2 et H3K79me3 (cf. Figure 27). Ces deux types de chromatines sont très similaires et partagent de nombreuses protéines communes telles que les HDAC RPD3 et SIR2, ainsi que la partenaire de RPD3, SIN3A. On retrouve également DF31, décrite comme pouvant décondenser la chromatine *in vivo* et permettant de maintenir accessible l'ADN (Schubert et al., 2012), mais aussi Ash2, méthyltransférase de la H3K4, membre de la famille Trithorax (Papoulas et al., 1998). Pour finir Max facteur de transcription connu pour s'oligodimériser avec Myc pour activer la transcription (Gallant, Shio, Cheng, Parkhurst, & Eisenman, 1996). La chromatine jaune et rouge se distinguent l'une de l'autre par leur vitesse de répllication, la chromatine rouge se répliquant plus rapidement de la jaune (cf. Figure 30-A). Ces deux types de chromatine se différencient aussi en fonction de l'enrichissement en MGR15 et H3K36me3 (cf. Figure 30-B).

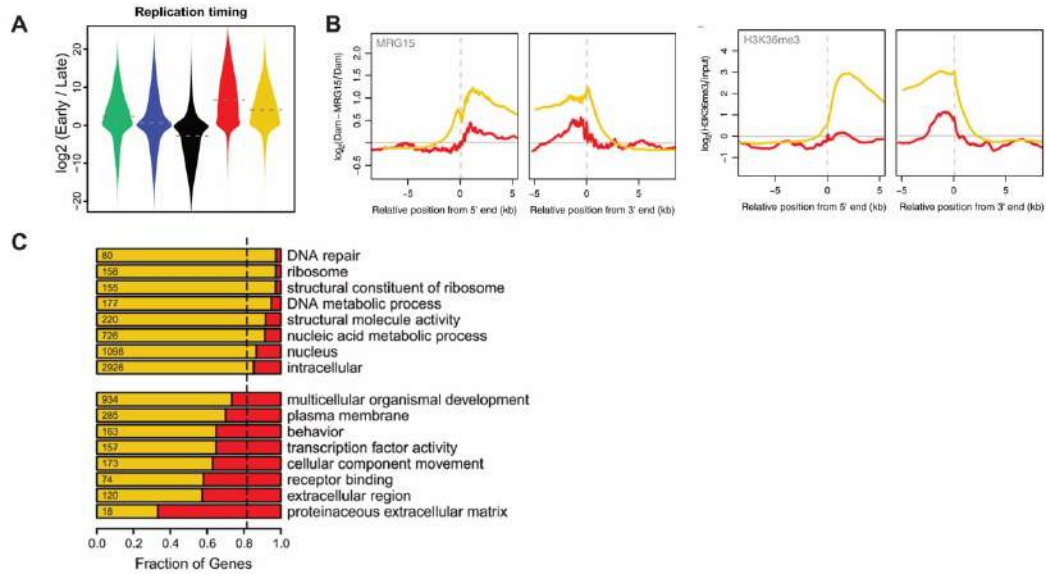


Figure 30 : Les chromatines jaune et rouge sont différentes.

A) Vitesse de réplication au sein de la chromatine jaune et rouge. La chromatine jaune se réplique plus rapidement que la rouge. B) Enrichissement ($\log_2(\text{Signal/Input})$) en MRG15 et H3K36me3 pour les deux types de chromatine. La chromatine jaune est fortement enrichie en MRG15 et H3K36me3. C) La chromatine jaune est enrichie en gènes associés à des GO Term des fonctions de ménage et la chromatine rouge est enrichie en GO Term associés à des fonctions plutôt développementales (Filion et al., 2010).

3.3.2.1.2.1 La chromatine jaune

La chromatine jaune est la chromatine qui contient le plus de gènes et, de surcroît, actifs dont un grand nombre de gènes de ménage impliqués dans des processus de base de la cellule comme la synthèse de ribosome ou la réparation de l'ADN (cf. Figure 30-C). La seule protéine spécifique à cette chromatine est la protéine MGR15 (cf. Figure 27-A ; Figure 30-A) qui est connue chez l'Humain pour être concomitante à la marque H3K36me3 (cf. Figure 30-B).

La chromatine jaune est donc définie par sa composition en gènes « de ménage », fortement traduits et un enrichissement spécifique en H3K36me3 et MGR15 (cf. Figure 31).

3.3.2.1.2.2 La chromatine rouge

La chromatine rouge est une zone de transcription active (cf. Figure 29-C) fortement enrichie en gènes correspondant aux termes « fixation au récepteur », « activité de facteur de transcription et « remodelage cellulaire » (cf. Figure 30-C). On y trouve notamment le facteur Su(Var)2-10, connu pour son implication dans le développement de l'œil²⁰, la sous-unité MED31 de complexe Médiateur, le récepteur à l'ecdysone EcR, des facteurs de type GAGA (GAF), Brahma et Lola-like, appartenant à la famille Trithorax (cf. Figure 27-A). Tous ces facteurs ont en commun d'être connus pour être impliqués dans l'activation de la transcription des gènes développementaux. Cette région de la chromatine est également fortement enrichie en histones H3 mono ou triméthylés sur la lysine 4 (H3K4me1/H3K4me3) ainsi qu'en histones H3 acétylés sur la lysine 27 (H3K27ac) (cf. Figure 27-B) permettant d'identifier les régions actives dont les enhancers.

²⁰ <https://flybase.org/reports/FBgn0003612.html>

La chromatine rouge est donc spécifique des gènes développementaux actifs et est associée aux protéines de la famille Trithorax ainsi qu'aux marques d'histones H3K4me1/me3 et H3K27ac (cf. Figure 31).

Ce modèle en cinq couleurs de chromatine est couramment utilisé mais comme tout modèle il tend à être amélioré et remis en question. Par approche similaire d'intégration de données, mais cette fois-ci de ChIP-seq disponible sur ModENCODE, *Kharchenko et al.* ont défini neuf états différents de la chromatine. Pour la définition de ces neuf groupes, les chercheurs ont, comme Filion *et al.*, comparé diverses protéines associées à la chromatine et des marques d'histones mais ils ont également tenu compte de l'annotation génique (intron, promoteur proximal, gène *etc.*). Cela permet ainsi d'extrapoler l'état de la régulation de la transcription (Kharchenko et al., 2011 b). Ainsi, les marques d'histones n'ont pas seulement permis d'appréhender les différentes catégories de chromatine mais permettent d'identifier différents éléments géniques comme les enhanceurs, les promoteurs ou bien le corps du gène.

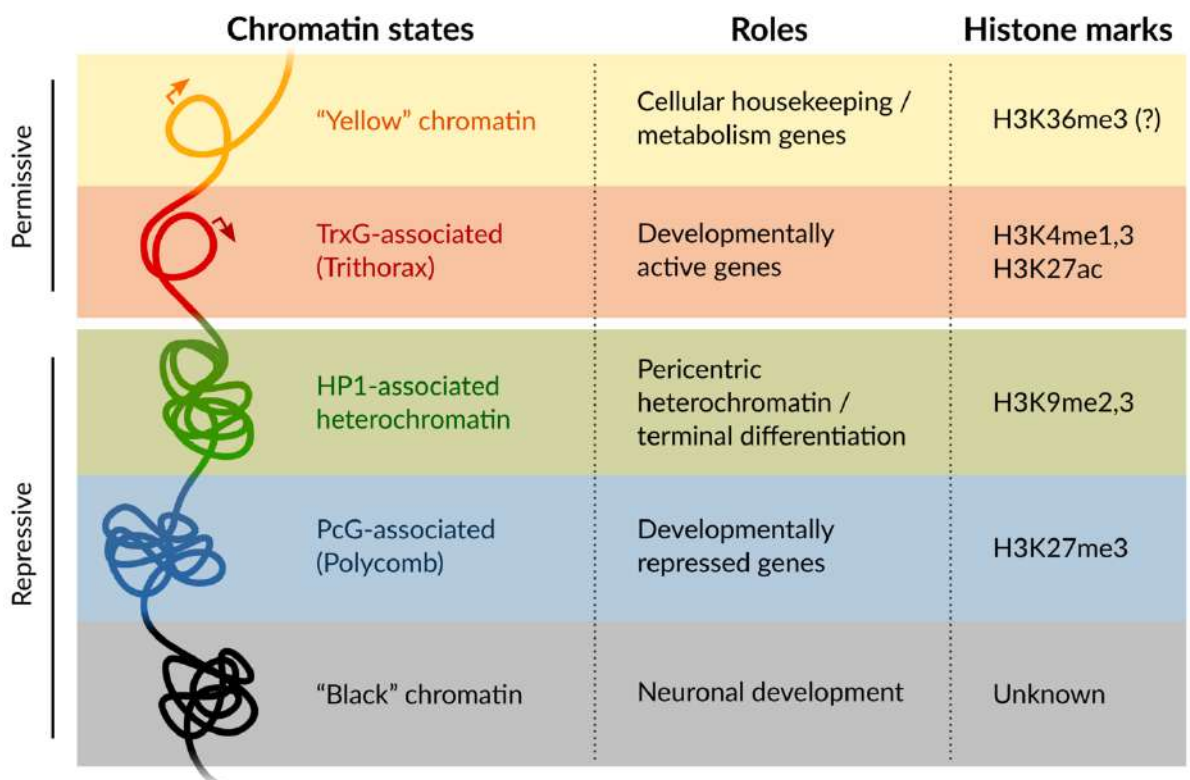


Figure 31 : Les différentes couleurs de la chromatine.
(Delandre et al., 2019)

3.3.2.2 Les histones comme signatures des régions cis-régulatrices

La disponibilité et la relative facilité d'obtention des données de fixation concernant de nombreuses marques d'histones a permis de définir des signatures spécifiques pour différents éléments géniques d'intérêt et en particulier en ce qui concerne les enhanceurs. Ainsi, grâce à des approches d'intégration (« machine learning », recherche de corrélation *etc*) de données de position des différentes marques d'histones avec les régions de fixation d'autres protéines connues comme les facteurs de transcription mais aussi avec des données d'expression, il a été défini des enrichissements caractéristiques/signatures de différents

éléments géniques et permettant même de définir leur état (enhancers actifs/inactifs, promoteurs actif/inactif) (Robertson et al., 2008; Cubeñas-Potts et al., 2017; Koenecke et al., 2017; Jiang et al., 2018 a). Ainsi, la marque d'histone H3K4me1 est associée aux promoteurs et aux enhancers qu'ils soient actifs ou non. La présence de la marque H3K27ac dans la zone H3K4me1 positive permet de prédire que cette région agit comme enhancer développemental actif (Cubeñas-Potts et al., 2017 ; S. Jiang & Mortazavi, 2018a). De même un élément génique est défini par plusieurs marques d'histone et protéines mais l'ajout d'une autre marque permet de savoir quel est son état vis-à-vis de la transcription. Par exemple l'identification d'un promoteur peut se faire grâce à la fixation d'ARN Polymérase II, la présence des marques H3K27ac, H3K36me3, H3K4me1, la fixation d'un ou plusieurs facteurs de transcription (et donc de la présence d'un site de fixation). Le ratio des marques H3K4me3/ H3K4me1 permet également de distinguer si la zone d'intérêt correspond à un enhancer ou un promoteur (cf. Figure 32). De plus, la présence des marques d'histones H3K9ac ou H3K12ac indique que ce promoteur est actif (Jiang et al., 2018 a).

Les enhancers sont identifiables, car ils sont présents dans une région de la chromatine accessible présentant des sites de fixation liés par des facteurs de transcription et une présence forte des marques d'histones H3K4me3 et H3K4me1 avec un ratio favorable à la marque H3K4me1. La présence importante de la marque H3K27ac permettra de définir les enhancers identifiés par les marques précédentes comme étant actifs. A l'inverse une présence modérée/faible de la marque H3K27ac est un indicateur de l'inactivité des enhancers dans les cellules testées (Lagha et al., 2012; Koenecke et al., 2017). Le profil de fixation de la marque d'histone a également son importance dans la définition de l'élément génique. Par exemple si le profil de la marque H3K4me3 ressemble à une montagne avec deux sommets, il est fort probable que l'on soit en train d'observer un enhancer alors que si elle présente un profil ressemblant à une dune, alors nous observons un promoteur (cf. Figure 32) (Cubeñas-Potts et al., 2017; Jiang et al., 2018 a).

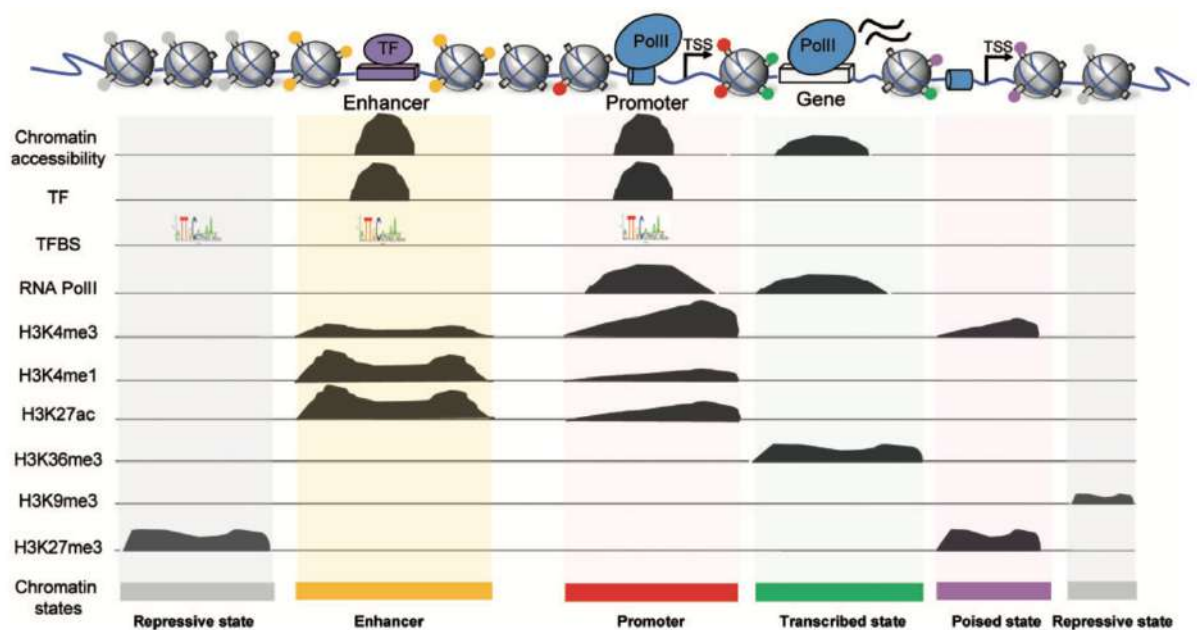


Figure 32 : Les marques d'histones comme signatures des éléments géniques. (S. Jiang & Mortazavi, 2018b)

C'est avec cette philosophie de définition de signatures que *Cubeñas-Potts et al.* ont décidé d'utiliser les données de STARR-seq pour caractériser les enhancers (Zabidi et al., 2015; Arnold et al., 2017; Cubeñas-Potts et al., 2017). Les enhancers des gènes développementaux (dCP) et de ménage (hkCP) ont été déterminés par le groupe d'Alexander Stark (Zabidi et al., 2015). Ce jeu de données comprend 4137 hkCP et 3586 dCP, les 1800 enhancers communs ont été retirés de l'étude. Ces enhancers sont également séparés en fonction de la distance au site d'initiation de la transcription (TSS) en enhancer proximal (présents à ± 250 pb d'un TSS) et enhancer distal (présent à >250 pb du TSS). Les données de STARR-seq ont ensuite été comparées aux données de ChIP-seq de quatre marques d'histones (H3K27ac ; H3K27me3 ; H3K4me1 et H3K4me3). Toutes les régions cis-régulatrices ayant un chevauchement avec de la marque H3K27ac (de minimum 1bp) sont considérées comme étant actives. Les autres présentent un enrichissement en H3K27me3, marque associée à du « silencing », ces enhancers sont donc qualifiés d'inactifs.

Les enhancers sont maintenant séparés en : enhancer dCP-distal actif ou inactif ; enhancer dCP-proximal actif ou inactif ; enhancer hkCP-distant actif ou inactif et enhancer hkCP-proximal actif ou inactif. Pour chacun de ces groupes, l'enrichissement en H3K4me1 et H3K4me3 est calculé. Les enhancer dCP qu'ils soient distaux ou proximaux présentent un enrichissement en H3K4me1. Les enhancers hkCP sont quant à eux enrichis en H3K4me3 bien que les proximaux soient plus enrichis que les distaux (cf. Figure 33). Cette dernière observation peut être due au biais de la proximité du promoteur.

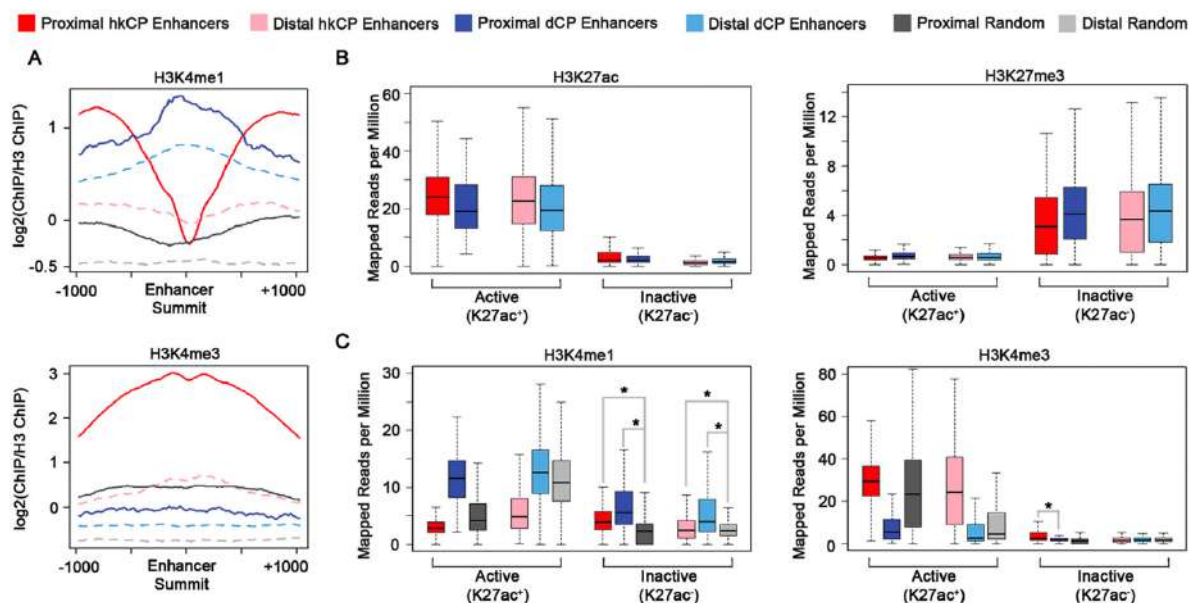


Figure 33 : Signature en histones des enhancers développementaux et de ménage.

A) Enrichissement ($\log_2(\text{Signal}/\text{Input})$) en H3K4me1 et H3K4me3 pour les différents types d'enhancers. B) Nombre de reads par Millions des histones H3K27ac ; H3K27me3 ; H3K4me1 ; H3K4me3 sur les différents groupes d'enhancers. (Cubeñas-Potts et al., 2017)

Pour conclure, il est intéressant de noter que la distribution des marques épigénétiques des histones peut être utilisée pour prédire la nature des éléments géniques ainsi que leur état transcriptionnel. De la même façon, ces marques associées à la distribution de protéines remodelleuses (Chromator, BEAF-32, Su(Hw), Cohésine, entre autres) peuvent être utilisées dans la lutte contre les cancers. En effet la dérégulation de ces marques d'histones (méthylation et/ou d'acétylation) sont associées à nombreux cancers dont des

cancers enfantins, des leucémies aiguës myéloïdes et des cancers gastriques ce qui fait des HDAC, HAT, HKMT de cibles thérapeutiques potentielles (Tam et al., 2013; Guo et al., 2015 b; Baylin et al., 2016 c; Hino et al., 2016; Jones et al., 2016 a; Gupta et al., 2019; Biran et al., 2020; Hillyar et al., 2020; Verza et al., 2020).

Objectifs

Une des fonctions majeures des facteurs OvoL codés, chez les vertébrés par trois locus distincts est de garantir le maintien du caractère épithélial des cellules. Ils sont considérés comme les gradients du caractère épithélial et luttent contre le processus de Transition Epithélio-Mésenchymateuse (EMT). Leur étude est donc essentielle pour la compréhension des mécanismes de plasticité épithéliale requise pour le développement et détournée dans certaines pathologies dont le cancer.

Chez la mouche du vinaigre, un seul locus *ovo/shavenbaby* code pour un facteur OvoL, facilitant ainsi son analyse fonctionnelle. Le locus *ovo/svb*, permet la production de trois isoformes protéiques distinctes : deux germinales OvoA et OvoB et une somatique SvB (Oliver et al., 1987, 1994; Garfinkel et al., 1992). SvB est tout d'abord synthétisée sous une forme longue répressive SvBRep qui en présence de smORF peptides Pri est dégradée partiellement, aboutissant à la formation d'une forme plus courte et activatrice SvBAct (Kondo et al., 2010; Zanet et al., 2015).

Mon travail de thèse a consisté à caractériser plus en profondeur, par des approches génomiques, le mode de fonctionnement spécifique de chaque forme de SvB grâce à la mise en place des lignées cellulaires stables exprimant l'une ou l'autre des formes. Mes objectifs étaient :

- de caractériser le mode de fixation de SvBREP et SvBACT (analyse d'expériences de CHIP-seq),
- de déterminer l'environnement de fixation (chromatinien) des deux formes de SvB en suivant les modifications post-traductionnelles d'histones (CHIP-seq) et d'accessibilité de la chromatine (ATAC-seq),
- d'identifier les gènes cibles directes de SvB (analyse intégrative de données de RNA-seq et CHIP-seq) dans les cellules S2 et dans les cellules souches intestinales,
- de comparer les différentes cohortes de gènes cibles de SvB identifiées dans distincts types cellulaires (cellules épidermiques de l'embryon, cellules souches intestinales et cellules S2).

En parallèle de ce projet principal, j'ai participé au projet de Jérôme Bohère dans la caractérisation de l'effet de Shavenbaby au sein des cellules souches rénales de la Drosophile, qui a donné lieu à une publication dans ***Nature Communication*** en 2018 (cf. Résultats Chapitre 4). J'ai également pris part au projet d'Hélène Chanut-Delalande pour la caractérisation des enhanceurs contrôlant l'expression du gène *pri*, étude publiée cette année dans ***Frontiers of Genetics*** (cf. Résultats Chapitre 5).

Dans la partie Résultats qui suit, vous trouverez les différents articles et manuscrits d'articles pour ces différents projets avec une remise en contexte et ma contribution pour chacun d'entre eux.

Résultats

Chapitre 4 Shavenbaby protège les cellules souches rénales en interagissant avec la voie Hippo

4.1 Contexte de la publication

Comme je l'ai mentionné dans la partie Introduction, la caractérisation fonctionnelle de Svb dans les cellules digestives adultes a donné lieu à deux études réalisées dans l'intestin par Sandy Al Hayek et dans les tubules de Malpighi par Jérôme Bohère.

La Drosophile adulte possède deux paires de tubules de Malpighi, sensiblement différentes, assurant des fonctions analogues aux reins des mammifères (Rodan, 2019). La paire de droite se projetant vers la région antérieure est plus longue et rattachée à l'intestin moyen antérieur. La paire de gauche est plus courte et est rattachée à l'intestin postérieur. Ces tubules, qui sont les seuls à ne pas se dégénérer lors de la métamorphose (LM, 1993), sont complètement immergés dans le système circulatoire. Ils assurent le maintien de la balance en eau, ions et l'homéostasie acido-basique ainsi que la détoxification de l'hémolymphe (Cohen et al., 2020). Outre leur fonction de régulation osmotique et ionique, les tubules de Malpighi semblent être impliqués dans des fonctions du système immunitaire. Ils peuvent notamment produire des peptides antimicrobiens comme la Diptericin et la Drosomycin. Des études sur le cycle circadien montrent une implication de ces tubules dans sa régulation (Rodan, 2019; Cohen et al., 2020)

Les tubules de Malpighi présentent une organisation très régionalisée et complexe suivant l'axe proximo-distal et chaque région possède des fonctions distinctes. Ils peuvent être catégorisés en deux grandes parties, elles-mêmes sous-divisées :

- **Les tubules inférieurs** comprenant l'uretère reliant les tubules à l'intestin et les tubules inférieurs proprement dits
- **Les tubules supérieurs** composées de trois segments dits principal, de transition, et initial (Figure 34-A)

Il est admis que les tubules de Malpighi sont composés deux types de cellules différenciées : les cellules principales (CP) d'origine ectodermique et les cellules étoilés (CE) d'origine mésodermique (Denholm et al., 2003) (cf. Figure 34-B). Ces deux types cellulaires se distinguent par leur forme, leur fonction et par les facteurs qu'elles expriment. Les CP, présentes sur toute la longueur des tubules de Malpighi, expriment le facteur de transcription Cut (Ct). Les CE, localisées uniquement dans les tubules supérieurs, expriment quant à elle le facteur de transcription Teashirt (Tsh). Ces gènes servent ainsi de marqueurs lors des approches de microscopies.

Les CP alimentent un gradient de protons pour faciliter le transport des cations Na^+ et K^+ par l'intermédiaire d'une pompe ATPase qui crée un potentiel électrochimique via le transport des ions H^+ permettant leur passage dans le lumen. Les CE assurent le passage de

l'eau de l'hémolymphe au lumen via des aquaporines ainsi que le passage des ions Cl⁻ (Dow et al., 2021). Comme pour l'intestin, cette vision est quelque peu restrictive et il existe des évidences de la diversité des cellules matures basées aussi bien sur leur morphologie que sur l'expression de marqueurs spécifiques (Sözen, PNAS).

Un troisième type cellulaire nommé **cellules réabsorbantes**, spécifique des tubules inférieurs et exprimant à la fois *cut* et *aph-4* codant pour la phosphatase alcaline 4, semble participer à la réabsorption de l'eau des tubules vers l'hémolymphe, une propriété spécifique des tubules inférieurs (Rheault et al., 2004; Beyenbach et al., 2010).

Il a longtemps été pensé que les cellules qui composent les tubules étaient générées au cours de l'embryogenèse et pour toute la vie de l'individu. Cette vision a été bousculée par l'identification des **cellules souches rénales** (CSR) en 2007 (Singh et al., 2007 a), peu de temps après les cellules souches intestinales. Il a été initialement proposé que ces cellules souches soient capables d'assurer le renouvellement des CP et CE (cf. Figure 34-B) (Singh et al., 2007 a). Des analyses effectuées dans l'équipe par Jérôme et Cédric ainsi qu'une étude récente fondée sur une analyse de transcriptome sur cellule unique (Wang et al., 2020 a) viennent contredire cette vision. Si les CSR sont bien des cellules souches quiescentes se divisant rarement et uniquement suite à un stress, elles assurent uniquement le renouvellement des cellules réabsorbantes situées dans les tubules inférieurs. Les CSR sont notamment incapables d'assurer le renouvellement des CE. L'analyse sur cellule unique a aussi permis de subdiviser les cellules réabsorbantes en trois types cellulaires distincts dont les fonctions et la filiation restent à définir.

Les CSR rénales sont issues de la migration des précurseurs des cellules souches intestinales (AMP) présents chez la larve. Lors de la pupaison, sous l'influence des voies de signalisation Wg (registre spatial) et ecdysone (registre temporel), une partie des AMP migre depuis l'intestin moyen pour coloniser l'uretère des tubules de Malpighi (C. Wang & Spradling, 2020a). Ainsi, les CSR et les CSI ont donc pour origine commune les AMP. Bien que dix fois moins nombreuses que les CSI (350 pour 3000), elles partagent avec ces dernières de nombreuses caractéristiques. Elles expriment bien entendu le gène *escargot*, et sont régulées par les voies de signalisation Jak/STAT, Notch, EGF, Wg et Hippo (223–225). Leur spécificité vient principalement de l'expression du facteur de transcription Cut qui détermine les CSR. Cut est activé lors de la sélection des AMP par l'intermédiaire d'un enhancer spécifique sensible aux signalisations Wg et Ecdysone. En effet, ces souches rénales dépourvues de *ct* donnent naissance à des EC intestinaux au sein des tubules de Malpighi (Wang et al., 2020 b).

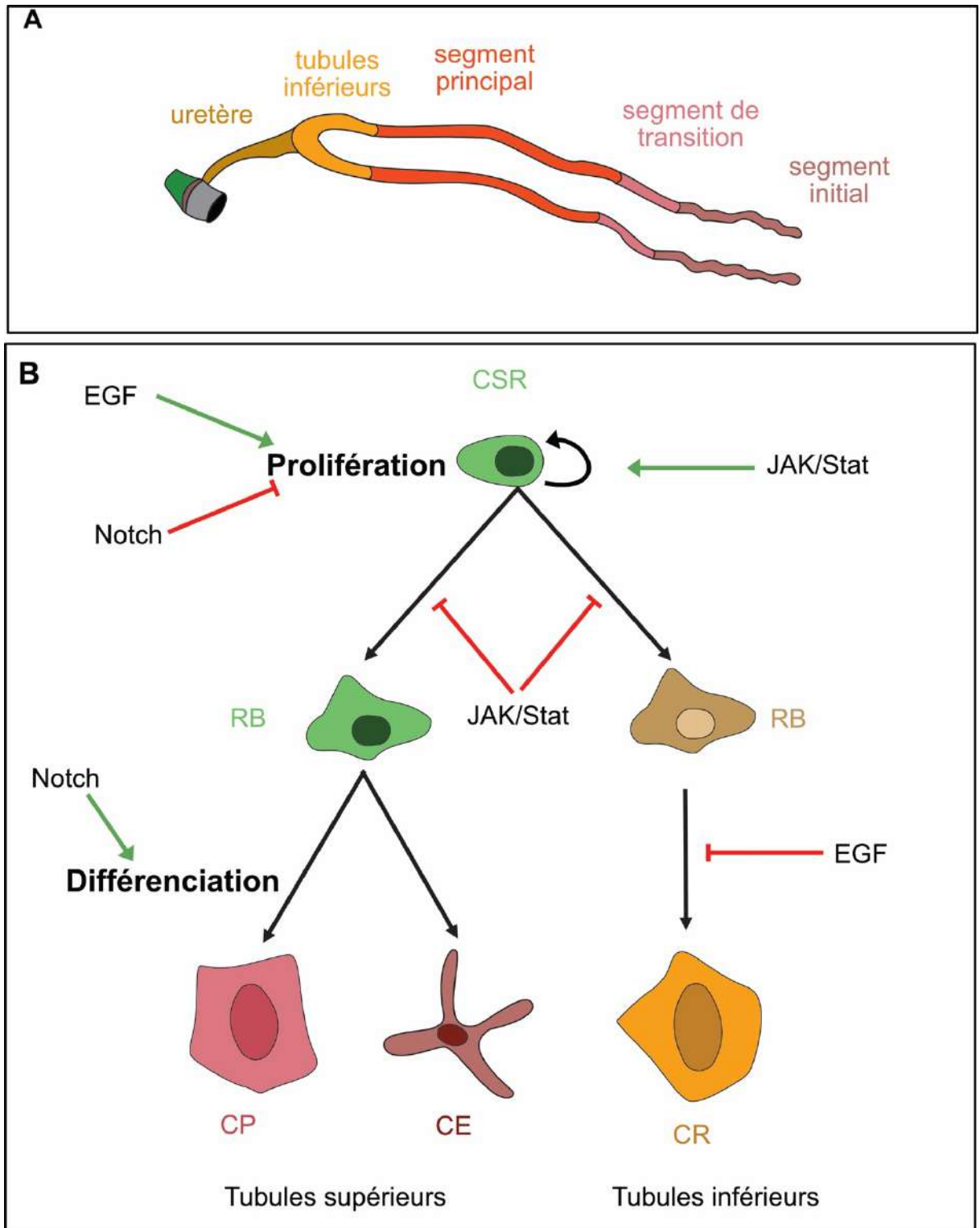


Figure 34 : Tubules de Malpighi et lignage cellulaire des cellules souches rénales.

Ainsi, bien que partageant des spécificités propres et originales telles que la migration, la quiescence ou la reprogrammation, les cellules souches rénales sont encore mal connues et relativement peu étudiées. Dans ce contexte, l'étude réalisée par Jérôme a mis en évidence le rôle de *svb* pour assurer la survie de ces cellules.

4.2 Résumé de la publication

Tout comme dans l'intestin, l'expression dans les CSR de *svb* est soutenue par l'activité des enhanceurs E3 et E6 qui assurent ici encore des fonctions redondantes. Jérôme a développé une méthode semi-automatique qui permet de quantifier précisément le nombre de CSR. Elles sont au nombre de 350 dans des conditions contrôles. La perte de fonction de *svb* entraîne une diminution progressive du nombre de CSR qui disparaissent après 30 jours de traitement à l'ARN interférant (RNAi). La présence de SvbRep ou la perte de *pri* et d'*ubr3* dans ces cellules ont un effet similaire. Ainsi dans ce système la maturation de Svb, Pri et Ubr3 sont nécessaires pour l'activité de Svb. Ici, et contrairement aux CSI, la présence de SvbAct n'est pas suffisante pour induire leur prolifération.

La disparition progressive des CSR en contexte *svb*-mutant est le fait d'un excès d'apoptose qui peut être contrebalancé par la présence de protéine anti-apoptotique p35 et DIAP1. L'originalité de ce travail a été de lier Svb avec la voie de signalisation Hippo connue pour contrôler la balance prolifération/apoptose. Jérôme a mis en évidence une interaction physique entre la protéine Svb et Yorkie (Yki), l'effecteur nucléaire de la voie Hippo (Ma et al., 2019). L'action concertée de ces deux protéines semble requise pour soutenir l'expression de la protéine anti-apoptotique DIAP1 et donc protéger les cellules souches de l'apoptose.

4.2.1 Contribution

Ma contribution à ce projet a été de déterminer si les facteurs Yki et Svb pouvaient potentiellement réguler le même ensemble de gènes. Notre stratégie a été d'utiliser les données de CHIP-seq disponibles pour ces deux facteurs et vérifier s'ils avaient des sites communs de fixation à l'ADN. Pour ce faire, j'ai utilisé les données de CHIP-seq pour Svb qui avaient été réalisées pour les cellules épidermiques embryonnaires (Menoret et al., 2013), ainsi que les données publiées de CHIP-seq de Yki réalisées à partir des disques imaginaux d'ailes (Oh et al., 2013). Après avoir aligné les données sur le génome de référence de *Drosophila melanogaster* (version dm6), j'ai déterminé les sites de fixation à l'ADN avec le logiciel de détection de pics de fixation MACS (Zhang et al., 2008) et comparé ces sites à l'aide de du package ChIPpeakAnno (Zhu et al., 2010). Grâce à cette approche, environ 1300 sites de fixation génomiques communs statistiquement pertinent ont été identifiés. Si Yki ne semble pas participer (absence de fixation de Yki) à la régulation des gènes cibles épidermiques de *svb*, la protéine Svb est fréquemment présente sur les éléments cis-régulateurs liés par Yki pour réguler ses gènes cibles dont *bantam*, *piwi*, *fat*, et également *DIAP1*.



Ces données suggèrent donc une coopération du facteur de transcription Svb et du cofacteur Yki dans la régulation des gènes critiques des CSR.

ARTICLE

DOI: 10.1038/s41467-018-07569-0

OPEN

Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis

Jérôme Bohère¹, Alexandra Mancheno-Ferris¹, Sandy Al Hayek^{1,2,3}, Jennifer Zanet¹, Philippe Valenti¹, Kohsuke Akino⁴, Yuya Yamabe⁴, Sachi Inagaki⁵, Hélène Chanut-Delalande¹, Serge Plaza^{1,6}, Yuji Kageyama^{4,5}, Dani Osman^{2,3}, Cédric Polesello¹  & François Payre¹ 

To compensate for accumulating damages and cell death, adult homeostasis (e.g., body fluids and secretion) requires organ regeneration, operated by long-lived stem cells. How stem cells can survive throughout the animal life remains poorly understood. Here we show that the transcription factor Shavenbaby (Svb, OvoL in vertebrates) is expressed in renal/nephric stem cells (RNSCs) of *Drosophila* and required for their maintenance during adulthood. As recently shown in embryos, Svb function in adult RNSCs further needs a post-translational processing mediated by the Polished rice (Pri) smORF peptides and impairing Svb function leads to RNSC apoptosis. We show that Svb interacts both genetically and physically with Yorkie (YAP/TAZ in vertebrates), a nuclear effector of the Hippo pathway, to activate the expression of the inhibitor of apoptosis *DIAP1*. These data therefore identify Svb as a nuclear effector in the Hippo pathway, critical for the survival of adult somatic stem cells.

¹Centre de Biologie du Développement (CBD), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Bat 4R3, 118 route de Narbonne, F-31062 Toulouse, France. ²Faculty of Sciences III, Lebanese University, Tripoli 1300, Lebanon. ³Azm Center for Research in Biotechnology and its Applications, LBA3B, EDST, Lebanese University, Tripoli 1300, Lebanon. ⁴Department of Biology, Graduate School of Science, Kobe 657-8501, Japan. ⁵Biosignal Research Center, Kobe University, 1-1 Rokko-dai, Nada, Kobe 657-8501, Japan. ⁶Present address: Laboratoire de Recherche en Sciences Végétales (LSRV), CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, 31326 Castanet-Tolosan, France. Correspondence and requests for materials should be addressed to C.P. (email: cedric.polesello@univ-tlse3.fr) or to F.P. (email: francois.payre@univ-tlse3.fr)

The family of OvoL/Ovo/Shavenbaby (Svb) transcription factors has been strongly conserved across evolution and is characteristic of animal species. Initially discovered in flies for a dual function in the development of the germline and of epidermal derivatives (Ovo/Svb)^{1,2}, mammalian orthologs (OvoL1-3) have soon been identified^{3,4}. *OvoL/svb* genes produce several protein isoforms and the existence of three partially redundant paralogs in mammals complicates their genetic analysis. There is a single gene in *Drosophila*, which expresses germline- (*ovo*) and somatic-specific (*svb*) transcripts from different promoters. Previous work has well-established the role of Svb in the development of embryonic epidermal tissues², where it triggers a tridimensional cell shape remodeling for the formation of actin-rich apical extensions, called trichomes. *Svb* expression is driven by a large array of *cis*-regulatory regions, which have become a paradigm for elucidating the function^{5,6} and evolution⁷⁻⁹ of developmental enhancers. *Svb* enhancers directly integrate multiple inputs from upstream regulatory pathways^{2,5,10} and often drive similar patterns^{5,6,11}, together conferring robustness to epidermal development in the face of varying environmental conditions and/or genetic backgrounds^{5,6}. During embryogenesis, the Svb transcription factor directly activates a battery of >150 target genes¹²⁻¹⁴ collectively responsible for actin and extra-cellular-matrix reorganization that underlies trichome formation. Recent studies have unraveled a tight control of the Svb protein activity in response to Polished rice peptides (Pri, also known as Tarsal-less), which belongs to a fast-growing family of peptides encoded from small open reading frames (smORF) hidden within apparently long noncoding RNAs^{15,16}. Svb is first translated as a long-sized protein that acts as a repressor (Svb^{REP})¹⁷. Pri smORF peptides then induce a proteolytic processing of Svb^{REP} leading to the degradation of its N-terminal region and releasing a shorter activator form, Svb^{ACT}^{17,18}. Further work has demonstrated that *pri* expression is directly regulated by periodic pulses of steroid hormones¹⁹, allowing a functional connection between hard-wired genetic regulatory networks (*svb* expression) and systemic hormonal control (*pri* expression) for a proper spatio-temporal control of epidermal cell morphogenesis¹⁵.

Recent studies suggest that OvoL/Svb factors display broader functions throughout epithelial tissues in both normal and various pathological situations. Molecular profiling of human tumors has revealed that OvoL deregulation is a feature of many carcinomas, directly linked to the metastatic potential of morbid cancers²⁰⁻²³, including kidney²⁴. OvoL factors have been proposed^{25,26} to counteract a conserved core of regulators composed of Snail/Slug and Zeb1-2 transcription factors, as well as the micro RNA *mir200*, well known to promote epithelial-mesenchymal transition (EMT)²⁷. The activity of OvoL might help stabilizing a hybrid phenotype between epithelial and mesenchymal states²⁵, providing many advantages for both tumors and normal stem cells²⁸. Indeed, recent data show that, like adult somatic stem cells, the most aggressive tumors often display a hybrid E/M phenotype²⁷, and the expression of specific OvoL isoforms can annihilate the metastatic potential of mammary tumors^{20,29}. In addition, OvoL/Svb factors have been linked to the control of various progenitors/stem cells, from basal invertebrates³⁰ to humans³¹⁻³³. Therefore, whereas a large body of evidence supports a key role for OvoL/Svb in the behavior of somatic stem cells, a functional investigation of their mode of action *in vivo* remains to be undertaken.

Here, we built on the knowledge and tools accumulated for the study of Svb function in flies to investigate its putative contribution to the behavior of somatic stem cells in the adult. We show that in Malpighian tubules, which ensure renal function in insects^{34,35}, *svb* is specifically expressed in the adult renal/nephric stem cells (RNSCs). We further find that the main function of Svb

in the kidney is to protect RNSCs from apoptosis by controlling the expression of the inhibitor of apoptosis, *DIAP1*, in interaction with Yorkie (Yki), a nuclear effector of the Hippo pathway.

Results

***svb* is expressed in RNSCs and controls their maintenance.** To assay whether *svb* might be expressed in the adult, we tested large genomic reporter constructs that cover each of the seven enhancers contributing to *svb* expression^{7,8}. We found that one enhancer, *svb^E*⁸, drove specific expression in tiny cells of the Malpighian tubules (Supplementary Figure 1a, b).

Malpighian tubules are mainly composed of two types of differentiated cells³⁵. The principal cells—characterized by the homeodomain Cut protein (Fig. 1a, b)—express the vacuolar-ATPase (V-ATPase) that establishes an H⁺ electrochemical potential promoting *trans*-epithelial secretion of Na⁺ and K⁺³⁴. The second main population of Malpighian tubules are termed stellate cells, featured by the expression of the Teashirt (Tsh) transcription factor (Fig. 1a), and that regulate the transport of Cl⁻ and water³⁴. While both principal and stellate cells display large-sized polyploid nuclei, a third population of small diploid cells, originally referred to as tiny cells with putative myoendocrine and/or neuroendocrine activity^{36,37}, are located in the lower tubules (Fig. 1a, b). Accumulated evidence now supports that these tiny cells ensure the renewal of at least some populations of kidney cells^{38,39} and correspond to adult RNSCs (see also Supplementary Figure 4). It has been shown that RNSCs derive from a subpopulation of intestinal stem cell precursors that migrate in Malpighian tubules during post-embryonic development⁴⁰. RNSCs are characterized by the expression of Escargot (Esg), a transcription factor of the Snail/SLUG family that is also expressed in intestinal stem cells⁴¹ where it acts to prevent stem cell differentiation^{42,43}. Co-localization with an *esg-LacZ* reporter confirmed that the *svb^E* enhancer was active in RNSCs (Fig. 1c and Supplementary Figure 1). To define the minimal region of *svb* responsible for the expression in RNSCs, we assayed a collection of overlapping constructs⁵. This identified two independent elements, the *svb^{E3N}* and *svb^{E6}* enhancers^{5,7}, which despite having distinct activities during embryogenesis⁷ both drive similar expression in adult RNSCs (Supplementary Figure 1c).

Having established that two enhancers drive specific expression of *svb* in the adult stem cells of the renal system, we next assayed consequences of depleting *svb* function in RNSCs. We used a well-controlled genetic system, hereafter referred to as *esg^{ts}*. *esg^{ts}* ensures RNAi-mediated gene depletion, specifically in stem cells⁴⁴, as monitored by GFP expression (Fig. 2a, b and Supplementary Figure 2a). In addition, the expression of *esg^{ts}*-driven transgene(s) and/or RNAi is tightly regulated in a temporal manner by a temperature shift, ensuring that stem cell manipulation is turned ON only in adults (3 days after eclosion) to rule out any earlier developmental defects (Fig. 2c). We also developed an image analysis pipeline, allowing automated quantification of the whole population of RNSCs (see Methods). In control conditions, the number of *esg*-positive RNSCs remains stable after adult hatching, with approx. 350 cells *per* tubules (Fig. 2a–c). We only noticed a weak reduction of RNSCs after 1 month. In contrast, *esg^{ts}*-driven RNAi depletion of *svb* in adults led to a progressive loss of RNSCs, which were completely absent after 32 days of treatment (Fig. 2a–c). Although not affecting RNSCs before (day 0, Supplementary Figure 2d)—or shortly after (2 days, Fig. 2c)—inducing transgene expression, the effects of *svb* depletion in the adults were already strong following 8 days of treatment, with a two-fold reduction in the number of RNSCs (Fig. 2a–c and Supplementary Figure 2). Similar results were observed when using either a second RNAi construct

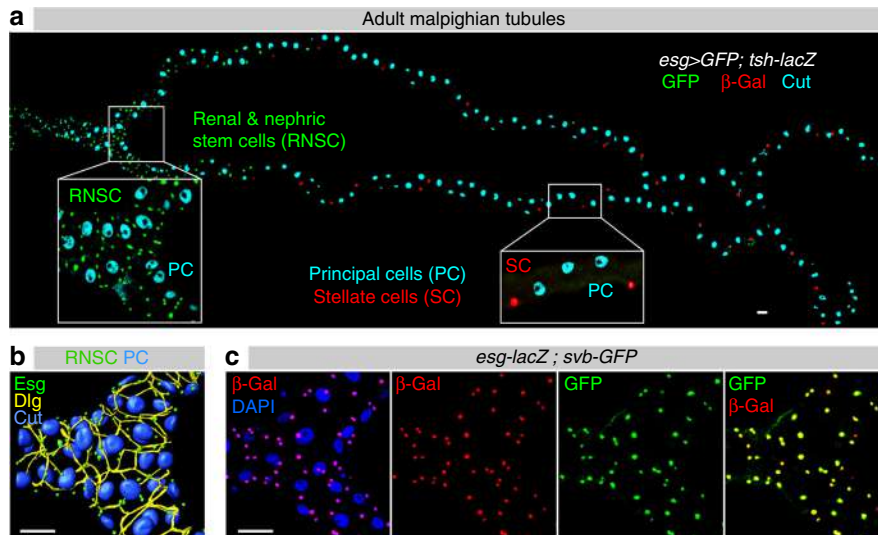


Fig. 1 *svb* is specifically expressed in renal stem cells. **a** Adult Malpighian tubules are composed of three types of cells. Principal cells (PC) are identified by immunostaining against Cut (cyan) and stellate cells (SC) by *tsh-lacZ* (red). RNSCs, located in the lower tubules, express *esg-Gal4*, *UAS-GFP* (green). **b** 3D reconstruction of the fork region of Malpighian tubules, with *esg*-positive RNSCs in green, immunostaining against Discs Large (Dlg) and Cut in yellow and blue, respectively. **c** Expression of *svb* and *esg* as monitored by co-staining for *svb-E-GFP* (green) and *esg-LacZ* (red) enhancers, respectively. Nuclei were counterstained with DAPI (blue). Scale bar is 30 μ m

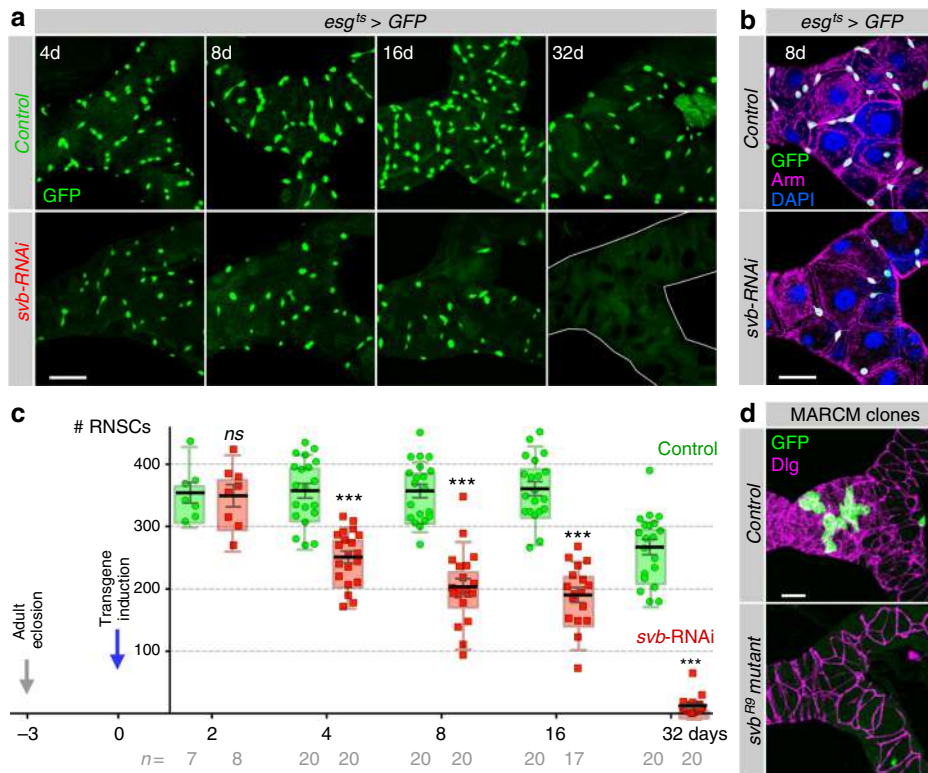


Fig. 2 *svb* is required for RNSC maintenance. **a** *esg^{ts}*-driven *svb-RNAi* leads to a progressive decrease of RNSC number compared to controls (*esg^{ts}* driving only GFP). **b** *svb* depletion eliminates RNSCs, without affecting typical features of remaining stem cells, e.g., *esg* expression and Arm accumulation (see also Supplementary Figure 2). **c** Quantification of the number of RNSCs (*esg*-positive) after 2 ($p = 0.9551$), 4 ($p < 0.0001$), 8 ($p < 0.0001$), 16 ($p < 0.0001$), and 32 days ($p < 0.0001$) of transgene induction in control (green) and *svb-RNAi* (red) conditions. **d** MARCM of control and *svb^{R9}* clones, positively labeled with GFP (green), 25 days after clone induction. See Appendix for full genotypes. In this and all subsequent figures, each dot corresponds to an independent sample; results were combined from at least two independent experiments. Values are presented as average \pm standard error of the mean (SEM) and boxes with whiskers (10–90 percentile). p -Values from Mann-Whitney test (ns, $p \geq 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). Scale bar is 30 μ m

(Supplementary Figure 2b), or an independent driver of RNSCs (*dome-MESO-gal4*) to knockdown *svb* (Supplementary Figure 2e). We identified another transcription factor, Hindsight (Hnt), as being specific of RNSCs within Malpighian tubules (Supplementary Figure 2c); and the loss of RNSCs upon *svb* depletion was confirmed by staining against Hnt (Supplementary Figure 2e). Finally, the key role of *svb* in the maintenance of adult RNSCs was further demonstrated by results from genetic mosaics (MARCM⁴⁵), showing that mutant cells bearing strong alleles of *svb* were unable to maintain RNSCs (Fig. 2d and Supplementary Figure 2f, g).

Taken together, these data thus reveal that *svb* is specifically expressed in RNSCs and critically required for the maintenance of the adult stem cell compartment.

Svb processing is essential for its activity in RNSCs. In the epidermis, Svb activity relies on a proteolytic processing that

causes a switch from a repressor to an activator form¹⁵ (see Fig. 3a). This processing is gated by Pri smORF peptides, which bind to and activate the Ubr3 ubiquitin E3-ligase that, in turn, triggers a limited degradation of Svb operated by the proteasome¹⁸. Thereby, *pri* mediates a systemic control of Svb maturation since the expression of *pri* is directly regulated by the ecdysone receptor (EcR)¹⁹.

To assess whether the function of Svb in Malpighian tubules also required its proteolytic maturation, we investigated a putative function of *pri* and *ubr3* in RNSCs. We screened a collection of *pri* reporter lines^{19,46} and identified two *cis*-regulatory regions driving expression in RNSCs (Fig. 3b and Supplementary Figure 3a, b). Consistently with the expression of *pri* in RNSCs, *pri* knockdown strongly impacted the population of RNSCs (Fig. 3c–e and Supplementary Figure 3c). The effects of RNAi-mediated depletion on RNSCs were even stronger for *pri* than for *svb*, as previously observed in the epidermis¹⁷. The lack of Pri peptides indeed leads to the accumulation of the Svb repressor^{14,17,19},

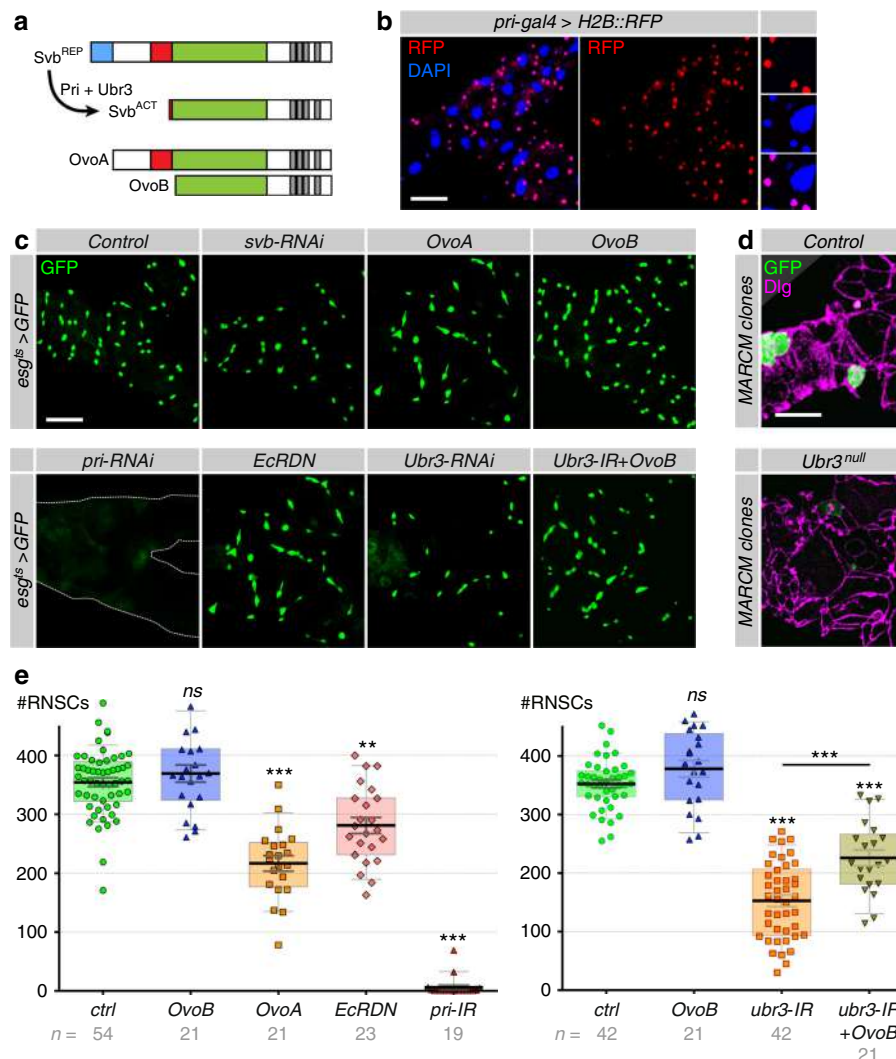


Fig. 3 Processing of Svb is essential for RNSC maintenance. **a** Schematic representation of Svb maturation, as well as the germinal isoforms OvoA and OvoB that act as constitutive (*pri*-independent) repressor and activator, respectively. **b** Expression of *pri* monitored by the activity of *pri* (*tal*)-*Gal4* driving the expression of H2B::RFP (red). Nuclei (DAPI) are in blue. **c** Fork region of Malpighian tubules, with *esg*^{TS}-driven expression of GFP and the indicated transgenes, after 8 days of induction. **d** MARCM clones (GFP-positive) of control and *ubr3*^B cells, 25 days after induction. **e** Quantification of the number of *esg*⁺ cells per tubule for the indicated genotype. Values are presented as average ± standard error of the mean (SEM) and boxes with whiskers (10–90 percentile). Unless otherwise indicated, *p* values correspond to comparison with control conditions: *OvoB* (*p* = 0.4899), *OvoA* (*p* < 0.0001), *EcRDN* (*p* < 0.0001), *pri*-RNAi (*p* < 0.0001); *OvoB* (*p* = 0.118), *ubr3*-RNAi (*p* < 0.0001), *ubr3*-RNAi vs *ubr3*-RNAi + *ovoB* (*p* < 0.0001). *p*-Values from Mann–Whitney test (ns, *P* > 0.05; ****P* < 0.001). Scale bar is 30 μm

which may explain a stronger downregulation of *Svb* target genes when compared to *svb* mutants¹⁴. In the case of RNSCs, it is also possible that *pri*-RNAi depletion is more efficient (smORF peptides are likely very unstable when compared to typical protein half-life), and/or that *Pri* peptides fulfil *Svb*-independent functions as also reported during embryogenesis¹⁷. In addition, a dominant negative form of the ecdysone receptor (EcRDN) that abolishes *pri* expression during both embryonic and post-embryonic development¹⁹ was sufficient to reduce the number of stem cells when specifically expressed in adult RNSCs (Fig. 3c–e). Furthermore, we found that *ubr3* was also required for RNSC maintenance, as deduced from results of RNAi-mediated depletion (using three non-overlapping constructs) or genetic nullification¹⁸ of *ubr3* activity (Fig. 3c–e, Supplementary Figure 3d). Finally, the expression of *OvoA* (see Fig. 3a) that behaves as a constitutive repressor isoform of *Svb*^{17,47,48} mimicked the effects observed in *svb* loss of function conditions (Fig. 3c, e). Reciprocally, the expression of *OvoB* (see Fig. 3a) that acts as a constitutive activator isoform of *Svb*^{17,47,48} was sufficient to rescue the lack of *ubr3* function (Fig. 3c, e), demonstrating that *Svb* function in RNSCs relies on its matured transcription activator form.

These results provide compelling evidence that the whole regulatory machinery discovered for its role in the development of epidermal cells^{17–19} is also at work in adult RNSCs. We therefore concluded that the post-translational maturation of the *Svb* transcription factor is essential for the maintenance of RNSCs.

***Svb* protects renal nephric stem cells from apoptosis.** The loss of RNSCs observed following the lack of *svb* function or maturation could theoretically result from at least three different causes: (i) lack of proliferation, (ii) precocious differentiation, or (iii) increased cell death. We therefore assayed the putative contribution of each of those aspects in the loss of stem cells resulting from *svb* inactivation.

Consistent with the quiescent behavior of RNSCs, we observed a low frequency of RNSC division in controls, as deduced from staining with the mitotic marker phospho-Histone H3 (Supplementary Figure 4a, b) and as previously noticed³⁸. Additional cell lineage analysis using the *esg-Gal4* and *dome-MESO-Gal4* drivers confirmed that adult renal stem cells/progenitors give rise to a progeny of large differentiated cells (Cut positive) in lower tubules (Supplementary Figure 4c, d). Similar analyses using the *Alkaline-Phosphatase 4* driver (*Aph4-Gal4*) that is specific of differentiated principal cells⁴⁹ in lower tubules showed no progenitor/progeny figures (Supplementary Figure 4c), supporting that only RNSCs are able to sustain cell renewal. Therefore, even a complete block of stem cell division cannot account for the disappearance of RNSCs observed in the absence of *svb*.

We next investigated a putative influence of *svb* on RNSC differentiation, making use of the lineage-tracing system called ReDDM that has been recently developed for intestinal stem cells⁵⁰. Based on differences in the stability of two fluorescent proteins, ReDDM allows marking renal progenitors that express both mCD8::GFP and H2B::RFP, while their progeny only maintain the very stable H2B::RFP (see Fig. 4a). In control conditions, we detected rare H2B::RFP progeny (Fig. 4a) confirming a low rate of cell renewal in Malpighian tubules^{38,39} (Supplementary Figure 4b). Recent work has shown that the expression of *mir-8* (the fly homolog of *mir-200* in vertebrates) downregulates the expression of EMT-inducing factors *Escargot* and *Zfh1* (the homolog of *Zeb1*), triggering a strong burst of stem cell differentiation in the intestine⁵⁰. Similarly, we found that *mir-8* expression in RNSCs forced *esg+* cells to

differentiate and only rare RNSCs persisted after 8 days of treatment (Fig. 4a, Supplementary Figure 5). Upon *mir-8* expression, the progeny (H2B::RFP-positive, GFP-negative cells) of RNSCs present in lower tubules also expressed *Alkaline Phosphatase 4* confirming that the depletion of RNSCs upon *mir-8* overexpression was caused by their premature differentiation (Supplementary Figure 5). In contrast, no significant modification of the progenitors/progeny ratio was observed in *svb*-RNAi conditions when compared to controls, showing that *svb* depletion did not trigger RNSC differentiation (Fig. 4a, Supplementary Figure 5).

Finally, we tested whether *svb*-depleted RNSCs were lost because they underwent apoptosis. As a first step, we assayed consequences of blocking programmed cell death by expressing the viral caspase inhibitor p35⁵¹. Although the expression of p35 had no detectable effect by itself on RNSCs, it rescued the number of RNSCs when *svb* was depleted (Fig. 4b). Next, we stained for cells undergoing apoptosis using the Terminal deoxynucleotidyl transferase dUTP Nick End Labelling (TUNEL), an assay allowing the detection of apoptotic DNA fragmentation. While apoptotic figures were almost absent in control conditions, expression of the pro-apoptotic gene *reaper* induced a strong increase in the number of TUNEL-positive cells (Fig. 4c). Similarly, knocking down *svb* in stem cells led to frequent apoptotic figures in RNSCs (Fig. 4c) consistently with the progressive decrease in RNSC number observed throughout adult life (Fig. 2c).

Taken together, these data show that the loss of RNSCs observed upon *svb* loss of function is primarily due to stem cell death, indicating that a main role of *Svb* is to protect adult stem cells from undergoing apoptosis.

***Svb* acts downstream of Hippo.** Previous work has shown that the Hippo pathway is a key regulator of the *Drosophila* gut homeostasis, controlling proliferation and survival of stem cells for tissue regeneration^{52,53}. Since the Hippo pathway^{54,55} is a key sensor of various stresses renowned to induce apoptosis^{56,57}, we then investigated its function and putative interplay with *Svb* in the control of RNSC behavior.

The core Hippo complex is composed of two kinases, Hippo (Hpo) and Warts (Wts) and two scaffolding proteins, Salvador and Mob As Tumor Suppressor^{54,55}. Activation of Hippo leads to the phosphorylation of the co-transcription factor Yorkie (Yki), preventing its positive action on the transcription of target genes such as *DIAP1* and *bantam*, favoring resistance to apoptosis and proliferation, respectively^{54,55}. Consistently, we found that activation of the pathway via Hpo overexpression induced a strong reduction in the number of RNSCs (Fig. 5a, b). Similarly, knockdown of Yki resulted in dramatic RNSC loss (Supplementary Figure 3d and 6a). Conversely, increased Yki activity was sufficient to induce a remarkable increase in the number of RNSCs (Fig. 5), providing additional evidence of the proliferative potential of RNSCs. These data showing the role of the Hippo pathway in the regulation of RNSC survival/proliferation, we next assessed how it interacted with *Svb* function. Co-expression of *OvoB*—mimicking a constitutive activator form of *Svb* (see Fig. 3a)—together with Hpo was sufficient to rescue the loss of RNSCs (Fig. 5). These results therefore suggested that *Svb* was acting downstream of Hpo. Indeed, the strong increase in the number of renal stem cells observed following increased Yki activity was entirely suppressed upon simultaneous expression either of *svb*-RNAi, or of the constitutive repressor *OvoA* (Fig. 5). Quantification indicated that *esg+* cells overexpressing Yki were even more sensitive to *svb* loss-of-function than otherwise normal RNSCs (Fig. 5b). Of note, Yki over-proliferating cells showed

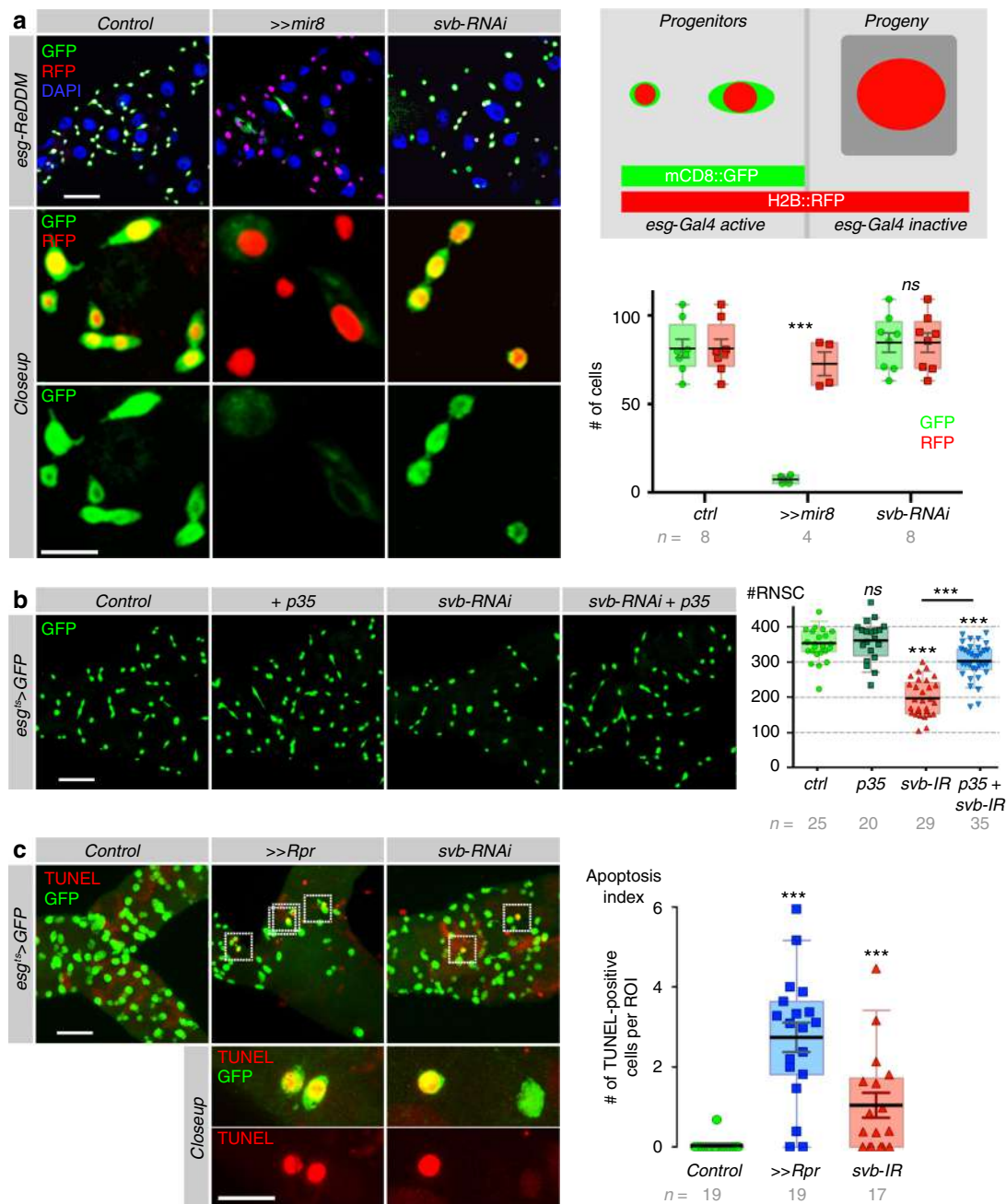


Fig. 4 *svb* protects RNSCs from apoptosis. **a** Lineage-tracing experiments (*esg*-ReDDM) at 8 days after induction. While stem cells (*esg*-positive) express both mCD8::GFP (green) and H2B::RFP (red), only the very stable H2B::RFP protein persists in their progeny (*esg* negative). Nuclei are in blue (DAPI). **b** Rescue of *svb*-depleted RNSCs by p35. *esg*^{ts} was used to drive the expression of indicated transgenes (together with GFP), during 8 days. Quantification of *esg*+ cells is shown at the right. Ctrl vs p35 ($p = 0.6613$), *svb*-RNAi ($p < 0.0001$), p35 + *svb*-RNAi ($p = 0.0006$), *svb*-RNAi vs p35 + *svb*-RNAi ($p < 0.0001$). **c** TUNEL assays (red) revealing the increase in RNSC apoptosis upon expression of the pro-apoptotic gene *reaper* (*Rpr*, $p < 0.0001$), or RNAi depletion of *svb* ($p = 0.0002$), driven by *esg*^{ts} > GFP (green). Renal stem cells undergoing apoptosis (green cytoplasm and red nuclei) are boxed, and better seen in closeups. The apoptotic index was calculated as the number of TUNEL-positive cells/GFP-positive cells ($\times 100$). Values are presented as average \pm standard error of the mean (SEM) and boxes with whiskers (10–90 percentile). p -Values are calculated using Mann-Whitney test (** $p < 0.001$). n = number of observed tubules. Scale bar is 30 μ m, except in closeup panels (10 μ m)

aberrant cell morphology reminiscent of tumors derived from intestine stem cells, which also display increased sensitivity to cell death when compared to normal stem cells⁵⁸, and that might explain why RNSCs with high Yki levels cannot survive upon Svb knockdown. Hence, the function of Yki in RNSCs requires Svb, suggesting that Svb was interacting with this nuclear effector of the Hippo pathway. Supporting this view, we found that RNSC survival could not be rescued either by expression of OvoB in the

absence of Yki (Supplementary Figure 6a), or by the over-expression of Yki in the absence of Svb (Fig. 5). In contrast, the artificial re-expression of a key target gene of Yki, *DIAP1*, was sufficient to compensate for *svb*-depletion, including in most extreme conditions, i.e., following 32 days of treatment (Figs. 2 and 5a).

In sum, both Svb and Yki are required for RNSC homeostasis, functionally interacting for the survival of adult stem cells. We

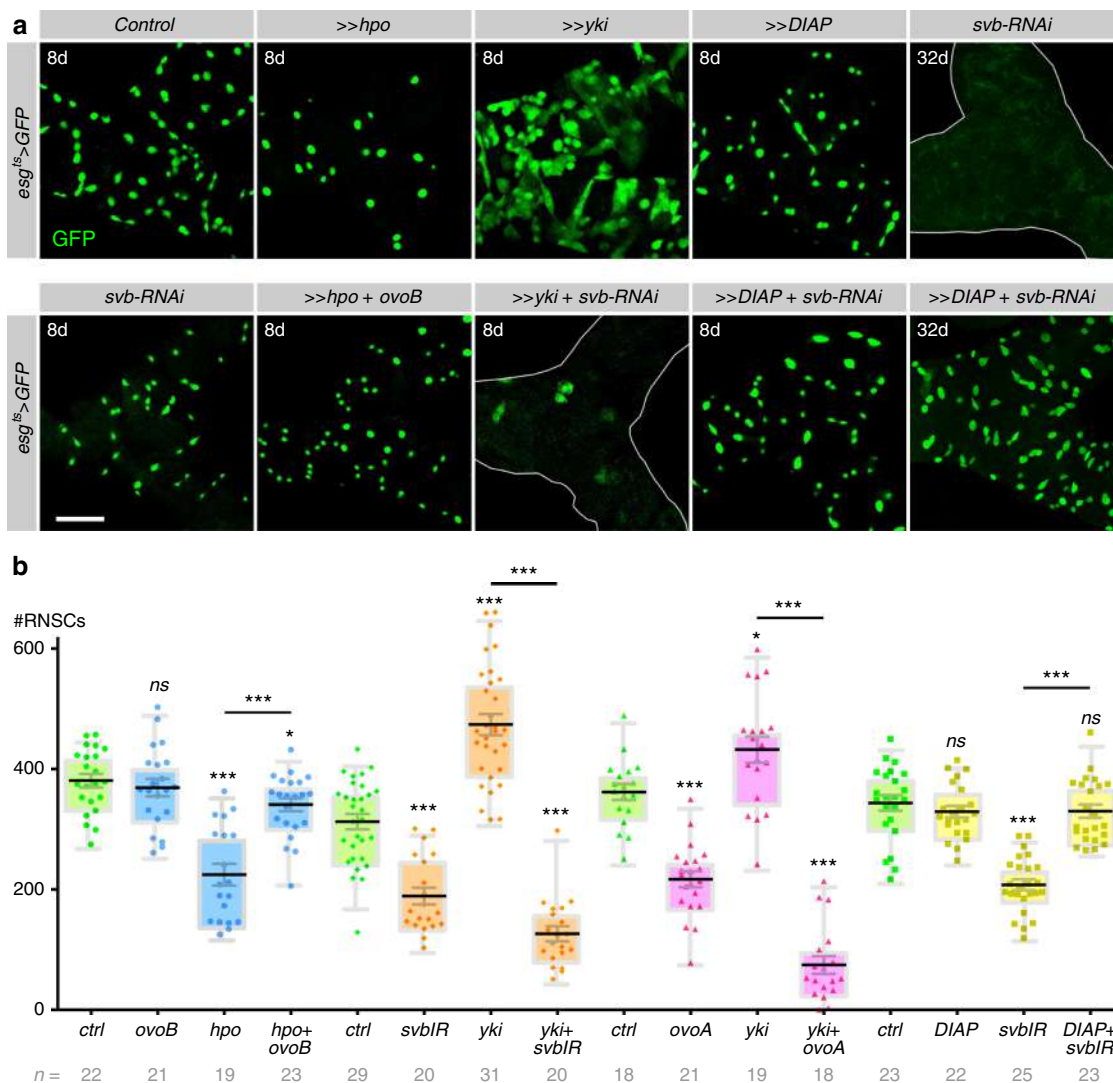


Fig. 5 *svb* is a member of the Hippo pathway. **a** Pictures of Malpighian tubules with *esg^{ts}*-driven expression of GFP (control) and indicated transgenes, at 8 days (8d) or 32 days (32d) after induction. **b** Quantification of *esg*-positive cells in corresponding genotypes. Values are presented as average ± standard error of the mean (SEM) and boxes with whiskers (10–90 percentile). Unless otherwise indicated, *p* values correspond to comparison with control conditions in a same series of experiments, as estimated by Mann–Whitney test (ns, *p* ≥ 0.05; **p* < 0.05; ***p* < 0.01; ****p* < 0.001). Scale bar is 30 μm

thus concluded that *Svb* acts downstream of Hippo cytoplasmic core components and, together with *Yki*, both nuclear factors are required to protect RNSCs from apoptosis.

***Svb* as a nuclear effector of the Hippo pathway.** Having established that *Svb* and *Yki* functionally interact, we sought to decipher the underlying molecular mechanisms and how these two regulators contribute to the control of relevant effector gene expression, e.g., *DIAP1*.

A first piece of evidence emerged from the comparison of in vivo chromatin immuno-precipitation (ChIP-seq) datasets between *Svb*¹⁴ and *Yki*⁵⁹. We found that *Svb* and *Yki* share >1,300 common genomic binding sites (Supplementary Figure 7a and Supplementary Table 1) and statistical tests established the significance of this overlap (Supplementary Figure 7b). Interestingly, co-binding of *Yki* was rare for the direct target genes of *Svb* identified in the epidermis^{12–14}, as illustrated by *shavenoid* or *dusky-like* that both lack *Yki* binding (Supplementary Figure 7c,d). In contrast, *Svb* was often bound to known *Yki* target genes, such as *bantam*, *piwi*, *fat*, or *nanos*⁶⁰ (Supplementary

Figure 7e, f). Importantly, ChIP-seq also revealed that *Svb* binds in vivo to an enhancer of *DIAP1* (Fig. 6a), previously identified as a *Yki* direct binding site^{61,62} and that drives specific expression in intestinal stem cells⁶³. We therefore tested if *Svb* might regulate *DIAP1* expression in adult RNSCs. Although weak in control conditions, we observed specific expression of the *DIAP1-LacZ* reporter in RNSCs, which was strongly enhanced upon *Yki* overexpression (Fig. 6b). This induction of *DIAP1* expression upon *Yki* overexpression was antagonized by *OvoA* (Fig. 6b). Similar results were obtained with the isolated *DIAP1-4.3-GFP*⁶² enhancer that contains binding sites for *Yki* and *Svb*. Although *OvoB* did not show significant influence (Supplementary Figure 6b), the expression of *DIAP1-4.3-GFP* was again enhanced by *Yki* overexpression and abrogated upon counteracting *Svb* activity (Fig. 6b). These results thus strengthen the conclusion that *Svb* and *Yki* functionally interact in RNSCs to prevent apoptosis, at least in part through promoting *DIAP1* expression.

The transcriptional activity of *Yki*/YAP/TAZ is well known to rely on the regulation of its subcellular localization between the cytoplasmic versus nuclear compartments. Assaying a putative

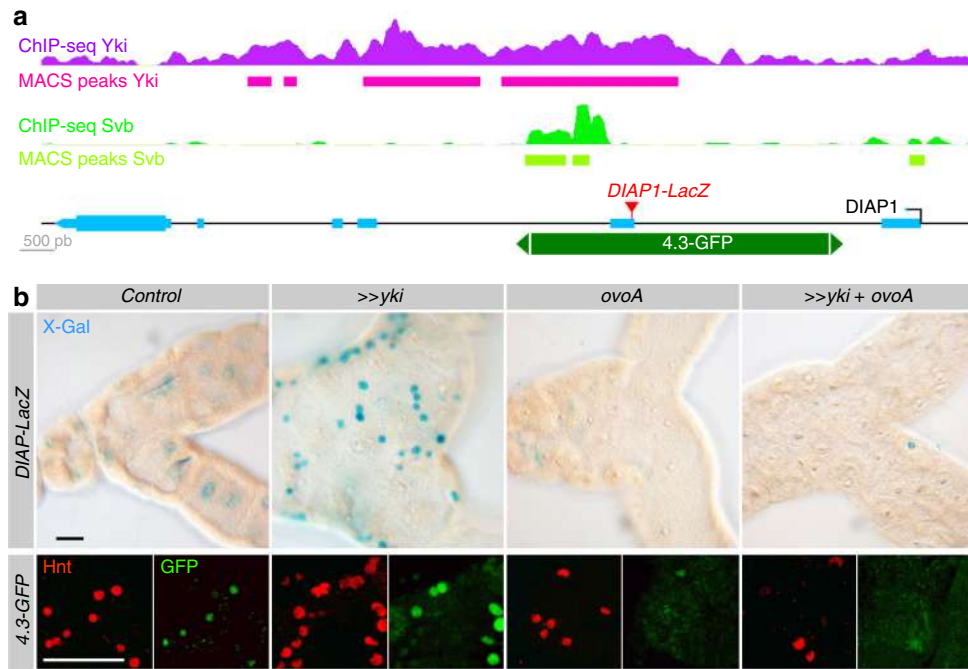


Fig. 6 Svb interacts with Yki to regulate *DIAP1* expression. **a** Drawing of the *DIAP1* locus. Exons are represented in cyan, the 4.3 enhancer in dark green and the insertion site of the *DIAP1-LacZ* reporter is in red. In vivo ChIP-seq profiles and MACS peaks bound by Svb and Yki are indicated in green and magenta, respectively. **b** *esg^{ts}* was used to drive the expression of *yki*, *svb^{REP}* (*ovoA*), and *yki* together with *svb^{REP}* in RNSCs. The expression of *DIAP1* was followed by the activity of *DIAP1-lacZ* (X-Gal staining, top panel). Expression of the 4.3-GFP enhancer of *DIAP1* (bottom panel) was followed by immuno-staining against GFP (green) and Hindsight (Hnt, red) revealing the influence of Yki and/or OvoA on RNSCs. Scale bar is 30 μ m

influence of Svb on Yki distribution, we did not however detect any consequences of Svb activity (either Svb^{ACT} or Svb^{REP}) on in vivo Yki nuclear accumulation within RNSCs (Supplementary Figure 6c). Yki is unable to bind DNA by itself and need to associate to sequence-specific transcription factors⁵⁵. Hence we assayed whether Yki could associate with Svb to explain the functional interactions observed in RNSCs between these two nuclear factors. Biochemical assays showed that Svb bound to Yki, as seen by reciprocal co-immunoprecipitation in protein extracts prepared from cultured cells (Fig. 7a). Both the full length repressor and matured Svb activator interacted with Yki (Fig. 7a). It has been shown that Yki contains two WW protein domains, mediating interaction with partners bearing PPxY motifs (such as Wts⁶¹ or Mad⁶⁴). Of note, we detected two PPxY motifs within the Svb protein, at position 523 (PPFY) and 881 (PPSY), i.e., within the region common to Svb repressor and activator forms. We found that the mutation of Yki WW motifs was sufficient to abrogate the interaction with both forms of Svb (Fig. 7a). Furthermore, point mutations of Svb PPxY motifs also impaired interaction with Yki, as monitored using immunoprecipitation assays (Fig. 7a). We next investigated the in vivo consequences of impairing Svb/Yki interaction, investigating the functional behavior of the Svb mutant lacking Yki interaction motifs (Svb-PPxY). When expressed in the embryonic epidermis, Svb-PPxY retained the ability to induce ectopic trichomes, consistent with a function of Svb being independent of Yki activity for the terminal differentiation of epidermal tissues (Supplementary Figure 8a). In contrast, Svb-PPxY behaved as a potent dominant negative mutant within adult stem cells. Previous work has shown that Svb^{ACT} and Svb^{REP} display distinct patterns of intra-nuclear distribution, characterized by diffuse nucleoplasmic localization or accumulation in dense foci, respectively¹⁷. Svb-PPxY::GFP displayed a discrete accumulation in foci, reminiscent of that observed for the repressor (Fig. 7b). Importantly, expression of Svb-PPxY in RNSCs significantly reduced their number, as

observed upon Svb or Yki knockdown (Fig. 7c). These data therefore indicate that Svb physically interacts with Yki (via PPxY and WW motifs, respectively) and impairing specifically this interaction alters adult stem cell behavior.

One important question was whether the interaction between Svb and the Hippo pathway also took place in other tissues. The function of Hippo has been initially described in larval imaginal discs, which give rise to most adult tissues⁶⁵ and Yki overexpression promotes cell proliferation in both wing and eye discs⁶¹. We tested Svb/Yki genetic interactions in the wing using *collier-Gal4* that drives expression in the medial (L3–L4) intervein region. *Yki* expression resulted in the expansion of this region due to tissue overgrowth (Fig. 8a). In contrast, *OvoA* leads to both a reduction of the L3–L4 region and the absence of epidermal trichomes. As in RNSCs, *OvoA* was epistatic to *Yki*, since the wing region expressing both *yki* and *ovoA* was smaller than in controls and lacked trichomes (Fig. 8a). As expected for a mutant unable to interact with Yki, we found that expression of Svb-PPxY did not affect wing cell proliferation. Additional evidence of interaction between the Hippo pathway and Svb/Pri came from genetic assays demonstrating the importance of relative levels of their activity. While lowering *wts* activity causes a wing overgrowth phenotype (Fig. 8b), we noticed that reducing *pri* activity suppressed this phenotype (Fig. 8b, c). Furthermore, *wts* mutation can lead to dramatic defects in wing development when introduced in a given genetic background, and decreasing *pri* dosage was sufficient to almost completely restore normal wings (Supplementary Figure 8b).

In the eye, overexpression of Yki using the *GMR-Gal4* driver promoted extra cell proliferation resulting in an increased eye size. Similar results were obtained following *pri* overexpression, and co-expressing *pri* and *yki* resulted in a synergistic eye growth (Supplementary Figure 9a). Northern blotting of RNAs extracted from adult heads revealed that *DIAP1* mRNA levels were increased following *pri* overexpression (Supplementary Figure 9b), while there was no effect on *yki* or *cycE* mRNA.

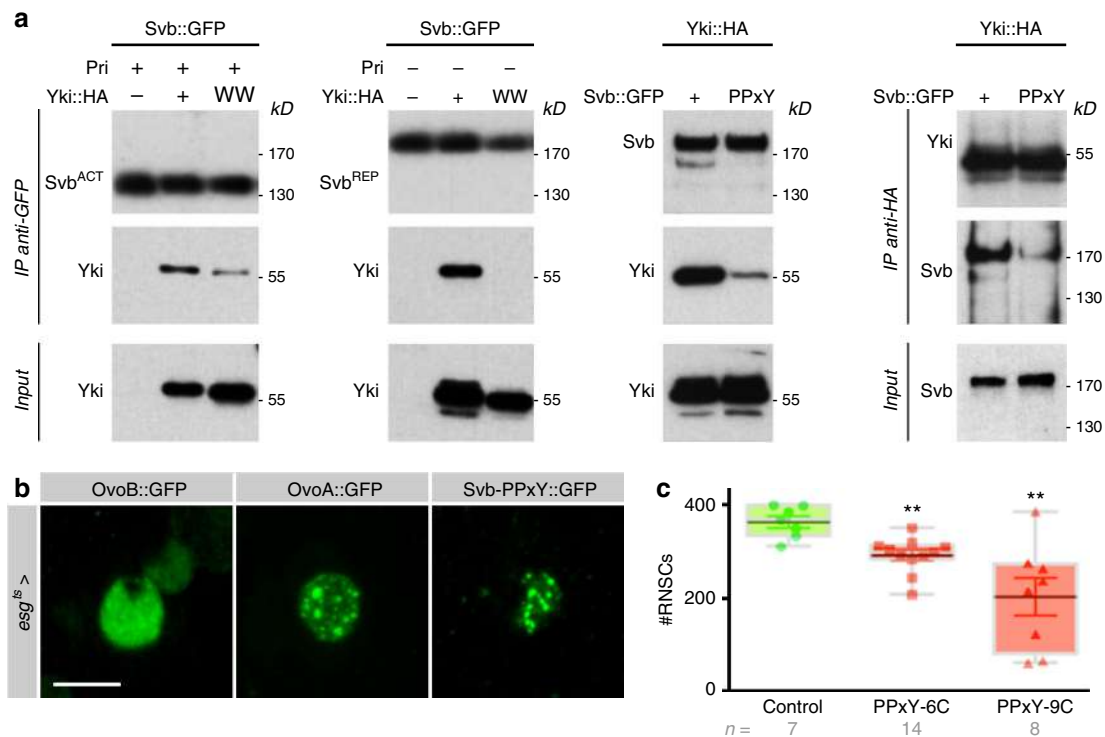


Fig. 7 The Svb and Yki proteins associate to maintain RNSCs. **a** Co-immuno-precipitation of Svb with Yki. Svb::GFP and Yki::HA were expressed in S2 cells, with or without Pri peptides, and protein extracts were immuno-precipitated using anti-GFP (left panels) or anti-HA antibodies (right panel). Point mutations substituting the WW domains of Yki (YkiWW::HA), or the PPxY motifs of Svb (Svb-PPxY::GFP), prevent Svb/Yki physical interaction. **b** Intranuclear distribution of the constitutive activator (OvoB), or repressor (OvoA) forms of Svb within RNSCs. Svb-PPxY displays a pattern of nuclear foci distribution, evoking that of the dominant negative form of Svb (OvoA). **c** Consequences of Svb-PPxY expression in adult RNSCs (driven by *esg^{ts}-Gal4*). Two different genomic insertions allowing the expression of Svb-PPxY::GFP (#6C and #9C) decrease the number of renal stem cells ($p = 0.0059$ and 0.0023). Values are presented as average \pm standard error of the mean (SEM) and boxes with whiskers (10–90 percentile), p -values from Mann-Whitney test (** $p < 0.01$). Scale bar is $3 \mu\text{m}$

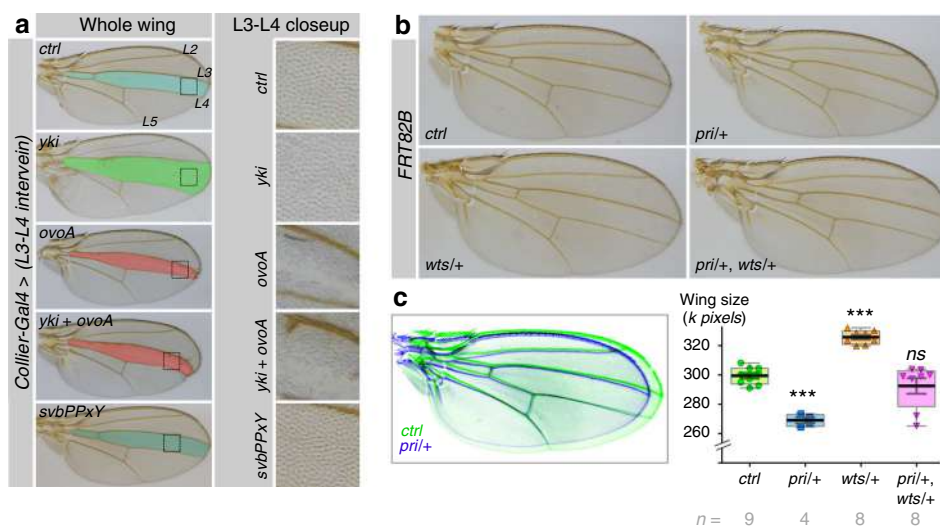


Fig. 8 Svb/Pri functionally interact with Yki in adult wings. **a** Adult wings expressing the indicated transgenes under the control of *col-Gal4*, a driver active in the L3-L4 region. **b** Genetic interaction between *pri^{S18.1}* and *wts^{X1}* alleles. **c** Overlay between *control* (green) and *pri-/+* (blue) wings (left), and quantification of wing area in the different conditions (right). Values are presented as average \pm standard error of the mean (SEM) and boxes with whiskers (10–90 percentile). p -Values from Mann-Whitney tests are *pri/+* (< 0.0001), *wts/+* (< 0.0001), *pri/+*, *wts/+* ($= 0.4234$) (ns, $p \geq 0.05$; ** $p < 0.01$; *** $p < 0.001$)

We interpret these results to imply that Svb functionally interacts with Yorkie, both in adult stem cells and in specific developing tissues, to regulate a subset of transcriptional targets of the Hippo pathway, including the activation of *DIAP1* expression.

Discussion

Our results show that Shavenbaby is expressed and required for the maintenance of adult renal stem cells in flies, supporting the conclusion that the OvoL/Svb family of transcription factors plays a key and evolutionarily-conserved role in the behavior of progenitors/stem cells.

The role of Svb in adult stem cell maintenance in flies requires both a fine control of its expression and of its transcriptional activity. *Svb* expression in RNSCs involves at least two separable enhancers, driving similar expression patterns. *Svb* was one of the first cases to reveal the functional importance of apparently redundant (or shadow) enhancers in the phenotypic robustness of regulatory networks^{5,6} across tissues and development stages¹¹. Our data suggest that a similar *cis*-regulatory architecture is also underlying the control of adult stem cells.

RNCS maintenance further requires a proper post-translational maturation of the Svb protein, in response to Pri smORF peptides. During both embryonic¹⁷ and post-embryonic development^{18,19}, the main role of Pri peptides is to provide a temporal control of Svb activity, conveying systemic steroid signaling¹⁹. It is therefore possible that Pri smORF peptides also connect genetic networks to hormonal control for the regulation of adult stem cells. Recent work has shown that various smORF peptides contribute to the regulation of developmental pathways, muscle formation and physiology, etc...^{15,66,67}, and our findings extend their influence to the control of adult stem cells. It has been proposed that the emerging field of smORF peptides may open innovative therapeutic strategies^{16,68}, for example peptidomimetic drugs, which might also be of interest for regenerative medicine.

Our results establish that a main function of Svb in adult stem cells is mediated by a functional interplay with the Hippo pathway, well established for its roles in the control of adult stem cells^{52–54}. Our results indicate that Svb behaves as a nuclear effector of Hippo, relying on a direct interaction with Yorkie in order to protect stem cells from apoptosis, at least in part through the regulation of *DIAP1* expression. Analysis of genome-wide binding events further suggests that the Svb/Yki interaction is involved in the control of a broader set of Hippo-regulated genes, including during development. Since both Pri and Ubr3 are also essential for the survival of adult stem cells, it is interesting to note that Ubr3 protects the *DIAP1* protein from degradation⁶⁹, and direct binding of Ubr3 on the activated form of *DIAP1* is elicited in the presence of Pri peptides¹⁸. Therefore, in addition to the control of *DIAP1* expression (via Svb), Ubr3 and Pri could also stabilize the *DIAP1* protein to protect stem cells from apoptosis. Although initially restricted to TEAD transcription factors, the number of Yorkie (YAP/TAZ) nuclear partners is rapidly growing⁵⁵. Interestingly, recent work has shown a direct interaction of YAP/TAZ with the pro-EMT factors Snail/Slug, in the control of stem cell renewal and differentiation⁷⁰. As previously reported for intestinal stem cells^{42,43,50}, we show that pro-EMT regulators are also required for preventing premature differentiation of renal stem cells. While pro-EMT and OvoL factors are often viewed as antagonistic factors^{20,25}, *in vivo* studies in *Drosophila* stem cells show that they both contribute to their maintenance, Svb/Yki preventing their apoptosis and EMT factors their differentiation. Since many studies have implicated the Hippo pathway, pro-EMT and OvoL/Svb factors in various

tumors, new insights into their functional interactions in adult stem cells may provide additional knowledge directly relevant to understand their connections in human cancers.

Methods

Drosophila work. Flies were cultured (unless otherwise noted) at 25 °C, using standard cornmeal food (per liter: 17 g inactivated yeast powder, 80 g corn flour, 9 g agar, 45 g white sugar, and 17 ml of Moldex). Similar results were also observed using a richer medium (same composition except 45 g of yeast powder). Female adult flies were used in all analyses throughout the study and placed on fly food supplemented with fresh yeast, which was changed every 2 days. Conditional expression in RNCSs was achieved by maintaining *tub-Gal80^{ts}* expressing flies at 18 °C, until adulthood. Eclosed females aged for 3-days were shifted to 29 °C for induction of gene expression and were kept at 29 °C for the indicated period of time (in most cases 8 days). Virgin females bearing *svb^{R9}*, *svb^{PL107}*, or *ubr3^B* mutations^{18,47} over FM0 balancers were mated with males of the following genotype: *y, w, hs-FLP, tub-Gal80, FRT19A; UAS::mcd8-GFP; tub-Gal4/TM6B, Tb*. Females of the correct phenotype (not *B* and not *Tb*) were heat shocked for 1 h at 37 °C. For G-TRACE experiments, females *w; UAS-RedStinger, UAS-FLP, Ubi-p63E (FRT.STOP)Stinger/CyO* were crossed with *esg-Gal4, dome-MESO-Gal4* or *c507-Gal4* males and raised at 18 °C until adulthood. Females were kept at 25 °C and dissected after 20 days. Flies were transferred on fresh food every 2 days and dissected at the indicated time. Detailed information about the genotype of each *Drosophila* stock is given in Supplementary Information.

Histology. Tissues were dissected in 1× PBS, fixed in 4% formaldehyde in PBS for 15 min at room temperature, washed for 5 min in PBS containing 0.1% Triton X-100 (PBT) and fixed again during 20 min. Following a 5 min wash in PBT, tissues were blocked for 30 min in PBT containing 1% BSA. Primary antibodies were incubated overnight at 4 °C. Anti-β-Galactosidase (Cappel Cat#08559761) antibody was used at 1:1000, anti-Cut (DSHB Cat#2B10), anti-GFP (Acris Antibodies Cat#TP401) at 1:200, anti-phospho-Histone H3 (Upstate Cat#TAKMA312B), anti-Disc-large (DSHB Cat#4F3 anti-discs large) at 1:500, anti-Arm (DSHB Cat# N2 7A1) at 1:100, anti-Hnt (DSHB Cat# 1G9) at 1:30. AlexaFluor-488 or -555 secondary antibodies (ThermoFischer Scientific Cat#Z25302 and Cat#Z25205) were incubated for 2 h at room temperature at 1:500. After three washes, tissues were mounted in Vectashield (Vector Laboratories). For X-gal staining, adult tissues were dissected in 1× PBS, fixed in 1% glutaraldehyde in PBS for 15 min at room temperature and washed in PBS. The staining solution was warmed up at 37 °C for 10 min plus another 10 min after addition of 8% X-Gal (5-bromo-3-indolyl-β-D-Galactopyranoside). The X-Gal solution used to reveal the β-Galactosidase activity was: 10 mM NaH₂PO₄·H₂O/Na₂HPO₄·2H₂O (pH = 7.2), 150 mM NaCl, 1 mM MgCl₂·6H₂O, 3.1 mM K₄ (FeII(CN)₆), 3.1 mM K₃ (FeIII (CN)₆), 0.3% Triton X-100. Bright-field pictures were acquired using a Nikon Eclipse 90i microscope.

TUNEL assays. TUNEL staining was performed following a protocol kindly provided by A. Bergmann. Tissues were dissected in 1× PBS, fixed in 4% formaldehyde in PBS for 30 min at room temperature, washed for 3 times 15 min in PBS containing 0.3% Triton X-100 (PBT) and blocked 15 min in PBT + 5% Normal Goat Serum (PBNT). Anti-GFP antibody was incubated overnight at 4 °C. The following day, tissues were washed with WB (50 mM Tris/HCl pH = 6.8, 150 mM NaCl, 0.5% NP40, 1 mg/ml BSA) and incubated overnight with the secondary antibody. Tubules were then washed 3 times in WB and incubated 30 min at 65 °C with 99 μl of 100 mM Na-Citrate + 1 μl of 10% Triton X-100. Following quick washes with WB, tubules were incubated in 45 μl of labelling solution (in situ cell death detection TMR Red kit, Roche), for 30 min at 37 °C. 5 μl of the enzyme solution was added for 2 h at 37 °C. Tubules were finally washed in PBT and mounted in Vectashield.

Microscopy, image, and statistical analysis. Images of whole Malpighian tubules were acquired on a LSM710 confocal scanning microscope (×20 objective), using automated multi-position scan. After stitching, tiled images of individual pairs of tubules were analyzed with IMARIS 8.0 to quantify the number of GFP-positive cells. Data of at least three independent experiments (approx. 20 tubules) were combined. All statistical analyses were carried out using Prism 5 (GraphPad).

Comparisons between normally distributed groups were carried out using Student's *t*-test, unpaired, two-tailed and incorporating Welch's correction to account for unequal variances. One-way ANOVA with Bonferroni correction was used when multiple comparisons were applied. In all figures, ns indicates $p \geq 0.05$, * indicates $0.05 > p \geq 0.01$, ** indicates $0.01 > p \geq 0.001$, and *** indicates $p < 0.001$. Close-up pictures were acquired using Leica SPE and Leica SP8 confocal laser scanning microscopes (×40 and ×63 objectives). Laser intensity and background filtering was set according to the control samples and remained the same for all subsequent samples. The intensity of pictures has been enhanced equally for all images within the same experiment using adjustments in Photoshop CS5. All images were processed using Adobe Photoshop and Illustrator CS5.

Western blotting and immunoprecipitation. *Drosophila* S2 cells (Drosophila Genomics Resource Center, stock number: 6) were grown in Schneider medium supplemented with 10% fetal calf serum and 1% penicillin/streptomycin (Invitrogen) at 25 °C. We used either S2 cells or stable cell lines co-expressing the copper-inducible constructs pMT-Svb::GFP and pMT-pri¹⁷. S2 cells were cultured in six-well plates (1.75 × 10⁶ cells/3 ml) and transfected in 100 µl of Opti-MEM (Invitrogen) with 3 µl of FugeneHD (Promega) and the indicated constructs. CuSO₄ (final concentration of 1 mM) was used to induce the expression of pMT plasmids. The following plasmids were used: pAc-Yki::HA and its related mutated version pAc-Yki-WW::HA, pAc-Svb::GFP, and pAc-Svb-PPxY::GFP. S2 cells were lysed on ice in 250 µl of ice-cold lysis buffer (150 mM NaCl, 50 mM Tris/HCl pH = 8.0), 0.5% NP40, 1 mM EGTA, and protease inhibitors (Roche). After clearing by centrifugation at 18,000g for 10 min, immuno-precipitations from transfected lysates were done in lysis buffer, using anti-GFP antibody (GFP-trap, Chromotek), or anti-HA antibody (Covance) and Fast-Flow Protein G-Sepharose (Sigma #P3296). Immuno-precipitated samples were separated by SDS-PAGE and transferred to PVDF membranes, then blotted by using the following antibodies: anti-GFP (Acris Antibodies, Cat#TP401, 1:10,000), anti-HA (Covance, Cat#BLE901513, 1:2000), and anti-Yki (1:4000, kind gift of K. Irvine). Secondary antibodies were anti-mouse or anti-rabbit IgG-HRP conjugates (Jackson Laboratory, Cat#515-035-003, 1:10,000), detected using the ECL Clarity (Bio-Rad). Uncropped scans of blots are supplied in Supplementary Information.

Northern blot analysis. Two days-old adults were frozen with liquid nitrogen and heads were sorted with sieves, followed by RNA purification with Isogen (Nippon Gene). 1 µg of RNA per lane was separated by formaldehyde-agarose gel electrophoresis and then transferred to a nylon membrane (Roche). Hybridization and wash procedures were carried out at 52 and 65 °C, respectively. The filters were reacted with an alkaline phosphatase-conjugated anti-DIG antibody (Roche) and chemiluminescent reactions with CPD-Star (Roche) were detected by LAS 4000mini (GE Healthcare).

Data availability

The ChIP-seq datasets analyzed during the current study are available in the EMBL-EBI repository: <https://www.ebi.ac.uk/biosamples/samples/SAMEA2439952>, <https://www.ebi.ac.uk/biosamples/samples/SAMN01041482>, <https://www.ebi.ac.uk/biosamples/samples/SAMN02231205>, <https://www.ebi.ac.uk/biosamples/samples/SAMN02231213>, <https://www.ebi.ac.uk/biosamples/samples/SAMN02231206>. The data that support the findings of this study are given in Supplementary Table 1. Detailed scripts and computer code used for ChIP-seq analysis are available from the corresponding author on request.

Received: 17 August 2017 Accepted: 12 November 2018

Published online: 30 November 2018

References

- Mevel-Ninio, M., Terracol, R., Salles, C., Vincent, A. & Payre, F. ovo, a *Drosophila* gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning. *Mech. Dev.* **49**, 83–95 (1995).
- Payre, F., Vincent, A. & Carreno, S. ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. *Nature* **400**, 271–275 (1999).
- Dai, X. et al. The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. *Genes Dev.* **12**, 3452–3463 (1998).
- Wells, J. et al. Ovol2 suppresses cell cycling and terminal differentiation of keratinocytes by directly repressing c-Myc and Notch1. *J. Biol. Chem.* **284**, 29125–29135 (2009).
- Crocker, J. et al. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).
- Frankel, N. et al. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).
- Frankel, N. et al. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474**, 598–603 (2011).
- McGregor, A. P. et al. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* **448**, 587–590 (2007).
- Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* **424**, 935–938 (2003).
- Preger-Ben, N. E., Davis, F. P. & Stern, D. L. Evolved repression overcomes enhancer robustness. *Dev. Cell* **39**, 572–584 (2016).
- Preger-Ben, N. E. et al. Comprehensive analysis of a cis-regulatory region reveals pleiotropy in enhancer function. *Cell Rep.* **22**, 3021–3031 (2018).
- Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F. & Plaza, S. Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biol.* **4**, e290 (2006).
- Fernandes, I. et al. Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes. *Dev. Cell* **18**, 64–76 (2010).
- Menoret, D. et al. Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome Biol.* **14**, R86 (2013).
- Zanet, J., Chanut-Delalande, H., Plaza, S. & Payre, F. Small peptides as newcomers in the control of *Drosophila* development. *Curr. Top. Dev. Biol.* **117**, 199–219 (2016).
- Plaza, S., Menschaert, G. & Payre, F. In search of lost small peptides. *Annu. Rev. Cell Dev. Biol.* **33**, 391–416 (2017).
- Kondo, T. et al. Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science* **329**, 336–339 (2010).
- Zanet, J. et al. Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* **349**, 1356–1358 (2015).
- Chanut-Delalande, H. et al. Pri peptides are mediators of ecdysone for the temporal control of development. *Nat. Cell Biol.* **16**, 1035–1044 (2014).
- Watanabe, K. et al. Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by Ovol2 transcriptional repressor. *Dev. Cell* **29**, 59–74 (2014).
- Roca, H. et al. Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer. *PLoS One* **8**, e76773 (2013).
- Roca, H. et al. A bioinformatics approach reveals novel interactions of the OVOL transcription factors in the regulation of epithelial–mesenchymal cell reprogramming and cancer progression. *BMC Syst. Biol.* **8**, 29 (2014).
- Wang, Z. H. et al. Ovol2 gene inhibits the epithelial-to-mesenchymal transition in lung adenocarcinoma by transcriptionally repressing Twist1. *Gene* **600**, 1–8 (2017).
- Ricketts, C. J. et al. Genome-wide CpG island methylation analysis implicates novel genes in the pathogenesis of renal cell carcinoma. *Epigenetics* **7**, 278–290 (2012).
- Jia, D. et al. OVOL guides the epithelial-hybrid-mesenchymal transition. *Oncotarget* **6**, 15436–15448 (2015).
- Li, S. & Yang, J. Ovol proteins: guardians against EMT during epithelial differentiation. *Dev. Cell* **29**, 1–2 (2014).
- Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).
- Jolly, M. K. et al. Coupling the modules of EMT and stemness: a tunable ‘stemness window’ model. *Oncotarget* **6**, 25161–25174 (2015).
- Wu, R. S. et al. OVOL2 antagonizes TGF-beta signaling to regulate epithelial to mesenchymal transition during mammary tumor metastasis. *Oncotarget* **8**, 39401–39416 (2017).
- Lapan, S. W. & Reddien, P. W. Transcriptome analysis of the planarian eye identifies ovo as a specific regulator of eye regeneration. *Cell Rep.* **2**, 294–307 (2012).
- Lee, B. et al. Transcriptional mechanisms link epithelial plasticity to adhesion and differentiation of epidermal progenitor cells. *Dev. Cell* **29**, 47–58 (2014).
- Kitazawa, K. et al. OVOL2 maintains the transcriptional program of human corneal epithelium by suppressing epithelial-to-mesenchymal transition. *Cell Rep.* **15**, 1359–1368 (2016).
- Kim, J. Y. et al. OVOL2 is a critical regulator of ER71/ETV2 in generating FLK1+, hematopoietic, and endothelial cells from embryonic stem cells. *Blood* **124**, 2948–2952 (2014).
- Beyenbach, K. W., Skaer, H. & Dow, J. A. The developmental, molecular, and transport biology of Malpighian tubules. *Annu. Rev. Entomol.* **55**, 351–374 (2010).
- Denholm, B. Shaping up for action: the path to physiological maturation in the renal tubules of *Drosophila*. *Organogenesis* **9**, 40–54 (2013).
- Sözen, M. A., Armstrong, J. D., Yang, M., Kaiser, K. & Dow, J. A. Functional domains are specified to single-cell resolution in a *Drosophila* epithelium. *Proc. Natl. Acad. Sci. USA* **94**, 5207–5212 (1997).
- Garayo, M., Villaro, A. C. & Sesma, P. Myoendocrine-like cells in invertebrates: occurrence of noncardiac striated secretory-like myocytes in the gut of the ant *Formica polyctena*. *Gen. Comp. Endocrinol.* **95**, 133–142 (1994).
- Singh, S. R., Liu, W. & Hou, S. X. The adult *Drosophila* malpighian tubules are maintained by multipotent stem cells. *Cell Stem Cell* **1**, 191–203 (2007).
- Xu, K., Liu, X., Wang, Y., Wong, C. & Song, Y. Temporospatial induction of homeodomain gene cut dictates natural lineage reprogramming. *eLife* **7**, e33934 (2018).
- Takashima, S., Paul, M., Aghajanian, P., Younossi-Hartenstein, A. & Hartenstein, V. Migration of *Drosophila* intestinal stem cells across organ boundaries. *Development* **140**, 1903–1911 (2013).
- Micchelli, C. A. & Perrimon, N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* **439**, 475–479 (2006).

42. Loza-Coll, M. A., Southall, T. D., Sandall, S. L., Brand, A. H. & Jones, D. L. Regulation of *Drosophila* intestinal stem cell maintenance and differentiation by the transcription factor Escargot. *EMBO J.* **33**, 2983–2996 (2014).
43. Korzelius, J. et al. Escargot maintains stemness and suppresses differentiation in *Drosophila* intestinal stem cells. *EMBO J.* **33**, 2967–2982 (2014).
44. Jiang, H. et al. Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. *Cell* **137**, 1343–1355 (2009).
45. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker (MARCM) for *Drosophila* neural development. *Trends Neurosci.* **24**, 251–254 (2001).
46. Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A. & Couso, J. P. Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* **5**, e106 (2007).
47. Delon, I., Chanut-Delalande, H. & Payre, F. The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. *Mech. Dev.* **120**, 747–758 (2003).
48. Andrews, J. et al. OVO transcription factors function antagonistically in the *Drosophila* female germline. *Development* **127**, 881–892 (2000).
49. Yang, M. Y., Wang, Z., MacPherson, M., Dow, J. A. & Kaiser, K. A novel *Drosophila* alkaline phosphatase specific to the ellipsoid body of the adult brain and the lower Malpighian (renal) tubule. *Genetics* **154**, 285–297 (2000).
50. Antonello, Z. A., Reiff, T., Esther, B.-I. & Dominguez, M. Robust intestinal homeostasis relies on cellular plasticity in enteroblasts mediated by *miR-8-Escargot* switch. *EMBO J.* **34**, 2025–2041 (2015).
51. Hay, B. A., Wolff, T. & Rubin, G. M. Expression of baculovirus P35 prevents cell death in *Drosophila*. *Development* **120**, 2121–2129 (1994).
52. Shaw, R. L. et al. The Hippo pathway regulates intestinal stem cell proliferation during *Drosophila* adult midgut regeneration. *Development* **137**, 4147–4158 (2010).
53. Staley, B. K. & Irvine, K. D. Warts and yorkie mediate intestinal regeneration by influencing stem cell proliferation. *Curr. Biol.* **20**, 1580–1587 (2010).
54. Pan, D. The hippo signaling pathway in development and cancer. *Dev. Cell* **19**, 491–505 (2010).
55. Staley, B. K. & Irvine, K. D. Hippo signaling in *Drosophila*: recent advances and insights. *Dev. Dyn.* **241**, 3–15 (2012).
56. Sansores-Garcia, L. et al. Modulating F-actin organization induces organ growth by affecting the Hippo pathway. *EMBO J.* **30**, 2325–2335 (2011).
57. Gaspar, P., Holder, M. V., Aerne, B. L., Janody, F. & Tapon, N. Zyxin antagonizes the FERM protein expanded to couple f-actin and yorkie-dependent organ growth. *Curr. Biol.* **25**, 679–689 (2015).
58. Ma, M. et al. Wildtype adult stem cells, unlike tumor cells, are resistant to cellular damages in *Drosophila*. *Dev. Biol.* **411**, 207–216 (2016).
59. Oh, H. et al. Genome-wide association of Yorkie with chromatin and chromatin-remodeling complexes. *Cell Rep.* **3**, 309–318 (2013).
60. Zhang, C. et al. The ecdysone receptor coactivator Taiman links Yorkie to transcriptional control of germline stem cell factors in somatic tissue. *Dev. Cell* **34**, 168–180 (2015).
61. Huang, J., Wu, S., Barrera, J., Matthews, K. & Pan, D. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the *Drosophila* homolog of YAP. *Cell* **122**, 421–434 (2005).
62. Zhang, L. et al. The TEAD/TEF family of transcription factor Scalloped mediates Hippo signaling in organ size control. *Dev. Cell* **14**, 377–387 (2008).
63. Poernbacher, I., Baumgartner, R., Marada, S. K., Edwards, K. & Stocker, H. *Drosophila* Pez acts in Hippo signaling to restrict intestinal stem cell proliferation. *Curr. Biol.* **22**, 389–396 (2012).
64. Oh, H. & Irvine, K. D. Cooperative regulation of growth by Yorkie and Mad through bantam. *Dev. Cell* **20**, 109–122 (2011).
65. Tapon, N. et al. Salvador Promotes both cell cycle exit and apoptosis in *Drosophila* and is mutated in human cancer cell lines. *Cell* **110**, 467–478 (2002).
66. Payre, F. & Desplan, C. Small peptides control heart activity. *Science* **351**, 226–227 (2016).
67. Pueyo, J. I., Magny, E. G. & Couso, J. P. New peptides under the s(ORF)ace of the genome. *Trends Biochem. Sci.* **41**, 665–678 (2016).
68. Couso, J. P. & Patraquim, P. Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol.* **18**, 575–589 (2017).
69. Huang, Q. et al. Ubr3 E3 ligase regulates apoptosis by controlling the activity of DIAP1 in *Drosophila*. *Cell Death Differ.* **21**, 1961–1970 (2014).
70. Tang, Y., Feinberg, T., Keller, E. T., Li, X.-Y. Y. & Weiss, S. J. {Snail/Slug} binding interactions with {YAP/TAZ} control skeletal stem cell self-renewal and differentiation. *Nat. Cell Biol.* **18**, 917–929 (2016).

Acknowledgements

We are grateful to FlyBase, Bloomington and Vienna stock centers, Developmental Studies Hybridoma Bank, *Drosophila* Genomics Resource Center, supported by NIH grant 2P40OD010949, as well as N. Tapon, J. Colombani, K.F. Harvey, D.J. Pan, J. Dow, H. Skaer, M. Crozatier, K. Irvine, A. Bergmann, and M. Dominguez, for providing flies, antibodies, and molecular reagents. We thank B. Ronsin (Toulouse RIO Imaging) for help with microscopy, A. Alsawadi, A. Dib, and M. Soulard for experimental support. We also thank all members of the Payre lab for critical reading of the manuscript. This work was supported by ANR (Chrononet), Fondation pour le Recherche Médicale (Equipe FRM DEQ20170336739), Agence Universitaire de la Francophonie (PCSI AUF-BMO) for D.O. and F.P., PRES IDEX Université de Toulouse for C.P., and by MEXT KAKENHI (JP26113006) for Y.K. J.B. was supported by fellowships from “Ministère de l’Enseignement et de la Recherche” and “La Ligue contre le Cancer”. S.A.H. and A.M.-F. were supported by fellowships from “FRM” and “Ligue contre le Cancer”, respectively.

Author contributions

C.P., S.P., Y.K., D.O., and F.P. conceived and directed the project, following initial observations made by D.O. J.B. and C.P. carried out most of the experiments presented here, and other experiments were conducted by S.A.H. (RNAi), J.Z. (immunoprecipitations), P.V. (molecular biology and transgenics), K.A., Y.Y., and S.I. (adult eyes), and H. C.-D. (genetics). A.M.-F. analyzed NGS data. J.B., C.P., K.A., Y.Y., S.I., H.C.-D., S.P., Y. K., D.O., and F.P. analyzed data and contributed to their interpretation. J.B., C.P., and F. P. prepared the figures and wrote the manuscript. All authors helped to write the paper.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-07569-0>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Chapitre 5 Les smORF peptides Pri sont de médiateurs de la signalisation de l'ecdysone, contribuant à façonner les réponses spatio-temporelles

5.1 Contexte de la publication

Les mécanismes de régulation de l'expression de *shavenbaby* (*svb*) ont été particulièrement bien étudiés à travers la dissection des enhancers (Chanut-Delalande et al., 2006; Menoret et al., 2013; Stern et al., 2013). Nous avons ainsi une bonne compréhension de la manière dont l'expression spatiale de *svb* est façonnée par les voies de signalisation Wg et notamment EGF. La caractérisation de la régulation post-traductionnelle de l'activité de Svb par les petits peptides Pri pose la question de la régulation temporelle de ce processus qui doit être régulé pour assurer la formation des trichomes à un instant bien défini du développement. Cette question a été adressée dans l'équipe par les travaux effectués par Hélène Chanut-Delalande (Chanut-Delalande et al., 2014). La réalisation d'un crible génétique permettant l'identification de gènes dont la perte de fonction affecte la formation des trichomes a permis l'identification du gène *EstG14* dont le produit participe à la synthèse de l'hormone stéroïdienne Ecdysone. Hormone bien connue pour contrôler les mues et la métamorphose de la drosophile (Yamanaka et al., 2012 a).

Chez la Drosophile, l'ecdysone est une hormone stéroïdienne synthétisée par la glande prothoracique à partir du cholestérol contenu dans la nourriture. Elle est synthétisée sous forme d'un précurseur qui subit des modifications post-traductionnelle par des enzymes codées par les gènes de la famille « halloween » (nom donné du fait de l'apparence fantomatique de l'exosquelette des jeunes larves mutantes pour ces gènes). Une fois synthétisée, l'ecdysone est libérée dans l'hémolymphe et transformée en 20-hydroxy-ecdysone (20E), sa forme active. Cette forme est capable de se lier à son récepteur hétérodimérique EcR/Usp (Ecdysone Receptor et Ultraspinacl) (Hill et al., 2012 b) pour activer l'expression de ses gènes cibles. Cette activation requiert le recrutement de corégulateur comme les remodeleurs de la chromatine Nucleosome Remodelling Factor (NURF), Trithorax Related (TRR) méthyltransférase ou encore CBP. Les cellules activées par l'ecdysone vont alors enclencher leur programme de différenciation tissus suite à des cascades de récepteurs nucléaires. L'expression des gènes cibles de l'ecdysone est finement régulée dans le temps. L'ecdysone est produite par pulsion à des temps précis du développement correspondant aux transitions d'un stade à une autre (Dubrovsky, 2005). De plus en effet, il est bien établi que certains gènes répondent précocement à l'ecdysone alors que d'autres répondent plus tardivement. Ce modèle de régulation temporel est connu sous le nom de modèle Ashburner (Hill et al., 2012 b) (cf. Figure 35).

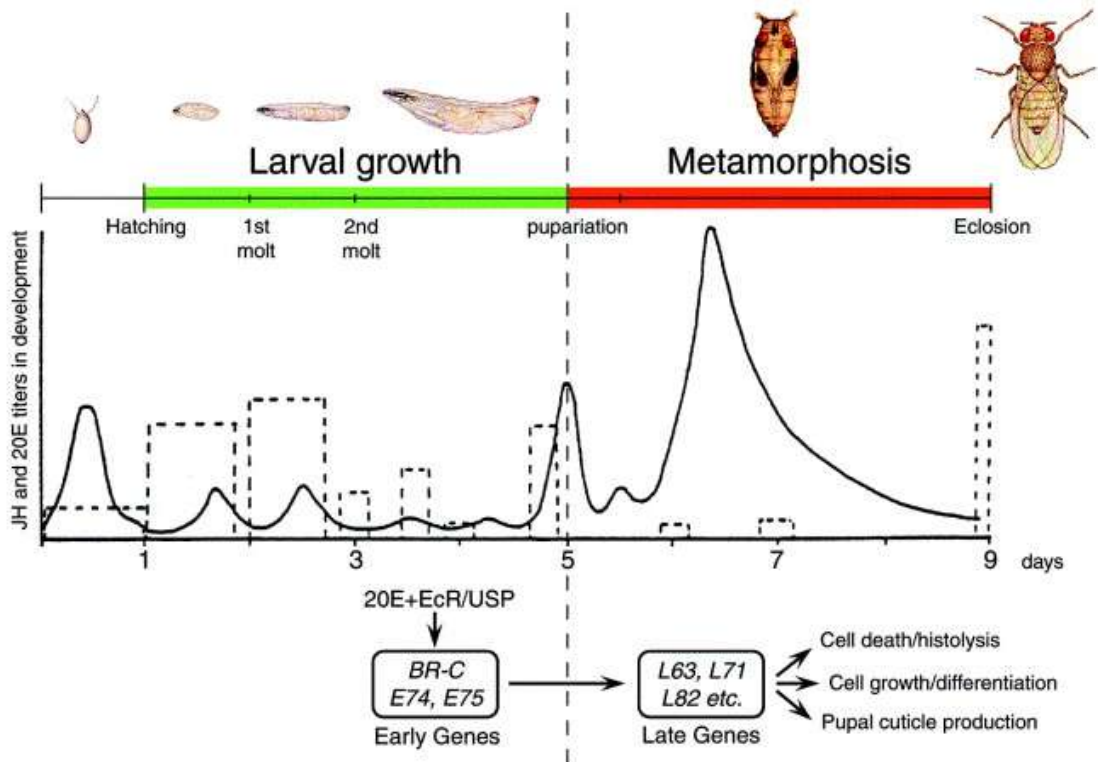


Figure 35 : L'hormone ecdysone et ses pulsions au cours du cycle de vie de la *Drosophile*.
(Dubrovsky, 2005)

Il a été montré que l'ecdysone est impliquée dans l'autophagie et l'apoptose de nombreux tissus lors de la morphogénèse comme les grandes salivaires, l'intestin et le corps gras (235). La synthèse de l'ecdysone s'effectue en réponse à différents stimuli comme la lumière (232), la nutrition et la croissance tissulaire (Colombani et al., 2005; Delanoue et al., 2010). Cette synthèse se fait sous l'influence de la voie insuline (Malita et al., 2021) et de signaux cellulaires autonomes comme le métabolisme du fer (Pan et al., 2021).

A l'aide de données de CHIP-seq permettant de suivre la fixation de EcR et de l'analyse de lignées rapportrices couvrant le locus du gène *pri*, Hélène a pu identifier différents enhanceurs contrôlant l'expression de ce gène dans l'épiderme (Chanut-Delalande et al., 2014). Elle a montré que l'expression de *pri* dans l'épiderme est dépendante de l'enhancer nommé G dont l'activité est régulée par l'Ecdysone.

Ainsi l'activité de *Svb* est régulée spécialement par les voies de signalisation Wnt et EGF et temporellement par l'Ecdysone à travers la régulation de *pri* pour assurer la formation des trichomes.

5.2 Résumé de la publication

Durant sa thèse, Azza Dib a poursuivi l'analyse de la régulation de l'expression de *pri* par l'Ecdysone. Le gène *pri* est exprimé dans divers organes et tissus au cours du développement en plus de l'épiderme embryonnaire comme les trachées (appareil respiratoire de la mouche) et les appendices (pattes et ailes). Cela soulevait la question de la dépendance de l'expression de *pri* dans ces différents tissus quant à sa dépendance à la voie Ecdysone. Cette étude a permis de montrer que *pri* est fortement activé en réponse à

l'Ecdysone et cela dans de nombreux types de lignées cellulaires de drosophile (S2, Kc, BG3...). Cela suggère que *pri* est un composant générique de cette voie de signalisation. Bien que l'ARN permettant la traduction des peptides Pri soit relativement court (1,5kb), ses éléments de cis-régulation sont dispersés sur 50kb et permettent l'expression de *pri* dans plusieurs tissus et cela à des temps de développement différents. L'analyse poussée de certains de ces enhanceurs a permis de montrer que certains d'entre eux comme pri-G sont directement régulés par EcR à différents stades développementaux.

En conclusion, cette étude montre que les peptides Pri ont un rôle essentiel dans la signalisation ecdysone et que sa régulation spacio-temporelle s'effectue par de nombreux enhanceurs au cours du développement ayant des fonctions pléiotropes et possiblement redondantes.

5.2.1 Contribution

Pour ce projet j'ai utilisé les données de transcriptomique produites par (240). Dans cette étude, le transcriptome de 41 lignées cellulaires de *Drosophila melanogaster*, en présence ou en absence d'ecdysone a été établi. Du fait de sa petite ORF (cadre de lecture), le niveau d'expression de *pri* n'est pas détecté par les méthodes classiques de comptage. Je me suis donc attachée à suivre la modulation de son expression dans ces différentes conditions expérimentales. La détection de *pri* nécessite un alignement des données brutes sur le génome de référence (*Drosophila melanogaster* dm6) (132), et comptage des reads correspondants à la région d'intérêt (3R:13,813,109 ..13,814,648) et cela pour chaque condition. Cette étape a été réalisée grâce à samtools view (H. Li et al., 2009).

L'étude statistique a été réalisée de deux façons, soit avec EdgeR (Robinson et al., 2010) par moi-même, soit avec DEseq2 (243) par un groupe de quatre étudiants-ingénieurs de l'INSA Toulouse spécialité Bio-informatique que j'ai co-encadré pendant quatre mois. Ce groupe d'étudiants devait développer une chaîne de traitement (pipeline) leur permettant d'obtenir le tableau de comptage pour réaliser l'étude statistique. J'avais déjà mis en place ce pipeline de mon côté et cet exercice leur permettait d'appréhender la difficulté de comptage des petits ORF. Puis, à partir du tableau de comptage obtenu, ils ont réalisé l'approche statistique de détermination de gènes différentiellement exprimés avec DEseq2. Ils ont ensuite comparé leurs résultats avec ceux que j'avais obtenus avec EdgeR. Cet exercice permet d'appréhender les différences entre les deux suites de logiciels.

Les deux méthodes statistiques bien que différentes arrivent à la même observation que les peptides Pri sont fortement exprimés en présence d'ecdysone alors qu'ils sont faiblement exprimés en son absence et cela dans de nombreux types cellulaires différents.

De plus, pour certaines lignées (Stoiber et al., 2016) les expériences de RNA-seq ont été réalisées à différents temps d'exposition à l'ecdysone. En utilisant le même « pipeline » d'analyse des données de transcriptomique et de détection de *pri* sur ces cellules, nous avons pu observer que l'expression de *pri* commence au bout de sept heures d'exposition. A titre de comparaison, celles des gènes *Eip75B* et *Hr4*, identifiés comme des gènes de réponse précoces à l'Ecdysone se produit en quatre heures.

L'ensemble de ces analyses convergent et nous permettent de conclure que *pri* est un gène canonique de la voie de signalisation Ecdysone exprimé tardivement.



OPEN ACCESS

Edited by:

Joseph Rothnagel,
The University of Queensland,
Australia

Reviewed by:

Robert Kenichi Maeda,
Université de Genève, Switzerland
Shalini Sharma,
University of Arizona, United States

***Correspondence:**

Hélène Chanut-Delalande
helene.chanut@univ-tlse3.fr
François Payre
francois.payre@univ-tlse3.fr

† These authors have contributed
equally to this work and share last
authorship

‡ Present address:

Serge Plaza,
Laboratoire de Recherche en
Sciences Végétales, CNRS, UPS,
Toulouse INP, Université de Toulouse,
Toulouse, France

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 24 May 2021

Accepted: 28 July 2021

Published: 30 August 2021

Citation:

Dib A, Zanet J,
Mancheno-Ferris A, Gallois M,
Markus D, Valenti P,
Marques-Prieto S, Plaza S,
Kageyama Y, Chanut-Delalande H
and Payre F (2021) Pri smORF
Peptides Are Wide Mediators
of Ecdysone Signaling, Contributing
to Shape Spatiotemporal Responses.
Front. Genet. 12:714152.
doi: 10.3389/fgene.2021.714152

Pri smORF Peptides Are Wide Mediators of Ecdysone Signaling, Contributing to Shape Spatiotemporal Responses

Azza Dib¹, Jennifer Zanet¹, Alexandra Mancheno-Ferris¹, Maylis Gallois¹, Damien Markus¹, Philippe Valenti¹, Simon Marques-Prieto¹, Serge Plaza^{1‡}, Yuji Kageyama^{2,3}, Hélène Chanut-Delalande^{1*†} and François Payre^{1*†}

¹ Molecular, Cellular and Developmental Biology Department (MCD), Centre de Biologie Intégrative (CBI), CNRS, UPS, University of Toulouse, Toulouse, France, ² Department of Biology, Graduate School of Science, Kobe University, Kobe, Japan, ³ Biosignal Research Center, Kobe University, Kobe, Japan

There is growing evidence that peptides encoded by small open-reading frames (sORF or smORF) can fulfill various cellular functions and define a novel class regulatory molecules. To which extend transcripts encoding only smORF peptides compare with canonical protein-coding genes, yet remain poorly understood. In particular, little is known on whether and how smORF-encoding RNAs might need tightly regulated expression within a given tissue, at a given time during development. We addressed these questions through the analysis of *Drosophila polished rice* (*pri*, a.k.a. *tarsal less* or *mille pattes*), which encodes four smORF peptides (11–32 amino acids in length) required at several stages of development. Previous work has shown that the expression of *pri* during epidermal development is regulated in the response to ecdysone, the major steroid hormone in insects. Here, we show that *pri* transcription is strongly upregulated by ecdysone across a large panel of cell types, suggesting that *pri* is a core component of ecdysone response. Although *pri* is produced as an intron-less short transcript (1.5 kb), genetic assays reveal that the developmental functions of *pri* require an unexpectedly large array of enhancers (spanning over 50 kb), driving a variety of spatiotemporal patterns of *pri* expression across developing tissues. Furthermore, we found that separate *pri* enhancers are directly activated by the ecdysone nuclear receptor (EcR) and display distinct regulatory modes between developmental tissues and/or stages. Alike major developmental genes, the expression of *pri* in a given tissue often involves several enhancers driving apparently redundant (or shadow) expression, while individual *pri* enhancers can harbor pleiotropic functions across tissues. Taken together, these data reveal the broad role of Pri smORF peptides in ecdysone signaling and show that the *cis*-regulatory architecture of the *pri* gene contributes to shape distinct spatial and temporal patterns of ecdysone response throughout development.

Keywords: smORF peptides, polished rice, tarsal less, mille pattes, ecdysone, steroids

INTRODUCTION

Recent advances in functional genomics indicate that apparently non-coding RNAs often encode smORF peptides (Andrews and Rothnagel, 2014; Pauli et al., 2015; Pueyo et al., 2016) and *Polished rice (pri)* represents a paradigm of this emerging field (Andrews and Rothnagel, 2014; Zanet et al., 2016). *Pri*, also known as *tarsal less* or *mille pattes*, was initially identified as a long non-coding RNA that exhibits highly dynamic expression during embryogenesis (Inagaki et al., 2005; Tupy et al., 2005). Further work has demonstrated that *pri* acts through the production of evolutionarily conserved peptides, encoded by smORFs. In *Drosophila*, *pri* encodes four smORF peptides (from 11 to 32 aa), all bearing the conserved LDPTGQY motif. Across insect species, *pri* is involved in multiple developmental processes and plays key roles in the production and patterning of posterior structures (Savard et al., 2006; Galindo et al., 2007; Kondo et al., 2007; Ray et al., 2019). Partial loss-of-function mutations have revealed the role of *pri/tarsal less* in the formation of adult appendages, mutant animals displaying atrophic legs with missing/fused distal segments (Galindo et al., 2007; Pueyo and Couso, 2008; Ray et al., 2019). The complete loss of *pri* function yet results in embryonic lethality; *pri* mutants display severe defects of the tracheal respiratory system (Kondo et al., 2007, 2010; Ozturk-Colak et al., 2016). Another prominent phenotype of *pri* embryos is a deep alteration of epidermal development, including the absence of trichomes (Galindo et al., 2007; Kondo et al., 2007). The insect body is decorated by a stereotyped pattern of epidermal cell extensions, called denticles or hairs, and collectively referred to as trichomes (Payre, 2004). Previous work has established the pivotal role of a transcription factor, Ovo/Shavenbaby (Svb), whose expression defines the trichome pattern (Payre et al., 1999; Delon et al., 2003; Stern and Frankel, 2013; Crocker et al., 2015). Svb induces the expression of cellular effectors that reorganize the cytoskeleton (Menoret et al., 2013), extracellular matrix (Fernandes et al., 2010), as well as cuticle composition (Andrew and Baker, 2008) or pigmentation (Chanut-Delalande et al., 2006, 2012). Pri peptides are essential for trichome formation, since they induce a posttranslational conversion of Svb from a large-sized repressor to a shorter activator, which is required to trigger the expression of trichome effectors (Kondo et al., 2010). Pri peptides act through their binding to an E3 ubiquitin-ligase, called Ubr3, which allows the recruitment of the Ubcd6/Ubr3 complex on the Svb N-terminal region. Ubiquitination of the Svb N-terminal region leads to limited proteasome degradation of the Svb repressor domain, ultimately releasing the processed Svb activator protein (Zanet et al., 2015). Thus, the program of trichome differentiation is set up with the accumulation of the Svb repressor and kept on hold until *pri* is expressed.

Accumulated evidences support that the developmental function of this sophisticated mechanism of *pri* action on Svb activity is to provide a strict temporal control of epidermal differentiation, in response to the systemic ecdysone hormone. Indeed, epidermal cells need to secrete a new cuticle exoskeleton before each molt (Moussian, 2010) and during the ultimate round of epidermal differentiation that occurs during metamorphosis for adult epidermal derivatives. Periodic pulses of ecdysone, the

main steroid hormone in insects, provide timing cues well-known to trigger larval molting and metamorphosis (Yamanaka et al., 2013). Ecdysone is synthesized from dietary cholesterol *via* a series of enzymatic reactions leading to 20-hydroxyecdysone (20E) (Petryk et al., 2003), which binds to, and activates, the nuclear ecdysone receptor (EcR), often dimerizing with another nuclear receptor ultraspiracle (Usp) (Yao et al., 1992; Hall and Thummel, 1998; Ghbeish et al., 2001). The mechanisms of the response to ecdysone have been extensively studied at the onset of metamorphosis, which relies on a temporal series of transcriptional regulation (Thummel, 2001; Ou and King-Jones, 2013; Cheate Jarvela and Pick, 2017). In the first step, 20E-bound EcR directly activates the expression of early genes that comprise Eip75B and broad-complex (Br) TFs (Thummel, 2001; Ou and King-Jones, 2013). Then, 20E/EcR acts together with early TFs to trigger the expression of additional tiers of early-late TFs, e.g., Hr4 and Hr3, activating, in turn, the expression of late genes (which are no longer under direct CONTROL OF EcR); inhibitory feedback loops ensuring tight temporal regulation (see **Figure 1A**). Ecdysone is also critical during embryonic development (Bender et al., 1997; Kozlova and Thummel, 2003), and EcR-responsive TFs are expressed in late embryos, following a similar temporal cascade (Ruaud et al., 2010). Consistently, we found that *pri* epidermal expression during both embryogenesis and metamorphosis is activated in response to ecdysone (Chanut-Delalande et al., 2014). Recent studies further show that *pri* expression is also controlled by ecdysone in the embryonic tracheal cells (Taira et al., 2021), as well as in the adult stem cells that regenerate the gut (Al Hayek et al., 2021) and kidneys (Bohère et al., 2018) to ensure homeostasis of vital functions across adulthood.

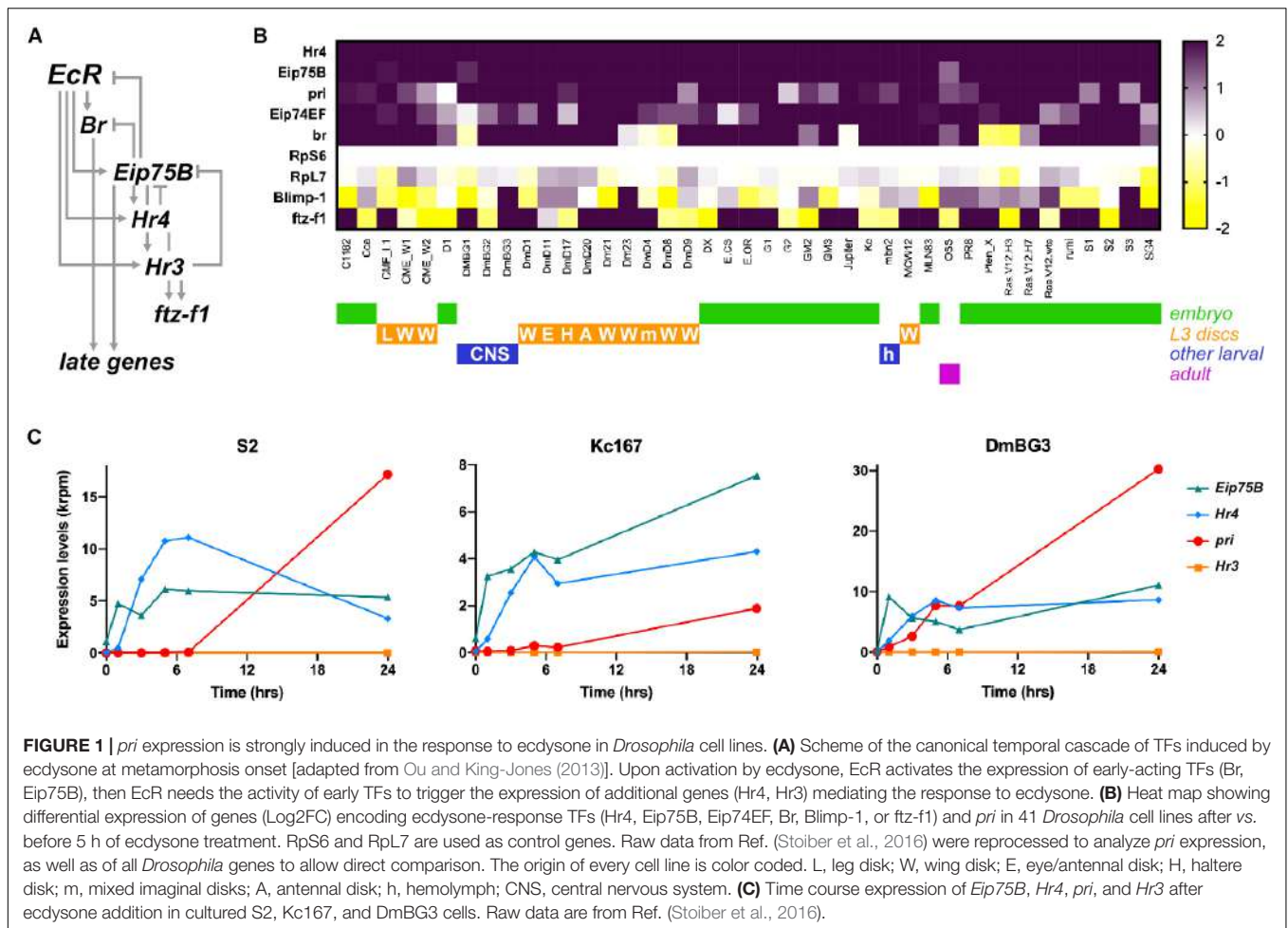
Here we show that Pri peptides further act as widespread mediators of ecdysone signaling across embryonic and post-embryonic development. Analysis of genome-wide transcriptional response to ecdysone reveals that the induction of *pri* expression is a hallmark of most cell types, when assayed *ex vivo*. *In vivo*, *pri* displays specific patterns of expression in various ecdysone-responsive tissues, during embryogenesis, larval/pupal stages, and in the adult. Genetic assays show that *pri* function relies on a large genomic region, which harbors multiple tissue- and stage-specific enhancers. Focusing on the embryonic epidermis and the adult leg primordium, we identified two *pri* enhancers whose activity is directly controlled by EcR, using distinct binding sites across tissues and between developmental stages.

MATERIALS AND METHODS

Fly Stocks

Most mutants were obtained from the Bloomington *Drosophila* Stock Center¹. The strains we used were *spo[1]/TM3,Dfd-YFP*, *phm[E7]/FM7a*, *Dfd-YFP*, *UAS-EcRDN* (Cherbas et al., 2003), *pri1* and *pri3* mutants (Kondo et al., 2007), *UAS-white-RNAi* (BL28980), *UAS-EcR-RNAi* (BL#50712, BL#29374, BL58286),

¹<http://flystocks.bio.indiana.edu/>



Ptc-Gal4, *en-Gal4*, *Dll-Gal4*, *UAS-pri* (Kondo et al., 2007), *UAS-mRFP* (BL#30567), and *tubGal80^{ts}* (BL#7017).

DNA Constructs and Transgenic Lines

LacZ reporter constructs were produced by inserting 5- to 6-kb DNA fragments of the *pri* genomic region into the pAttB-LacZ (Menoret et al., 2013) vector. GFP reporters correspond to the same *pri* regulatory regions inserted into the pS3AG vector (gift from T. Williams). All constructs were verified by sequencing. Transgenic lines, including BAC constructs [from P(acman) Resources²], were generated using the PhiC31 system, and all inserted at the 86F position (BestGene). BAC transgenic lines were recombined with *pri*¹ mutants. *w*⁺ candidate recombinants were screened for *pri*¹ mutation by PCR using a primer located in the promoter region of *pri* and a primer in the residual P element present in the *pri*¹ deficiency. Additional details are provided in **Supplementary Tables 1,2**.

The function of EcR was depleted by expressing *UAS-EcRDN* driven by *ptc-Gal4* in the embryonic epidermis, and *UAS-EcRDN* driven by *dll-Gal4* and *UAS-EcRDN* or *UAS-EcR-RNAi* driven by *en-gal4*, *UAS-mRFP*; *tubGal80^{ts}* in larval/pupal legs.

²<http://flypush.imgen.bcm.tmc.edu/pacmanfly/>

Analysis of ChIP-seq and RNA-seq Data

RNA-seq raw data (GSE11167) from Stoiber et al. (2016) were downloaded and reprocessed for bioinformatic and statistical analysis. Reads were mapped on the *Drosophila* genome (release r6.13) using STAR (Dobin et al., 2013) (v 2.5.2b, default parameters). Read count was done with HTseq-count (Anders et al., 2015) (v0.6.0, -t exon -r pos -i gene_symbol). We determined counts for *pri* with SAMtools (Li et al., 2009) (v 1.11), using values corresponding to the transcribed region of *pri* (samtools view "3R:13813109-13814648"). Statistical analyses were performed with edgeR (Robinson et al., 2010), using negative binomial generalized log linear model to the normalized (TMM) read count for each gene (glmLRT). Using R, we next calculated log₂FC for either one sample for most cell lines, or by averaging values when two samples or more were available. Log₂FC was considered when expression levels were ≥5 in control conditions. Heat map and plots were drawn using Prism 8.

Embryo, Larval, and Pupal Staining

Embryos were dechorionated by bleach treatment, fixed in heptane saturated in paraformaldehyde for 20 min, devitellinized with heptane/methanol, and stored in methanol. Homozygous

mutant embryos were identified by the absence of balancer chromosome (marked with GFP/YFP or LacZ). Sibling controls and mutant embryos were processed in the same batch; a typical collection includes >300 embryos in total. Staining was performed as previously described (Fernandes et al., 2010) using anti- β gal (1/400), biotinylated goat anti-rabbit or anti-mouse (1/1,000, Vector Laboratories) preincubated with streptavidin-HRP (Vector Laboratories), and revelation was performed with DAB (3,3'-diaminobenzidine, Sigma). For phalloidin staining, embryos were treated with 80% ethanol instead of methanol to preserve actin organization. We used TRITC-phalloidin (Sigma) and AlexaFluor-488 (1/1,000, Molecular Probes) secondary antibodies. *In situ* hybridization to *pri* mRNA was processed using a DIG-labeled RNA antisense probe synthesized *in vitro* and reacted with alkaline phosphatase-conjugated anti-DIG antibody (Roche) as described previously (Chanut-Delalande et al., 2006).

For experiments with imaginal disks, expression of UAS constructs was induced at the L3 stage for 24 h at 29°C to inactivate *Gal80^{ts}*. Imaginal disks from mid L3 larvae and pupae (between 0 and 4 h APF) were dissected in PBS and fixed in 4% paraformaldehyde/PBS at room temperature for 20 min. Tissues were rinsed three times in PBS for 10 min and permeabilized in PAT (1 × PBS, 0.3% Triton 100X, BSA 0.3%) for 1 h. For LacZ reporters, tissues were incubated with anti- β gal (1:1,000, Torrey Pines) overnight at 4°C in PAT, then rinsed in PAT and incubated for 2 h with Alexa-488 anti-rabbit secondary antibody (1:1,000, Jackson ImmunoResearch). For GFP reporters, we directly observed the fluorescence of GFP. Nuclei were stained with TOPRO or DAPI, and samples were further dissected and mounted in Vectashield (Vectorlabs). Images were acquired with a SP8 Leica confocal microscope.

Cuticle Preparation and Rescuing Assays

To assay the rescuing ability of BAC transgenes, *pri³/TM3-Dfd-LacZ* individuals were crossed by *pri¹,BACxyz(mw+)/TM3-Dfd-LacZ* recombinants. To score for trichome formation, embryos were genotyped using X-Gal staining, and cuticles were prepared in Hoyers/lactic acid (1/1). Each rescuing experiment has been performed, independently, at least three times. To score for full rescue up to viable adults, the progeny of *pri³/TM3-Dfd-LacZ* X *pri¹,BACxyz(mw+)/TM3-Dfd-LacZ* crosses was analyzed and the percentage of rescue estimated by the ratio between numbers of *w+* individuals (rescued) and *w+/TM3* siblings.

RESULTS

Pri Is a Widespread Ecdysone Response Gene

To investigate the role of *pri* in the response to ecdysone, we first analyzed RNA-seq data profiling the transcriptional changes induced by ecdysone, across a broad set of cell lines from various origins (Stoiber et al., 2016). Each out of the 41 analyzed cell lines displays large-scale response to ecdysone, with several hundred

of genes being differentially expressed upon ecdysone treatment (Stoiber et al., 2016). Unexpectedly, only a few genes exhibit general ecdysone response across cell lines, including *Hr4* and *Eip75B* that are activated in all cell lines, or *Eip74EF* and *broad* that are activated in most of them (Stoiber et al., 2016). *ftz-f1* displays more varying responses to ecdysone, while *Blimp-1* is often down regulated (Figure 1B).

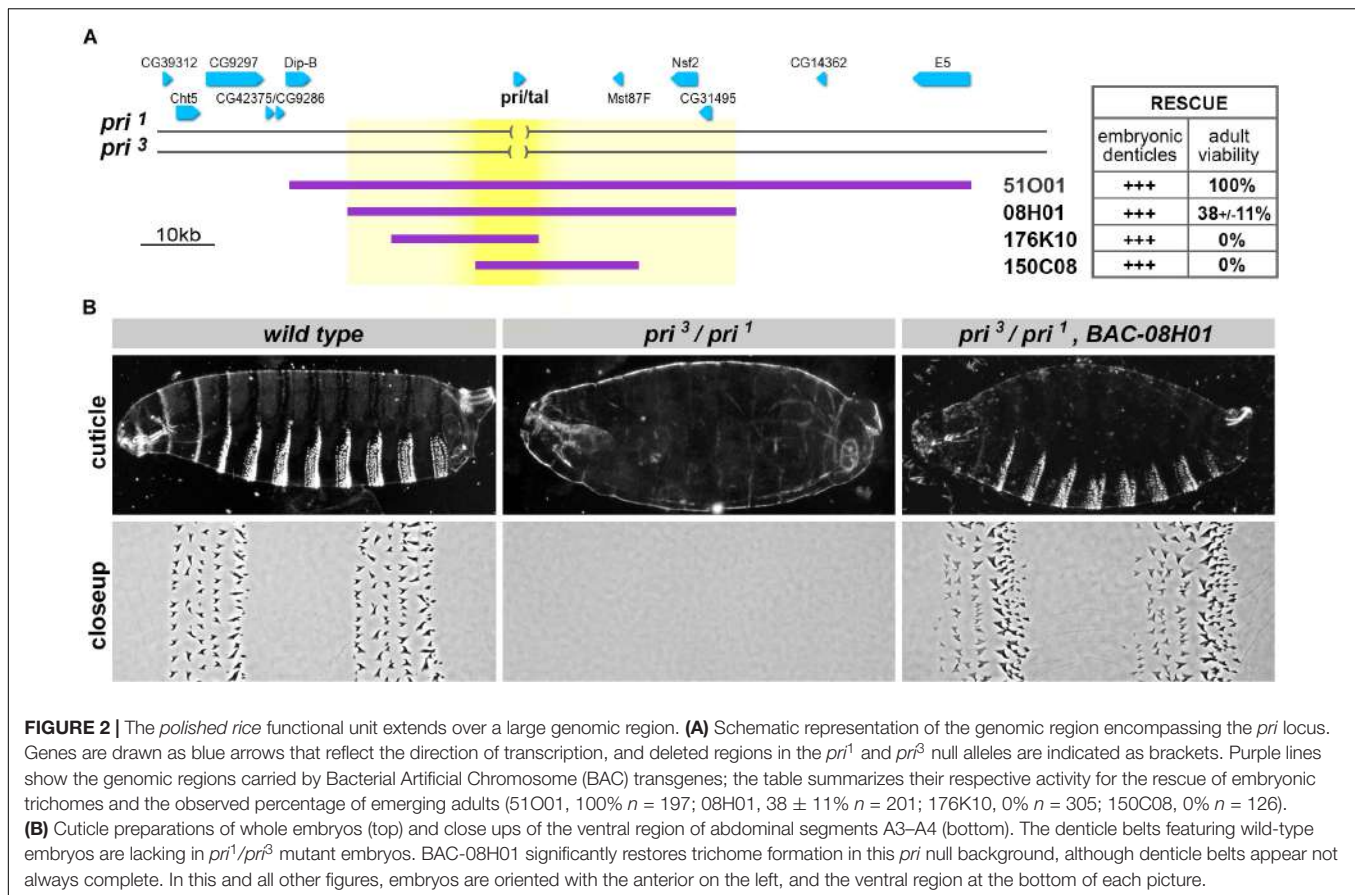
However, the study of Stoiber et al. (2016) did not report the expression of *pri*, probably since this atypical polycistronic gene is often mis-annotated and/or not properly taken into account by standard bioinformatic pipelines. We, thus, reanalyzed raw data and mapped sequence reads that match the transcribed region of *pri* (3R:13,813,109..13,814,648) in every cell line, with or without ecdysone treatment. The results show that *pri* is generally not expressed, or with barely detectable levels, prior to ecdysone addition. In contrast, *pri* levels are robustly induced upon ecdysone treatment ($\log_2FC > 1$) in the vast majority (35 out of 41) of cell lines, derived from embryos, larval imaginal disks or central nervous system (Figure 1B). In most cases, the activation of *pri* expression was very strong (24 lines with $\log_2FC > 2$), *pri* belonging to the top five most induced genes in eight cell lines (induction > 85×), and showing the strongest upregulation in DmD4 cells (>800×).

In addition, Stoiber et al. (2016) performed an extended time course analysis of the response to ecdysone in three cell lines (S2, Kc, and BG3-c2). Consistent with the temporal cascade that occurs during the larval/pupal transition, the induction of *Eip75* is seen soon after ecdysone addition, while *Hr4* induction starts slightly later and peaks around 7–10 h of treatment. The expression of *pri* starts later and keeps increasing to reach very high levels, as seen 24 h after ecdysone exposure (Figure 1C). These data, thus, suggest that *pri* expression requires both 20E-bound EcR and other early TFs, as expected for early-late response genes (Ou and King-Jones, 2013). It is, thus, likely that 5 h of treatment (the conditions used in all other cell lines) is not optimal to capture the induction of *pri* by ecdysone, suggesting that longer exposure times might reveal *pri* response in even broader contexts.

Together, these data indicate that *pri* expression is induced by ecdysone in a large variety of cell types. They further suggest that *pri* is a broad target of ecdysone signaling, representing a novel member of the core response to steroids.

Pri Function Relies on a Large Genomic Region

To delineate the regulatory landscape driving *pri* transcription, we next sought to define the extent of the genomic locus required for its function throughout development. The *pri* gene is transcribed as an intron-less polyadenylated RNA of 1.5 kb in length and separated from neighboring genes by large upstream and downstream intergenic regions (Figure 2A). We generated a series of transgenic lines bearing overlapping genomic constructs [selected from libraries of Bacterial Artificial Chromosomes (BAC) (Venken et al., 2009)] and assayed their rescuing activity when reintroduced in a genetic background lacking *pri* function. We used a *trans* allelic combination of two null *pri* alleles



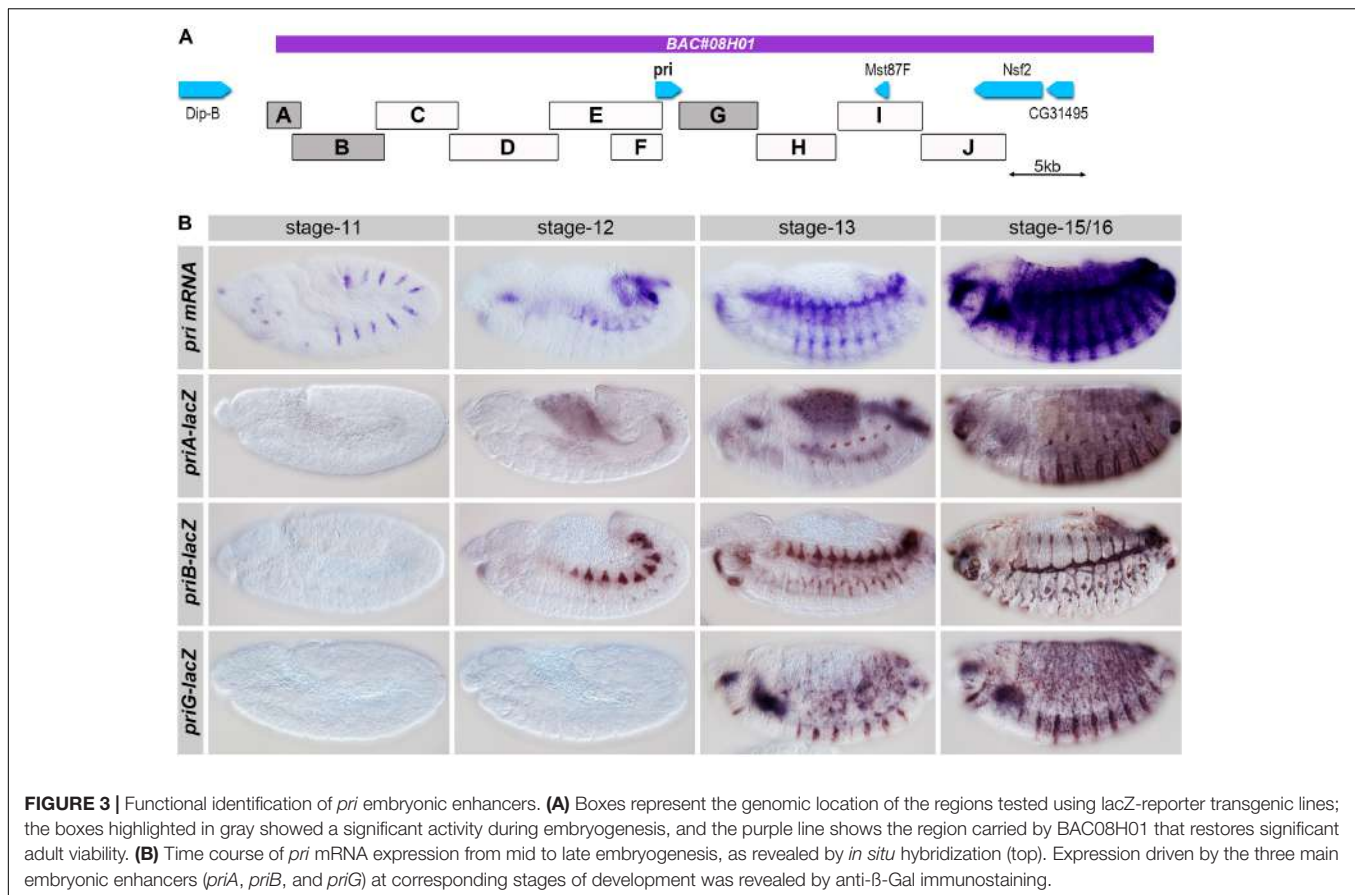
(*pri*¹/*pri*³) and tested whether BAC constructs could restore (i) differentiation of embryonic trichomes and (ii) full development up to viable and fertile adults. All BAC constructs were able to significantly restore the formation of trichomes (**Figures 2A,B**). This result suggested that the minimal genomic sequence required for *pri* function in epidermal trichome formation was included in the region shared by all rescue constructs, which is 8.7 kb in length (bright yellow, **Figure 2A**). However, none of the two smaller BAC constructs (176K10 and 150C8, of 20.4 and 21.9 kb, respectively) was sufficient to restore viability and the progression throughout later developmental stages. In contrast, the large 51O01 construct (98.7 kb) fully rescued the emergence of adults when introduced in a *pri* null genetic context. A significant yet weaker rescuing activity was also observed for BAC 08H01 (**Figure 2B**), restricting the minimal genetic interval for *pri* function to a 52.6-kb DNA sequence (light yellow, **Figure 2A**).

These data, therefore, show that despite a compact transcribed region, the developmental functions of *pri* relies on the activity of large upstream and downstream genomic sequences.

Multiple Enhancers Drive Tissue-Specific Expression of *pri* in Embryos

Having defined the functional genetic unit of *pri* for *Drosophila* development, we next attempted to delineate the *cis*-regulatory

sequences controlling its expression in embryo. We used a series of transgenic LacZ reporter lines (Chanut-Delalande et al., 2014) to systematically examine the whole region. The *pri* locus was dissected in DNA fragments of approximately 5 kb long, which displayed overlapping regions to ensure coverage of the locus (**Figure 3A**). We performed LacZ staining on staged embryos to define the spatial and temporal pattern(s) of expression driven by each region. Three regions drove expression in subsets of embryonic cells, reproducing parts of the endogenous pattern of *pri* mRNA. *priG* was expressed in the epidermis from stage 13 mostly in ventral cells, and its activity extended to the dorsal region in stage 15/16 (**Figure 3B**). A second remote enhancer, *priA*, also exhibited epidermal expression in the dorsal and ventral regions. Following a faint onset of epidermal activity at stage 13, *priA* displayed a strong activity in the epidermis, which peaked at stage 15/16 (**Figure 3B**). Both *priA* and *priG* enhancers displayed stronger activity in presumptive trichome cells (as best seen in the ventral region), consistently with the endogenous pattern of *pri* mRNA (Galindo et al., 2007; Kondo et al., 2007). Finally, we observed that the *priB* enhancer drove strong expression in the embryonic tracheal system from stage 12–16 (**Figure 3B**). This activity was yet slightly delayed when compared with *pri* endogenous expression in the tracheal system, the onset of which being visible at stage 11. Other DNA regions, including those comprising the *pri* promoter, displayed either barely detectable activity or expression in tissues (posterior gut, groups



of mesodermal cells, hemocytes, see **Supplementary Figure 1**) that might reveal additional aspects of *pri* expression, but which currently lack functional evidence.

All together, these results identify three main enhancers that control *pri* expression across embryonic development. Two separate *cis*-regulatory regions, *priA* and *priG*, drive *pri* expression in the epidermis and a third one, *priB*, in the tracheal system.

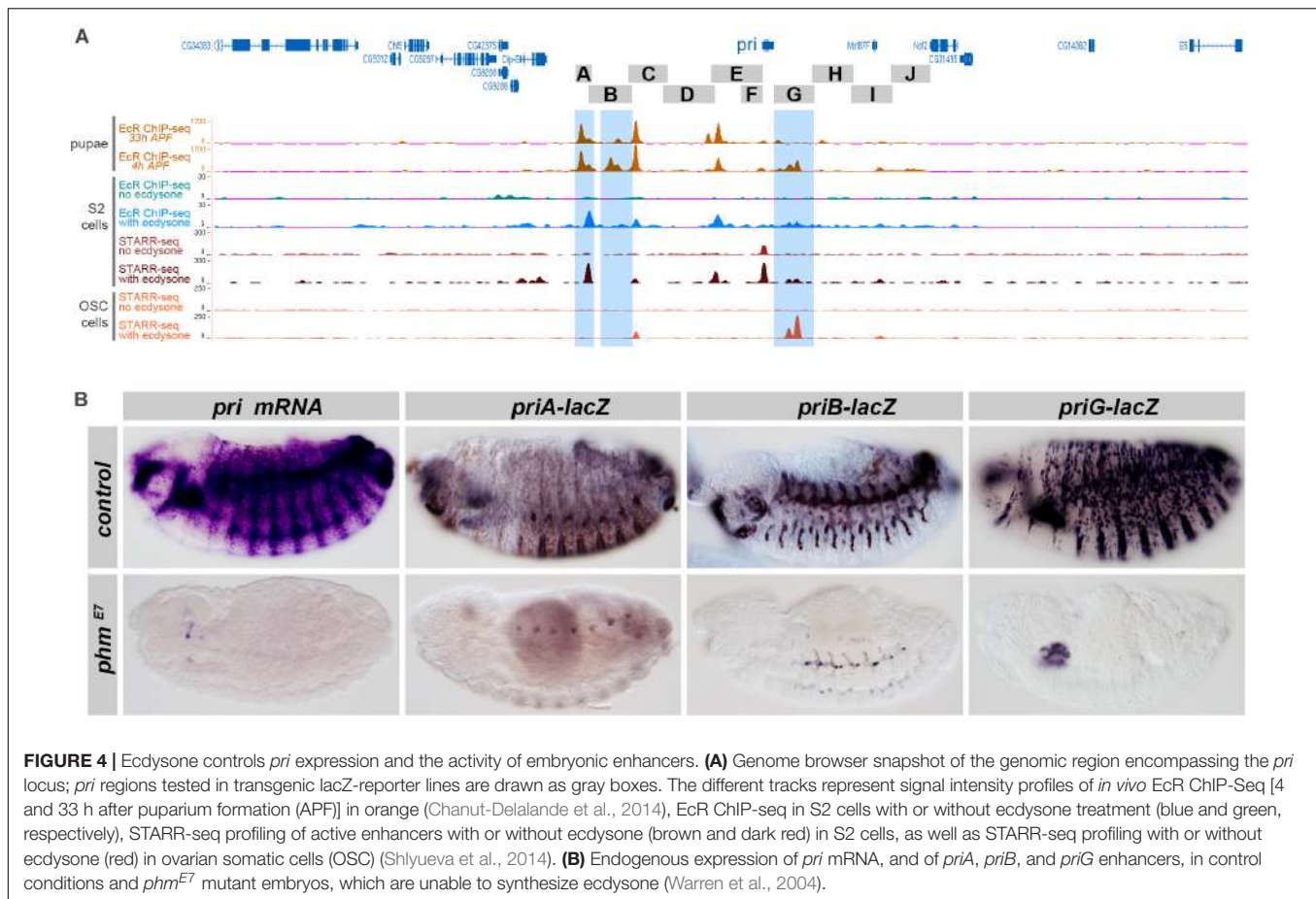
Ecdysone Controls the Activity of Separate *pri* Embryonic Enhancers

Previous work has shown that the expression of *pri* in the embryonic epidermis requires ecdysone signaling (Chanut-Delalande et al., 2014). Since the embryonic expression of *pri* is driven by three major enhancers, *priA*, *priB*, and *priG*, we next investigated whether these enhancers could be regulated by ecdysone and EcR.

ChIP-seq data showed that the *pri* locus is prominently bound by EcR in pupae (Chanut-Delalande et al., 2014), and each of the three embryonic enhancers *priA*, *priB*, and *priG* contain major EcR peaks (highlighted in blue in **Figure 4A**). In addition, profiling of EcR binding in cultured S2 cells (Shlyueva et al., 2014) also detected a strong EcR peak in the *priA* sequence. Although displaying a weaker intensity, EcR binding is also seen within *priB* and *priG* sequences. These EcR binding events

are likely to be relevant, since the signal is only seen in the presence of ecdysone (**Figure 4A**). The same study used high-throughput STAR-seq profiling to functionally define enhancers whose activity was specifically activated upon ecdysone treatment in cultured cells (Shlyueva et al., 2014). Of note, the *priA* region contains the strongest ecdysone-responsive enhancer of the third chromosome in S2 cells. A weaker enhancer activity was also found in the *priG* region in S2 cells, while *priB* corresponded to a major ecdysone-responsive enhancer (**Figure 4A**) in OSC cells (Shlyueva et al., 2014).

During embryogenesis, the expression of *pri* mRNA in the tracheal system, epidermis, gut, and pharynx was abolished in *phantom* (*phm*) mutant embryos (**Figure 4B**), which are unable to synthesize ecdysone (Warren et al., 2004). We next tested whether the lack of ecdysone also influenced the activity of *pri* embryonic enhancers. In embryos deprived of ecdysone, the activity of *priG* was strongly impaired, preventing the detection of any epidermal expression (**Figure 4B**). As an additional test, we drove a dominant negative version of EcR (EcR^{DN}) in a subset of epidermal cells using *ptc-gal4*, leading to a strong reduction of *priG* epidermal expression in EcR^{DN} cells, as also observed for endogenous *pri* mRNA (**Supplementary Figure 2**). The activity of *priA* was also dramatically decreased in the absence of ecdysone (**Figure 4B**). Finally, although not entirely abolished, expression of the *priB* enhancer was reduced in the tracheal system (**Figure 4B**). In



contrast, other *pri* regions that contain major pupal peaks of EcR binding in pupae (*priBC*, *priDE*) only displayed weak embryonic expression without obvious reduction in *phm* mutant embryos (Supplementary Figure 3).

Taken together, these data indicate that ecdysone is a requisite for the activity of *priA* and *priG* epidermal enhancers in the embryo, and to a lesser extent of the tracheal enhancer *priB*.

EcR Binds to *pri* Enhancers in a Tissue-Specific Manner

Although the 20E/EcR complex regulates the expression of many genes (Li and White, 2003; Beckstead et al., 2005; Gauhar et al., 2009; Shlyueva et al., 2014; Stoiber et al., 2016; Uyehara and McKay, 2019), experimental evidence of its direct role in the *in vivo* activity of enhancers remains limited. We aimed to get a deeper understanding of *pri* regulation in the epidermis by ecdysone, and the *priA* and *priG* enhancers appeared as good candidates for direct regulation by EcR.

We carried out systematic dissection of these two enhancers to delineate the respective minimal region sufficient to drive proper expression in embryonic epidermis. The *priA* element (Figures 5A–C) contains a sequence of 288 bp that drives ecdysone-dependent expression in S2 cells and whose activity was abolished the mutation of two EcR binding

sites (Shlyueva et al., 2014). These two sites are evolutionarily conserved across *Drosophila* species (Figure 5C) and a 3' deletion of *priA* that lacks this region, *priAb1*, was devoid of activity in embryos (not shown). Reciprocally, the *priAb2* region displayed epidermal expression, although with a significantly decreased intensity when compared with the full *priA* enhancer (Figure 5B). However, the short enhancer delineated in S2 cells (*priAs*) was not sufficient to drive a proper *in vivo* expression in the embryonic epidermis. Instead, we observed ectopic expression in scattered cells that likely represents blood cells. Similar results were also observed with a construct bearing an extended version *priAse* (421 bp, Figures 5A,B). These data indicate that the *in vivo* activity of *priA* requires additional *cis*-regulatory elements present in the 5' region of the *priA* sequence, which may include other EcR binding sites. Accordingly, mutation of both EcR binding sites in the backbone of the full *priA* enhancer did not significantly impact on its epidermal expression (Figure 5C).

The *priG* enhancer was also dissected in a series of overlapping DNA fragments of approximately 1 kb in length (*priG1-5*, Figure 5D). Transgenic reporters displayed no significant activity for *priG1* or *priG4*, *priG3* and *priG5* displaying only weak expression in embryos (Figure 5E). In contrast, *priG2* exhibited strong epidermal activity comparable with the entire *priG* enhancer. Using the JASPAR database of transcription factor

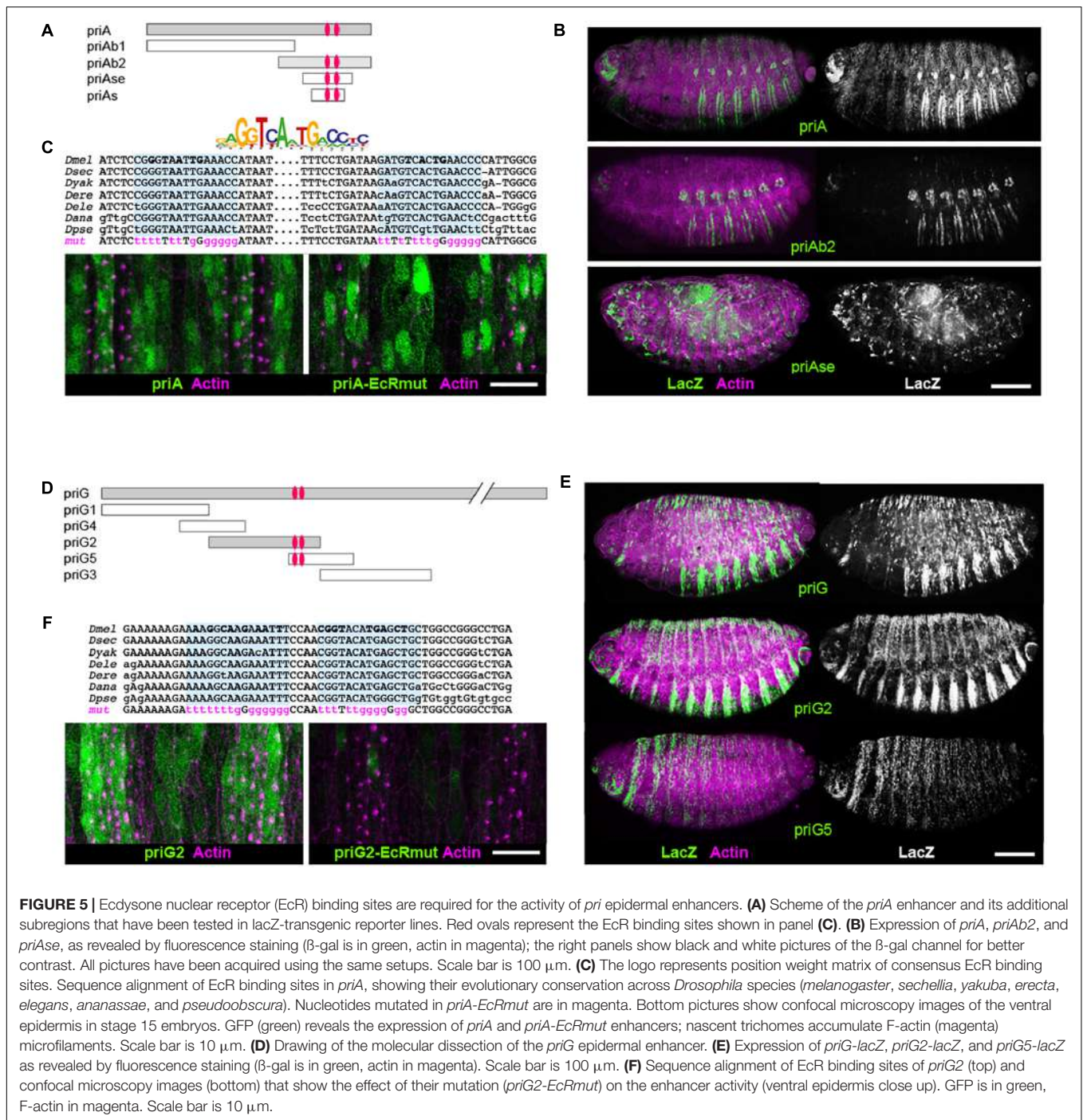


FIGURE 5 | Ecdysone nuclear receptor (EcR) binding sites are required for the activity of *pri* epidermal enhancers. **(A)** Scheme of the *priA* enhancer and its additional subregions that have been tested in lacZ-transgenic reporter lines. Red ovals represent the EcR binding sites shown in panel **(C)**. **(B)** Expression of *priA*, *priAb2*, and *priAse*, as revealed by fluorescence staining (β -gal is in green, actin in magenta); the right panels show black and white pictures of the β -gal channel for better contrast. All pictures have been acquired using the same setups. Scale bar is 100 μ m. **(C)** The logo represents position weight matrix of consensus EcR binding sites. Sequence alignment of EcR binding sites in *priA*, showing their evolutionary conservation across *Drosophila* species (*melanogaster*, *sechellia*, *yakuba*, *erecta*, *elegans*, *ananassae*, and *pseudobscura*). Nucleotides mutated in *priA-EcRmut* are in magenta. Bottom pictures show confocal microscopy images of the ventral epidermis in stage 15 embryos. GFP (green) reveals the expression of *priA* and *priA-EcRmut* enhancers; nascent trichomes accumulate F-actin (magenta) microfilaments. Scale bar is 10 μ m. **(D)** Drawing of the molecular dissection of the *priG* epidermal enhancer. **(E)** Expression of *priG-lacZ*, *priG2-lacZ*, and *priG5-lacZ* as revealed by fluorescence staining (β -gal is in green, actin in magenta). Scale bar is 100 μ m. **(F)** Sequence alignment of EcR binding sites of *priG2* (top) and confocal microscopy images (bottom) that show the effect of their mutation (*priG2-EcRmut*) on the enhancer activity (ventral epidermis close up). GFP is in green, F-actin in magenta. Scale bar is 10 μ m.

binding sites (Mathelier et al., 2014), we found two juxtaposed putative EcR sites in *priG2*, which are evolutionarily conserved (Figure 5F). We inactivated these EcR binding sites in the *priG2* construct by site-directed mutagenesis, and the disruption of both sites abolished the activity of the mutated enhancer, *priG2-EcRmut*, when assayed in embryos (Figure 5F).

Hence, we identified a 1-kb minimal regulatory region, *priG2*, sufficient to drive expression in epidermal cells. Its epidermal activity is dependent on ecdysone signaling, likely through a

direct action of EcR since the mutation of EcR binding sites is sufficient to inactivate this enhancer.

Distinct Enhancers Drive *pri* Expression in the Leg Primordium

During post-embryonic development, *pri* is essential for adult leg morphogenesis, and the loss of *pri* leads to strong defects in tarsal segments, which fuse and do not develop (Galindo et al., 2007;

Pueyo and Couso, 2008, 2011; Pi et al., 2011). *Pri* expression in the leg disk is controlled by ecdysone (Chanut-Delalande et al., 2014), and we assayed whether EcR function was required for tarsal development. As expected, driving EcR^{DN} in the distal region of legs using *dll-gal4* provoked strong phenotypes, in which tarsal segments were mostly absent with only a small region likely representing atrophic and fused remnants of those (Figure 6A). Importantly, we found that *pri* was sufficient to suppress the defects resulting from EcR^{DN} expression, therefore, suggesting that *pri* is a major target of ecdysone signaling for adult appendage development.

Using *in vivo* reporter assays, we next searched for genomic regions driving *pri* expression during the formation of adult structures, which occurs from larval to pupal stages. To monitor the activity of *pri* enhancers during post-embryonic development, we generated a series of transgenic lines that drive a nuclear-GFP reporter. We did not observe expression for *priC*, *DE*, *F*, *H*, and *J* enhancers in the leg primordium, neither in the third instar imaginal disks nor after puparium

formation (Supplementary Figure 4 and data not shown). In contrast, five separate *pri* enhancers (*priA*, *priB*, *priBC*, *priG2*, and *priI*) drove significant expression in leg disks, with a marked temporal regulation of their activity. Notably, *priI* drove expression in the larval leg disk from the early-mid third instar stage (L3), which was then downregulated in late L3, and came up with a high expression from 1 to 6 h after puparium formation (APF) (Figure 6B), i.e., a pattern reminiscent of the endogenous expression of *pri* mRNA (Galindo et al., 2007; Pueyo and Couso, 2008, 2011; Chanut-Delalande et al., 2014). *priB* displayed a very limited expression in larval leg disks, which was substantiated at pupal stages (Supplementary Figure 4). *priA*, *priBC*, and *priG2* enhancers, which were not active in larval leg disks, drove salt-and-pepper patterns of expression in pupal legs (Figure 6B and Supplementary Figure 4).

To decipher how *pri* expression was regulated by ecdysone signaling, we tested whether the loss of EcR function affected the expression of *pri* enhancers in leg disks. We used targeted

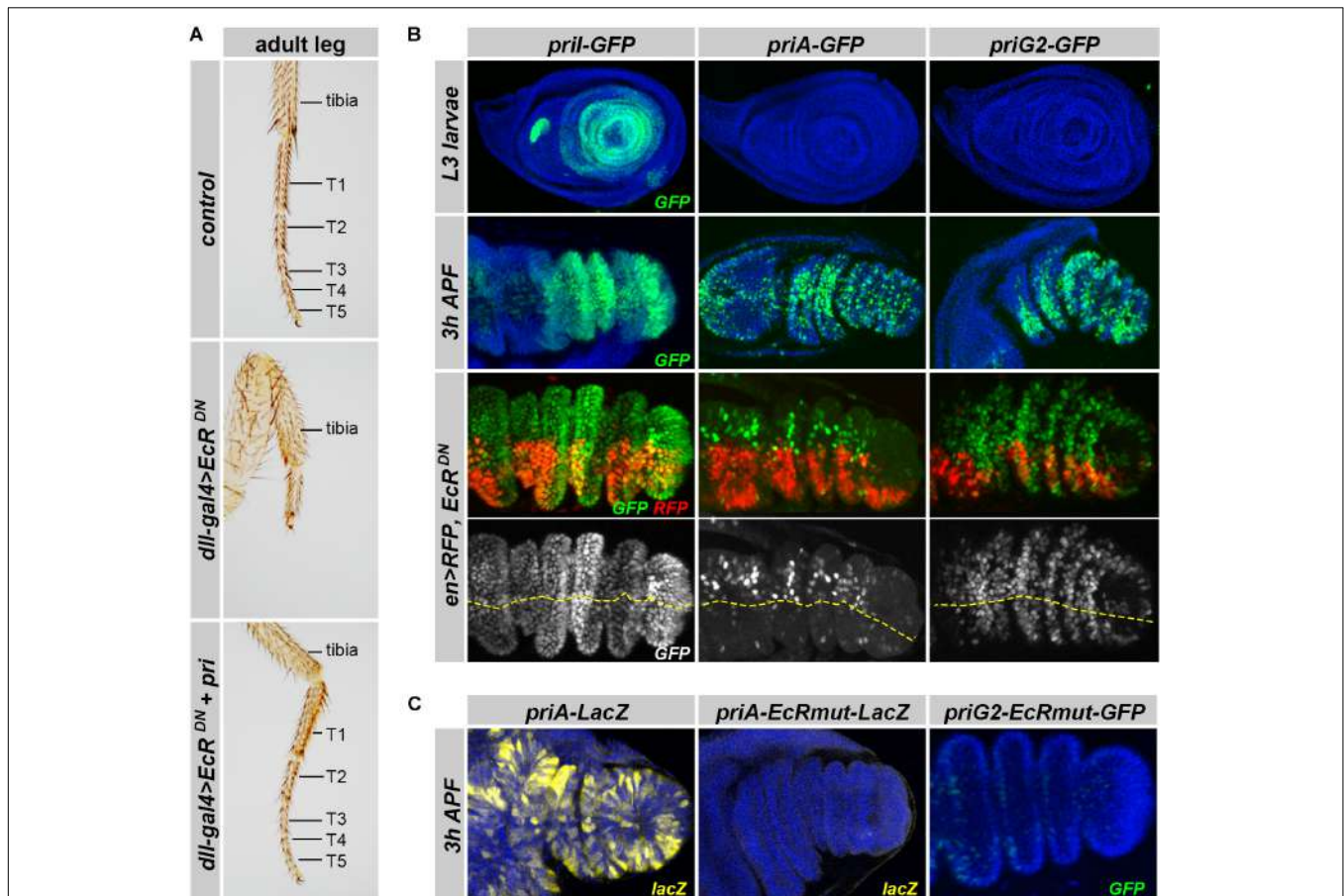


FIGURE 6 | Expression of *pri* enhancers active in leg primordia at the larval/prepupal transition. **(A)** Pictures of adult forelegs in control individuals or animals expressing EcR^{DN} under the control of *dll-gal4*, with or without simultaneous expression of *pri* mRNA. T1–5 represents tarsal segments. **(B)** Activity of *priI*, *priA*, and *priG2* enhancers (green) in control leg disks (upper panels) at mid L3 and pupal (3–5 h APF) stages (two top panels). The two bottom panels show the effect of expressing EcR^{DN} under the control of *en-Gal4* [that also drives RFP (red)] on the activity of *priI*, *priA*, and *priG2* enhancers in pupal leg disks. DAPI is in blue. The yellow line highlights the boundary between RFP-positive (posterior compartment) and RFP-negative (anterior) cells. **(C)** Effect of inactivating EcR binding sites on the activity of *priA* (β -Gal immunostaining, yellow) and of *priG2* (GFP, green) in the pupal leg primordium. Nuclei are visualized by DAPI (blue).

expression of EcR^{DN} , or EcR-RNAi , driven in the posterior compartment by en-Gal4 , the anterior region providing internal control (**Figure 6B** and **Supplementary Figure 4**). Although some pri enhancers were not (priI) or only slightly (priG2) affected by EcR inactivation (**Figure 6B** and **Supplementary Figure 4**), the activity of priA was clearly reduced in the posterior cells of pupal leg disks, when compared with wild-type anterior cells. Furthermore, point mutations that disrupt the two EcR binding sites (required in S2 cells but not in embryos, see **Figure 5**) were sufficient to abrogate priA activity in pupal legs (**Figure 6C**). Of note, we found that the priG2 enhancer in which the binding sites for EcR have been killed also displayed reduced activity in the pupal leg disks, suggesting that our EcR^{DN} conditions represented only a partial inactivation of ecdysone signaling.

Pri function is also required for the proper development of adult wings (Pi et al., 2011). Soon after puparium formation, pri is highly expressed in wing disks (Pi et al., 2011) and accumulates in L2–L5 protein regions (**Figure 7A**). Among pri enhancers, we found that priG2 drove a pattern mimicking endogenous expression of pri mRNA (**Figure 7B** and **Supplementary Figure 5**). In the pupal wing, the activity of priG2 was clearly decreased by EcR^{DN} expression (**Figure 7C**). Furthermore, inactivation of EcR binding sites abrogated activity of the priG2 enhancer in pupal wing disks (**Figure 7D**). We concluded that priG2 is a major enhancer that drives pri expression in the wing, likely under direct control of EcR .

Taken together, these data show that pri expression is driven by multiple enhancers during the morphogenesis of

adult tissues, with one, priA , directly responding to EcR in developing legs, while one other, priG2 , requires EcR activity in the wing primordium.

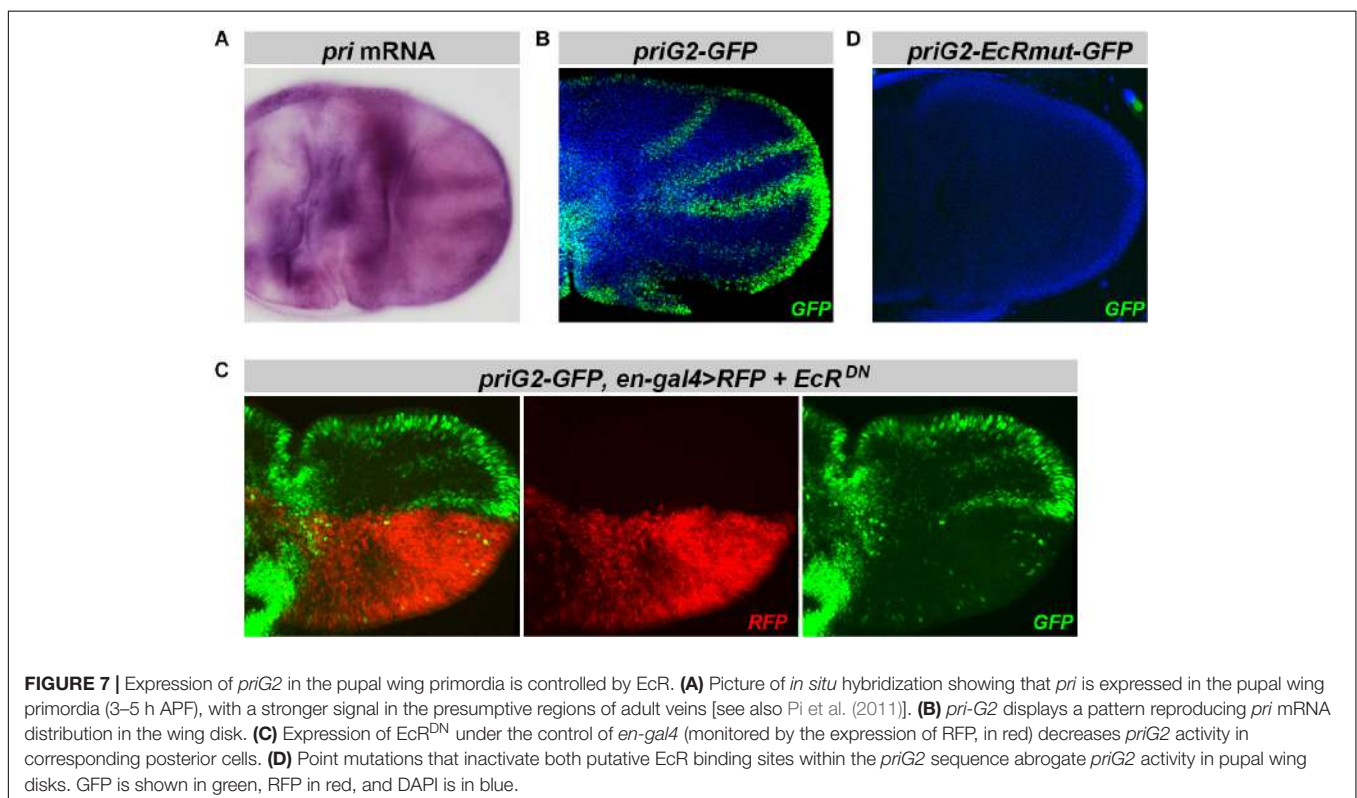
DISCUSSION

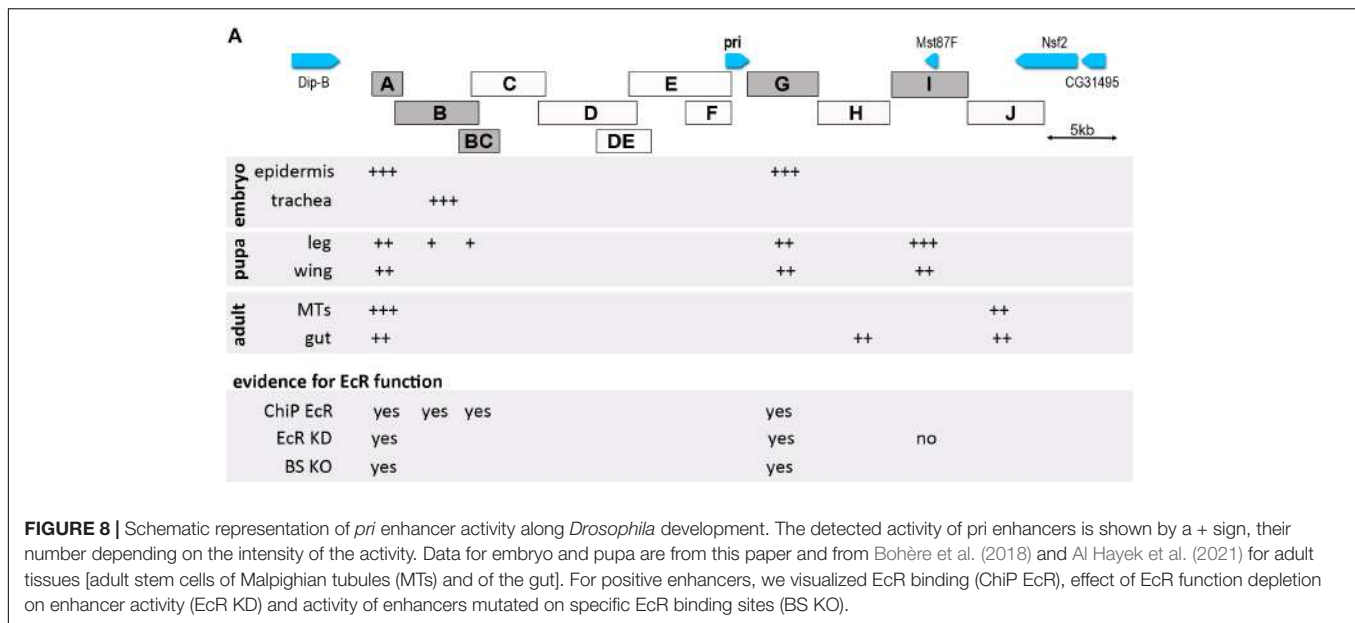
Our data show that the pri gene encoding smORF peptides is a major core component of the response to ecdysone. We found that pri expression involves a wide array of enhancers, some of them being likely directly controlled by the ecdysone nuclear receptor EcR (**Figure 8**). These results further suggest that the cis -regulatory architecture of the pri gene contributes to shape tissue- and stage-specific patterns of ecdysone response during development.

The Pri Gene Displays Hallmarks of a Master Developmental Gene

The expression of eukaryotic genes is primarily determined by the activity of proximal promoter elements, whose activity is further regulated by enhancers that can act at a distance. Genes can schematically be separated in two classes, those initiating transcription over a broad region (dispersed transcription) and generally expressed in most cell types, and those initiating transcription at a focused position, which often involves a TATA box element and remote regulatory enhancers, such as important developmental genes (Vo Ngoc et al., 2017; Furlong and Levine, 2018).

Contrasting with the compact size of its transcribed region, pri functions throughout development involve an unexpectedly





large genomic locus, as defined by rescuing assays. The smallest DNA fragment that allows full rescue of *pri* activity spans over 50 kb, indicating that *pri* expression relies on a large array of *cis*-regulatory regions scattered over this region, which underlie the highly dynamic expression of *pri* across tissues and developmental stages (Savard et al., 2006; Galindo et al., 2007; Kondo et al., 2007). Consistent with this conclusion, we identify remote enhancers located as far as 20 kb upstream and 10 kb downstream of the *pri* transcriptional unit. The attribution of *cis*-regulatory elements located within large intergenic regions to their respective target genes yet remains difficult. For example, the *priI* enhancer overlaps with *Mst87F*, a male-specific transcript expressed in the germline from late pupal stages onward (Kuhn et al., 1991). The remote *priI* enhancer, nevertheless, reproduces the expression of *pri* in the developing larval/pupal leg disk. Also, the strong EcR binding peak contained in the *priA* enhancer was previously attributed to *Dip-B* (Shlyueva et al., 2014), which is located closer than *pri*. Several pieces of evidence are, however, consistent with its role in the regulation of *pri* expression. First, *priA* drives a stage-specific embryonic expression in epidermal cells like *pri* mRNA, while *Dip-B* is mostly expressed in the hindgut and Malpighian tubules³. Second, *priAb2* that exhibits epidermal expression is contained in the smaller *pri* rescuing BAC. Third, levels of *pri* mRNA induction-triggered ecdysone treatment is several times higher than for *Dip-B*, as seen in Kc cells (Skalska et al., 2015), well in line with the strength of this ecdysone-responsive element (Shlyueva et al., 2014). It remains possible that this enhancer contributes, at least in part, to the control of *Dip-B* expression. The *priA* region also represents an interesting case to compare the function of an enhancer between cultured cells and *in vivo* tissues. While *ex vivo* assays have demonstrated the strong activity of a short region of *priA* directly bound by EcR (Shlyueva et al., 2014), neither this sequence nor an

extended version of it is sufficient to faithfully drive *pri* (or *Dip-B*) expression in the embryo. Instead, it leads to ectopic expression in blood cells. It is interesting to note that the activity of this short element in S2 cells also requires binding sites for GATA factors (Shlyueva et al., 2014), and the GATA factor Serpent is well known to play a key in the blood cell lineage (Rehorn et al., 1996; Muratoglu et al., 2007).

Distinct *pri* enhancers can drive similar and/or overlapping patterns, for example, *priA* and *priG* in the embryonic epidermis, or *priI*, *priA*, *priG*, and *priB* in the leg primordium. The expression of major developmental genes often involves apparently redundant enhancers, sometimes called shadow enhancers, which ensure a robust spatiotemporal gene expression, in particular, when development proceeds under non-optimal environmental conditions or in compromised genetic backgrounds (Frankel et al., 2010; Perry et al., 2010; Osterwalder et al., 2018). Genome-wide studies revealed that many developmental genes have shadow enhancers driving a similar activity (Menoret et al., 2013; Cannavo et al., 2016). Therefore, multiple enhancers active in the same tissue could collectively provide robustness against genetic and/or environmental variations to ensure the proper expression of *pri*, throughout embryonic and post-embryonic development.

Individual *pri* enhancers can also be active in different tissues and/or developmental stages; this feature is called pleiotropy (Preger-Ben Noon et al., 2018). *PriA* represents a striking case of a pleiotropic enhancer that is expressed in both the embryonic epidermis and the leg primordium, as well as in renal (Bohère et al., 2018) and intestinal (Al Hayek et al., 2021) adult stem cells. Of note, other genes of the ecdysone cascade possess pleiotropic enhancers, as seen with *EcR*, *Eip75B*, and *ftz-f1* that are driven by enhancers active in both somatic and germline tissues during oogenesis (McDonald et al., 2019). Pleiotropic enhancers may serve as a platform to drive gene expression in several spatiotemporal contexts (Sabaris et al., 2019), and future

³<http://insitu.fruitfly.org/cgi-bin/ex/insitu.pl>

work will be required to better understand their behavior and functions during development (Preger-Ben Noon et al., 2018).

Pri Is a Core Element of the Response to Ecdysone

Animal development requires a precise temporal coordination, across the whole organism, of genetic programs underlying the formation of the different tissues and organs, as well exemplified by major morphological changes such as the transition of immature juveniles into reproductive adults (Rougvie, 2001). The steroid hormone ecdysone plays a key role in the temporal control of major developmental transitions, acting as a systemic signal timely released in response to various inputs from both the internal milieu and the environment (Ou and King-Jones, 2013; Yamanaka et al., 2013). While we now have a wealth of information on the regulation of ecdysone production (Andersen et al., 2013; Ou and King-Jones, 2013; Yamanaka et al., 2013; Niwa and Niwa, 2014), how ecdysone signaling is integrated within terminal differentiation programs remains to be fully elucidated.

Previous work on Pri peptides has provided a molecular connection between the intimate mechanisms of epidermal differentiation and ecdysone signaling (Chanut-Delalande et al., 2014). Several arguments support that *pri* is directly activated by hormone-bound EcR. First, the expression of *pri* is abrogated in mutant embryos lacking ecdysone. Of note, the inactivation of any enzyme of the ecdysone biosynthesis pathway causes a similar phenotype, referred to as Halloween (Warren et al., 2002, 2004; Petryk et al., 2003; Gilbert, 2004; Niwa et al., 2004, 2010; Namiki et al., 2005; Ono et al., 2006; Chanut-Delalande et al., 2014; Enya et al., 2014), which is characterized by a poorly differentiated cuticle and the complete lack of trichomes. Second, there are prominent peaks of EcR binding to *pri* genomic regions *in vivo* (Chanut-Delalande et al., 2014), as well as in different cell lines following ecdysone treatment that induces a burst of *pri* transcription (Shlyueva et al., 2014; Skalska et al., 2015). RNA-seq data from cell lines also provide information on the kinetics of *pri* induction, which starts after that of *Eip75B* and *Hr4*, therefore, suggesting that the activation of *pri* expression requires both 20E/EcR binding and the production of early TFs, which is representative of the so-called early-late response genes (Ou and King-Jones, 2013). When assayed *in vivo*, several regions of the *pri* locus act as tissue-specific enhancers responsive to ecdysone during development. The activity of *priA*, *priB*, and *priG* enhancers is compromised in embryos lacking ecdysone, and targeted expression of EcR^{DN} represses the expression of *priG* and *priA* enhancers in corresponding cells, in embryos and larval/pupal tissues. Finally, point mutations of EcR binding sites are sufficient to kill the activity of *priG2* and *priA*, in the embryo and in pupal wings or legs, as expected if the transcription of *pri* involves a direct control by the 20E/EcR complex.

Our results further suggest that *pri* is a key and broad mediator of the response to ecdysone. Indeed, the re-expression of *pri* is sufficient to compensate for deficient ecdysone signaling across various tissues. In embryos, *pri* expression in epidermal cells restore trichomes in the complete absence of ecdysone, showing that *pri* is a pivotal target of ecdysone action in the epidermis (Chanut-Delalande et al., 2014). The same is also true in tracheal

cells, as shown recently (Taira et al., 2021). In a similar way, the expression of Pri peptides can counteract the defects resulting from targeted expression of EcR^{DN} in pupae, restoring both proper development and adult viability (Chanut-Delalande et al., 2014). Recent studies have revealed that ecdysone signaling is essential to adapt the behavior of intestinal stem cells in adult females, as part of a post-mating response (Ahmed et al., 2020; Zipper et al., 2020). We found that Pri peptides are playing an important role in this ovary-to-gut interorgan communication, being able to substitute for ecdysone signaling in intestinal stem cells (Al Hayek et al., 2021). Hence, Pri peptides appear as key mediators of ecdysone, in a broad variety of tissues throughout development, as well as in adult stem cells. We propose that *pri* may belong to a small number of genes, representing a core module for the response to ecdysone. Supporting this model, *pri* expression is strongly induced by ecdysone in almost all cell lines that have been profiled by the work of Stoiber et al. (2016), presenting a behavior similar to the canonical *Broad* and *Eip74EF* factors. Future work will be needed to thoroughly test the full extent of Pri peptides in mediating response to ecdysone during development and adulthood.

Pri Cis-Regulatory Architecture Contributes to Shape Stage and Tissue-Specific Patterns of the Response to Steroids

One most intriguing question about ecdysone signaling is how this systemic hormone can trigger various patterns of tissue-specific responses, including different timing across target tissues. Several mechanisms are likely at work in implementing specific responses to ecdysone. They include regulation of ecdysone uptake in peripheral tissues (Okamoto et al., 2018; Okamoto and Yamanaka, 2020), the existence of distinct EcR isoforms with tissue-specific functions (Cherbas et al., 2003; Braun et al., 2009; Gautam et al., 2015), the variation of partners associating with EcR [e.g., ultraspiracle that generally dimerizes with EcR (Hodin and Riddiford, 1998) is not required for activating glue genes in salivary glands (Costantino et al., 2008)], as well as chromatin factors that regulate accessibility of EcR binding sites across cell types (Gutierrez-Perez et al., 2019). Since *pri* appears required for the response to ecdysone in various cell contexts, the dynamic pattern of *pri* expression might also contribute to determine which and when tissues become competent to respond to the systemic hormonal signal.

To explain how *pri* is activated by ecdysone in so many different contexts, a simple model would be that EcR might act on *pri* promoter and/or proximal elements, distant enhancers providing tissue-specific regulatory cues. Instead, our results indicate that independent remote *pri* enhancers likely require the direct binding of EcR for their activity, since their activity is abolished upon the mutation of EcR binding sites. Therefore, the *cis*-regulatory landscape of *pri* that involves multiple enhancers appears well suited to sculpt various patterns of ecdysone response throughout development. If, as supported by the current data, *pri* is a major mediator of ecdysone, one would expect that changes within the *pri* locus may provide efficient means to adapt

the organism to environmental conditions, at different time-scales throughout evolution. Although this hypothesis remains to be tested, it is interesting to note that *pri* has been identified as a candidate gene to be involved in the adaptation to darkness (Izutsu et al., 2015).

CONCLUSION

In this work, we depicted the complex regulation of the *pri* gene by ecdysone to sustain strong levels of expression that dynamically varies among tissues and along development. Such a complexity in transcription regulatory mechanisms is well established for developmental genes encoding regular proteins, but it was yet unknown for genes only encoding smORF peptides. Our data suggest that smORF peptide genes can behave as *bona fide* key developmental regulators, opening novel research paths for the functional exploration of this recently emerging family of genes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

HC-D, YK, and FP contributed to the conception and design of the study and wrote the first draft of the manuscript. SP helped

in the student supervision. AM-F performed the computational analysis. AD, JZ, MG, DM, PV, SM-P, and HC-D contributed to the experimental work and interpretation. All authors read and approved the submitted version.

FUNDING

This work was supported by ANR grants (FP: Chrononet and JZ: Morphosmorf) and funding from Azm and Saade Association to AD, and JST PRESTO program and MEXT KAKENHI (20370091) to YK.

ACKNOWLEDGMENTS

We would like to thank Flybase, the Bloomington, Vienna, and Kyoto stock centers, as well as H. Bellen for BAC constructs. We would also like to thank the Toulouse RIO Imaging platform for help with microscopy and M. Aguirrebengoa for bioinformatics (BigA). We would also further like to thank other members of the FP lab for helpful suggestions and critical reading of the manuscript. Parts of the thesis of AD are present in this manuscript (Dib, 2016).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.714152/full#supplementary-material>

REFERENCES

- Ahmed, S. M. H., Maldera, J. A., Krunic, D., Paiva-Silva, G. O., Penalva, C., Teleman, A. A., et al. (2020). Fitness trade-offs incurred by ovary-to-gut steroid signalling in *Drosophila*. *Nature* 584, 415–419. doi: 10.1038/s41586-020-2462-y
- Al Hayek, S., Alsawadi, A., Kambris, Z., Boquete, J. P., Bohere, J., Immarigeon, C., et al. (2021). Steroid-dependent switch of OvoL/Shavenbaby controls self-renewal versus differentiation of intestinal stem cells. *EMBO J.* 40:e104347.
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Andersen, D. S., Colombani, J., and Leopold, P. (2013). Coordination of organ growth: principles and outstanding questions from the world of insects. *Trends Cell Biol.* 23, 336–344. doi: 10.1016/j.tcb.2013.03.005
- Andrew, D. J., and Baker, B. S. (2008). Expression of the *Drosophila* secreted cuticle protein 73 (*dsc73*) requires Shavenbaby. *Dev. Dyn.* 237, 1198–1206. doi: 10.1002/dvdy.21512
- Andrews, S. J., and Rothnagel, J. A. (2014). Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Genet.* 15, 193–204. doi: 10.1038/nrg3520
- Beckstead, R. B., Lam, G., and Thummel, C. S. (2005). The genomic response to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. *Genome Biol.* 6:R99.
- Bender, M., Imam, F. B., Talbot, W. S., Ganetzky, B., and Hogness, D. S. (1997). *Drosophila* ecdysone receptor mutations reveal functional differences among receptor isoforms. *Cell* 91, 777–788. doi: 10.1016/s0092-8674(00)80466-3
- Bohère, J., Mancheno-Ferris, A., Al Hayek, S., Zanet, J., Valenti, P., Akino, K., et al. (2018). Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis. *Nat. Commun.* 9:5123.
- Braun, S., Azoitei, A., and Spindler-Barth, M. (2009). DNA-binding properties of *Drosophila* ecdysone receptor isoforms and their modification by the heterodimerization partner ultraspiracle. *Arch. Insect Biochem. Physiol.* 72, 172–191. doi: 10.1002/arch.20328
- Cannavo, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., et al. (2016). Shadow enhancers are pervasive features of developmental regulatory networks. *Curr. Biol. CB* 26, 38–51. doi: 10.1016/j.cub.2015.11.034
- Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F., and Plaza, S. (2006). Shavenbaby couples patterning to epidermal cell shape control. *PLoS Biol.* 4:e290. doi: 10.1371/journal.pbio.0040290
- Chanut-Delalande, H., Ferrer, P., Payre, F., and Plaza, S. (2012). Effectors of tridimensional cell morphogenesis and their evolution. *Semin. Cell Dev. Biol.* 23, 341–349. doi: 10.1016/j.semcdb.2012.03.002
- Chanut-Delalande, H., Hashimoto, Y., Pelissier-Monier, A., Spokony, R., Dib, A., Kondo, T., et al. (2014). Pri peptides are mediators of ecdysone for the temporal control of development. *Nat. Cell Biol.* 16, 1035–1044. doi: 10.1038/ncb3052
- Cheatle Jarvela, A. M., and Pick, L. (2017). The function and evolution of nuclear receptors in insect embryonic development. *Curr. Top. Dev. Biol.* 125, 39–70. doi: 10.1016/bs.ctdb.2017.01.003
- Cherbas, L., Hu, X., Zhimulev, I., Belyaeva, E., and Cherbas, P. (2003). EcR isoforms in *Drosophila*: testing tissue-specific requirements by targeted blockade and rescue. *Development (Cambridge, England)* 130, 271–284. doi: 10.1242/dev.00205

- Costantino, B. F., Bricker, D. K., Alexandre, K., Shen, K., Merriam, J. R., Antoniewski, C., et al. (2008). A novel ecdysone receptor mediates steroid-regulated developmental events during the mid-third instar of *Drosophila*. *PLoS Genet.* 4:e1000102. doi: 10.1371/journal.pgen.1000102
- Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., et al. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* 160, 191–203. doi: 10.1016/j.cell.2014.11.041
- Delon, I., Chanut-Delalande, H., and Payre, F. (2003). The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. *Mech. Dev.* 120, 747–758. doi: 10.1016/s0925-4773(03)0081-9
- Dib, A. (2016). *Pri a Novel Target of Ecdysone for the Temporal Control of Drosophila Development*. Ph.D. thesis. Toulouse: Université Paul Sabatier.
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Enya, S., Ameku, T., Igarashi, F., Iga, M., Kataoka, H., Shinoda, T., et al. (2014). A Halloween gene noppera-bo encodes a glutathione S-transferase essential for ecdysteroid biosynthesis via regulating the behaviour of cholesterol in *Drosophila*. *Sci. Rep.* 4:6586.
- Fernandes, I., Chanut-Delalande, H., Ferrer, P., Latapie, Y., Waltzer, L., Affolter, M., et al. (2010). Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes. *Dev. Cell* 18, 64–76. doi: 10.1016/j.devcel.2009.11.009
- Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., and Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466, 490–493. doi: 10.1038/nature09158
- Furlong, E. E. M., and Levine, M. (2018). Developmental enhancers and chromosome topology. *Science (New York, NY)* 361, 1341–1345. doi: 10.1126/science.aau0320
- Galindo, M. I., Pueyo, J. I., Fouix, S., Bishop, S. A., and Couso, J. P. (2007). Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* 5:e106. doi: 10.1371/journal.pbio.0050106
- Gauhar, Z., Sun, L. V., Hua, S., Mason, C. E., Fuchs, F., Li, T. R., et al. (2009). Genomic mapping of binding regions for the Ecdysone receptor protein complex. *Genome Res.* 19, 1006–1013. doi: 10.1101/gr.081349.108
- Gautam, N. K., Verma, P., and Tapadia, M. G. (2015). Ecdysone regulates morphogenesis and function of Malpighian tubules in *Drosophila melanogaster* through EcR-B2 isoform. *Dev. Biol.* 398, 163–176. doi: 10.1016/j.ydbio.2014.11.003
- Ghbeish, N., Tsai, C. C., Schubiger, M., Zhou, J. Y., Evans, R. M., and McKeown, M. (2001). The dual role of ultraspiracle, the *Drosophila* retinoid X receptor, in the ecdysone response. *Proc. Natl. Acad. Sci. U.S.A.* 98, 3867–3872. doi: 10.1073/pnas.061437798
- Gilbert, L. I. (2004). Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in *Drosophila melanogaster*. *Mol. Cell. Endocrinol.* 215, 1–10. doi: 10.1016/j.mce.2003.11.003
- Gutierrez-Perez, I., Rowley, M. J., Lyu, X., Valadez-Graham, V., Vallejo, D. M., Ballesta-Illan, E., et al. (2019). Ecdysone-Induced 3D chromatin reorganization involves active enhancers bound by pipsqueak and polycomb. *Cell Rep.* 28, 2715–2727.e5.
- Hall, B. L., and Thummel, C. S. (1998). The RXR homolog ultraspiracle is an essential component of the *Drosophila* ecdysone receptor. *Development (Cambridge, England)* 125, 4709–4717. doi: 10.1242/dev.125.23.4709
- Hodin, J., and Riddiford, L. M. (1998). The ecdysone receptor and ultraspiracle regulate the timing and progression of ovarian morphogenesis during *Drosophila* metamorphosis. *Dev. Genes Evol.* 208, 304–317. doi: 10.1007/s004270050186
- Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A., et al. (2005). Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*. *Genes Cells* 10, 1163–1173. doi: 10.1111/j.1365-2443.2005.00910.x
- Izutsu, M., Toyoda, A., Fujiyama, A., Agata, K., and Fuse, N. (2015). Dynamics of dark-fly genome under environmental selections. *G3 (Bethesda)* 6, 365–376. doi: 10.1534/g3.115.023549
- Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S., and Kageyama, Y. (2007). Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* 9, 660–665. doi: 10.1038/ncb1595
- Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., et al. (2010). Small peptides switch the transcriptional activity of Shavenbaby during *Drosophila* embryogenesis. *Science (New York, NY)* 329, 336–339. doi: 10.1126/science.1188158
- Kozlova, T., and Thummel, C. S. (2003). Essential roles for ecdysone signaling during *Drosophila* mid-embryonic development. *Science (New York, NY)* 301, 1911–1914. doi: 10.1126/science.1087419
- Kuhn, R., Kuhn, C., Borsch, D., Glatzer, K. H., Schafer, U., and Schafer, M. (1991). A cluster of four genes selectively expressed in the male germ line of *Drosophila melanogaster*. *Mech. Dev.* 35, 143–151. doi: 10.1016/0925-4773(91)90064-d
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, T. R., and White, K. P. (2003). Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. *Dev. Cell* 5, 59–72. doi: 10.1016/s1534-5807(03)00192-8
- Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 42, D142–D147. doi: 10.1093/nar/gkt997
- McDonald, S. I., Beachum, A. N., Hinnant, T. D., Blake, A. J., Bynum, T., Hickman, E. P., et al. (2019). Novel cis-regulatory regions in ecdysone responsive genes are sufficient to promote gene expression in *Drosophila* ovarian cells. *Gene Expr. Patterns* 34:119074. doi: 10.1016/j.gep.2019.119074
- Menoret, D., Santolini, M., Fernandes, I., Spokony, R., Zanet, J., Gonzalez, I., et al. (2013). Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome Biol.* 14:R86.
- Moussian, B. (2010). Recent advances in understanding mechanisms of insect cuticle differentiation. *Insect Biochem. Mol. Biol.* 40, 363–375. doi: 10.1016/j.ibmb.2010.03.003
- Muratoglu, S., Hough, B., Mon, S. T., and Fossett, N. (2007). The GATA factor Serpent cross-regulates lozenge and u-shaped expression during *Drosophila* blood cell development. *Dev. Biol.* 311, 636–649. doi: 10.1016/j.ydbio.2007.08.015
- Namiki, T., Niwa, R., Sakudoh, T., Shirai, K., Takeuchi, H., and Kataoka, H. (2005). Cytochrome P450 CYP307A1/Spook: a regulator for ecdysone synthesis in insects. *Biochem. Biophys. Res. Commun.* 337, 367–374. doi: 10.1016/j.bbrc.2005.09.043
- Niwa, R., and Niwa, Y. S. (2014). Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. *Biosci. Biotechnol. Biochem.* 78, 1283–1292. doi: 10.1080/09168451.2014.942250
- Niwa, R., Matsuda, T., Yoshiyama, T., Namiki, T., Mita, K., Fujimoto, Y., et al. (2004). CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of Bombyx and *Drosophila*. *J. Biol. Chem.* 279, 35942–35949. doi: 10.1074/jbc.m404514200
- Niwa, R., Namiki, T., Ito, K., Shimada-Niwa, Y., Kiuchi, M., Kawaoka, S., et al. (2010). Non-molting glossy/shroud encodes a short-chain dehydrogenase/reductase that functions in the 'Black Box' of the ecdysteroid biosynthesis pathway. *Development (Cambridge, England)* 137, 1991–1999. doi: 10.1242/dev.045641
- Okamoto, N., and Yamanaka, N. (2020). Steroid hormone entry into the brain requires a membrane transporter in *Drosophila*. *Curr. Biol. CB* 30, 359–366.e3.
- Okamoto, N., Viswanatha, R., Bittar, R., Li, Z., Haga-Yamanaka, S., Perrimon, N., et al. (2018). A membrane transporter is required for steroid hormone uptake in *Drosophila*. *Dev. Cell* 47, 294–305.e7.
- Ono, H., Rewitz, K. F., Shinoda, T., Itoyama, K., Petryk, A., Rybczynski, R., et al. (2006). Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. *Dev. Biol.* 298, 555–570. doi: 10.1016/j.ydbio.2006.07.023
- Osterwalder, M., Barozzi, I., Tissieres, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243. doi: 10.1038/nature25461
- Ou, Q., and King-Jones, K. (2013). What goes up must come down: transcription factors have their say in making ecdysone pulses. *Curr. Top. Dev. Biol.* 103, 35–71.
- Ozturk-Colak, A., Moussian, B., Araujo, S. J., and Casanova, J. (2016). A feedback mechanism converts individual cell features into a supracellular ECM structure in *Drosophila* trachea. *ELife* 5:e09373.

- Pauli, A., Valen, E., and Schier, A. F. (2015). Identifying (non-)coding RNAs and small peptides: challenges and opportunities. *Bioessays* 37, 103–112. doi: 10.1002/bies.201400103
- Payre, F. (2004). Genetic control of epidermis differentiation in *Drosophila*. *Int. J. Dev. Biol.* 48, 207–215. doi: 10.1387/ijdb.15272387
- Payre, F., Vincent, A., and Carreno, S. (1999). ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. *Nature* 400, 271–275. doi: 10.1038/22330
- Perry, M. W., Boettiger, A. N., Bothma, J. P., and Levine, M. (2010). Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* CB 20, 1562–1567. doi: 10.1016/j.cub.2010.07.043
- Petryk, A., Warren, J. T., Marques, G., Jarcho, M. P., Gilbert, L. I., Kahler, J., et al. (2003). Shade is the *Drosophila* P450 enzyme that mediates the hydroxylation of ecdysone to the steroid insect molting hormone 20-hydroxyecdysone. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13773–13778. doi: 10.1073/pnas.2336088100
- Pi, H., Huang, Y. C., Chen, I. C., Lin, C. D., Yeh, H. F., and Pai, L. M. (2011). Identification of 11-amino acid peptides that disrupt Notch-mediated processes in *Drosophila*. *J. Biomed. Sci.* 18:42. doi: 10.1186/1423-0127-18-42
- Preger-Ben Noon, E., Sabaris, G., Ortiz, D. M., Sager, J., Liebowitz, A., Stern, D. L., et al. (2018). Comprehensive analysis of a cis-regulatory region reveals pleiotropy in enhancer function. *Cell Rep.* 22, 3021–3031. doi: 10.1016/j.celrep.2018.02.073
- Pueyo, J. I., and Couso, J. P. (2011). Tarsal-less peptides control Notch signalling through the Shavenbaby transcription factor. *Dev. Biol.* 355, 183–193. doi: 10.1016/j.ydbio.2011.03.033
- Pueyo, J. I., and Couso, J. P. (2008). The 11-amino acid long Tarsal-less peptides trigger a cell signal in *Drosophila* leg development. *Dev. Biol.* 324, 192–201. doi: 10.1016/j.ydbio.2008.08.025
- Pueyo, J. I., Magny, E. G., and Couso, J. P. (2016). New peptides under the s(ORF)ace of the genome. *Trends Biochem. Sci.* 41, 665–678. doi: 10.1016/j.tibs.2016.05.003
- Ray, S., Rosenberg, M. I., Chanut-Delalande, H., Decaras, A., Schwertner, B., Toubiana, W., et al. (2019). The mlpt/Ubr3/Svb module comprises an ancient developmental switch for embryonic patterning. *ELife* 8:e39748.
- Rehorn, K. P., Thelen, H., Michelson, A. M., and Reuter, R. (1996). A molecular aspect of hematopoiesis and endoderm development common to vertebrates and *Drosophila*. *Development (Cambridge, England)* 122, 4023–4031. doi: 10.1242/dev.122.12.4023
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rougvie, A. E. (2001). Control of developmental timing in animals. *Nat. Rev. Genet.* 2, 690–701. doi: 10.1038/35088566
- Ruad, A. F., Lam, G., and Thummel, C. S. (2010). The *Drosophila* nuclear receptors DHR3 and betaFTZ-F1 control overlapping developmental responses in late embryos. *Development (Cambridge, England)* 137, 123–131. doi: 10.1242/dev.042036
- Sabaris, G., Laiker, I., Preger-Ben Noon, E., and Frankel, N. (2019). Actors with multiple roles: pleiotropic enhancers and the paradigm of enhancer modularity. *Trends Genet.* 35, 423–433. doi: 10.1016/j.tig.2019.03.006
- Savard, J., Marques-Souza, H., Aranda, M., and Tautz, D. (2006). A segmentation gene in tribolium produces a polycistronic mRNA that codes for multiple conserved peptides. *Cell* 126, 559–569. doi: 10.1016/j.cell.2006.05.053
- Shlyueva, D., Stelzer, C., Gerlach, D., Yanez-Cuna, J. O., Rath, M., Boryn, L. M., et al. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. *Mol. Cell* 54, 180–192. doi: 10.1016/j.molcel.2014.02.026
- Skalska, L., Stojnic, R., Li, J., Fischer, B., Cerda-Moya, G., Sakai, H., et al. (2015). Chromatin signatures at Notch-regulated enhancers reveal large-scale changes in H3K56ac upon activation. *EMBO J.* 34, 1889–1904. doi: 10.15252/embj.201489923
- Stern, D. L., and Frankel, N. (2013). The structure and evolution of cis-regulatory regions: the shavenbaby story. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 368:20130028. doi: 10.1098/rstb.2013.0028
- Stoiber, M., Celniker, S., Cherbas, L., Brown, B., and Cherbas, P. (2016). Diverse hormone response networks in 41 independent *Drosophila* cell lines. *G3 (Bethesda)* 6, 683–694. doi: 10.1534/g3.115.023366
- Taira, Y., Wada, H., Hayashi, S., and Kageyama, Y. (2021). polished rice mediates ecdysone-dependent control of *Drosophila* embryonic organogenesis. *Genes Cells* 26, 269–281. doi: 10.1111/gtc.12841
- Thummel, C. S. (2001). Molecular mechanisms of developmental timing in *C. elegans* and *Drosophila*. *Dev. Cell* 1, 453–465. doi: 10.1016/s1534-5807(01)00060-0
- Tupy, J. L., Bailey, A. M., Dailey, G., Evans-Holm, M., Siebel, C. W., Misra, S., et al. (2005). Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5495–5500. doi: 10.1073/pnas.0501422102
- Uyehara, C. M., and McKay, D. J. (2019). Direct and widespread role for the nuclear receptor EcR in mediating the response to ecdysone in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 9893–9902. doi: 10.1073/pnas.1900343116
- Venken, K. J., Carlson, J. W., Schulze, K. L., Pan, H., He, Y., Spokony, R., et al. (2009). Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. *Nat. Methods* 6, 431–434. doi: 10.1038/nmeth.1331
- Vo Ngoc, L., Wang, Y. L., Kassavetis, G. A., and Kadonaga, J. T. (2017). The punctilious RNA polymerase II core promoter. *Genes Dev.* 31, 1289–1301. doi: 10.1101/gad.303149.117
- Warren, J. T., Petryk, A., Marques, G., Jarcho, M., Parvy, J. P., Dauphin-Villemant, C., et al. (2002). Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 11043–11048. doi: 10.1073/pnas.162375799
- Warren, J. T., Petryk, A., Marques, G., Parvy, J. P., Shinoda, T., Itoyama, K., et al. (2004). Phantom encodes the 25-hydroxylase of *Drosophila melanogaster* and *Bombyx mori*: a P450 enzyme critical in ecdysone biosynthesis. *Insect Biochem. Mol. Biol.* 34, 991–1010. doi: 10.1016/j.ibmb.2004.06.009
- Yamanaka, N., Rewitz, K. F., and O'Connor, M. B. (2013). Ecdysone control of developmental transitions: lessons from *Drosophila* research. *Annu. Rev. Entomol.* 58, 497–516. doi: 10.1146/annurev-ento-120811-153608
- Yao, T. P., Segreaves, W. A., Oro, A. E., McKeown, M., and Evans, R. M. (1992). *Drosophila* ultraspiracle modulates ecdysone receptor function via heterodimer formation. *Cell* 71, 63–72. doi: 10.1016/0092-8674(92)90266-f
- Zanet, J., Benrabah, E., Li, T., Pelissier-Monier, A., Chanut-Delalande, H., Ronsin, B., et al. (2015). Pri sORF peptides induce selective proteasome-mediated protein processing. *Science (New York, NY)* 349, 1356–1358. doi: 10.1126/science.aac5677
- Zanet, J., Chanut-Delalande, H., Plaza, S., and Payre, F. (2016). Small peptides as newcomers in the control of *Drosophila* development. *Curr. Top. Dev. Biol.* 117, 199–219. doi: 10.1016/bs.ctdb.2015.11.004
- Zipper, L., Jassmann, D., Burgmer, S., Gorlich, B., and Reiff, T. (2020). Ecdysone steroid hormone remote controls intestinal stem cell fate decisions via the PPARGgamma-homolog Eip75B in *Drosophila*. *ELife* 9:e55795.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Dib, Zanet, Mancheno-Ferris, Gallois, Markus, Valenti, Marques-Prieto, Plaza, Kageyama, Chanut-Delalande and Payre. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Chapitre 6 OvoL/Shavenbaby régule des ensembles distincts de gènes cibles en fonction du type cellulaire dans lequel il opère

6.1 Résumé de la publication

Nous avons vu que Svb est requis dans le remodelage tridimensionnel des cellules épidermiques, la prolifération et la survie des cellules souches digestives ainsi que pour le maintien de l'intégrité épithéliale des entérocytes intestinaux (Payre, 2004; Chanut-Delalande et al., 2006; Menoret et al., 2013; Bohère et al., 2018). La maturation de Svb dépendante des peptides Pri pose la question de la spécificité fonctionnelle des deux formes de protéines SvbRep et SvbAct dont les études réalisées dans l'intestin ont mis en lueur leur antagonisme et confirmés leur différence de localisation subnucléaire (al Hayek et al., 2021). La forme répressive, SvbRep, est accumulée sous forme de foci et la forme transcriptionnellement active, SvbAct est présente de façon homogène dans tout le noyau (Zanet et al., 2015; Bohère et al., 2018; al Hayek et al., 2021).

Mon travail de thèse a consisté à caractériser les comportements moléculaires des deux formes de Svb plus en profondeur et à l'échelle du génome. Pour cela, j'ai basé une grande partie de mes études sur l'utilisation de lignées stables issus de la transformation de cellules S2, permettant d'exprimer spécifiquement SvbRep ou SvbAct (Zanet et al., 2015). Dans ces cellules initialement dépourvues de Svb, tous les éléments nécessaires à la maturation de Svb sont présents et SvbAct est capable d'activer des enhanceurs cibles présents dans des épisomes (Zanet et al., 2015). Je me suis en premier lieu attachée en utilisant des approches de génomique (ChIP-seq) et de transcriptomique (RNA-seq) à déterminer le profil de fixation à l'ADN de SvbAct et SvbRep mais aussi à définir quels en étaient ses gènes cibles dans ce type cellulaire.

L'intégration de ces données a mis en évidence que SvbRep et ACT en se fixant sur les mêmes régions cis-régulatrices riche en motif Ovo (OvoQ6 et SvbF7) régulent de manière antagoniste (répression par SvbRep et activation pour SvbAct) 250 gènes cibles directs. Ces gènes présentent une modification de leur expression en présence de Svb et la fixation du facteur de transcription. Bénéficiant des nombreuses données disponibles pour les cellules S2, j'ai caractérisé l'environnement chromatinien dans lequel s'effectue la fixation de Svb. J'ai pu ainsi déterminer que la fixation de Svb s'effectue dans des régions enrichies en H3K4me1, H3K27ac. Ces marques d'histones sont connues pour être associées avec des régions cis-régulatrices (enhancers) actives. Cette analyse a également mis en évidence que la fixation de Svb est contrainte par la structure chromatinienne, Svb ne la modifiant pas.

Dans un deuxième temps, nous avons comparé les groupes de gènes régulés par Svb dans trois contextes cellulaires différents : les cellules S2, les cellules épidermiques embryonnaires et les cellules souches intestinales. Cet aspect du projet a bénéficié de l'arrivée dans l'équipe de Clément Immarigeon (post-doc), qui a réalisé les expériences de RNA-seq au sein des cellules souches intestinales. Pour l'analyse de ces données nous avons adapté le pipeline d'analyse et intégré des approches non-supervisée : du clustering. Ces analyses de clustering ont été réalisés par Alexia Rivero, étudiante en Master 2 Bioinformatique et Biologie des Systèmes que j'ai co-encadrée. Grâce à l'utilisation d'une classification hiérarchique

ascendante et d'un clustering k-moyennes, nous avons déterminé qu'au sein des cellules souches intestinales Svb régulent l'expression de 195 gènes cibles.

La comparaison des trois groupes de gènes cibles de Svb spécifiques de chaque type cellulaire a mis en évidence que Shavenbaby régule des groupes de gènes différents en fonction du contexte cellulaire dans lequel il évolue. Si la structure de la chromatine participe à cette diversité, elle ne peut, à elle seule, expliquer ces différences. En effet, la structure de la chromatine permet de comprendre la différence entre les cellules épidermiques et les cellules S2 mais pas entre les cellules S2 et les cellules souches intestinales pour lesquelles la fixation de Svb est très similaire. Cela suggère qu'un ou des facteurs partenaires de Shavenbaby, encore inconnus, participent à la régulation de l'expression génique dépendante de Svb.

OvoL/Shavenbaby regulates distinct sets of target genes between embryonic and adult stem cells

Alexandra Mancheno-Ferris¹, Clément Immarigeon¹, Alexia Rivero¹, Priscilla Lhoumaud^{1,2}, Julien Anglade¹, Serge Plaza^{1,3}, Olivier Cuvier¹, Cédric Polesello¹ and François Payre^{1*}

¹ Molecular, Cellular and Developmental biology department (MCD), Centre de Biologie Intégrative (CBI), University of Toulouse, CNRS, UPS, 31062 Toulouse, France

Present addresses:

² NYU Langone Medical Center, 55 First Avenue, Smilow 311, New York, NY 10016

³ Laboratoire de Recherche en Sciences Végétales (LSRV), CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, 31326, Castanet-Tolosan, France.

* correspondence: francois.payre@univ-tlse3.fr

ABSTRACT

Background: Adult stem cells are critical to the homeostasis of vital functions and their regenerative/differentiation potential suggest that they resemble embryonic cells. Although developmental signaling pathways play critical roles in stem cells, the question whether adult stem cells reuse embryonic transcriptional programs yet remains elusive. OvoL transcription factors that mediate functions of several pathways during embryogenesis have recently emerged as key regulators of stem cells across animals.

Results: Using *Drosophila* that encodes a single OvoL factor, Shavenbaby, we investigated putative conservation of its targets between embryonic and adult stem cells. Posttranslational modification of Shavenbaby produces two protein isoforms, Svb-ACT and Svb-REP, which display antagonistic transcriptional activities and promote intestinal stem cell renewal or differentiation, respectively. Capitalizing on cell lines engineered to express one or the other isoform, we profiled the genomic repertoire and chromatin context of their binding sites, and identified downstream target genes. We find that, within a given cell type, Svb-ACT and Svb-REP bind the same sites and regulate the expression of a same set of target genes, the choice of which is dominated by chromatin state. We find that unsupervised analyses of gene expression in response to Svb isoforms allow faithful identification of direct target genes. Using this framework *in vivo*, we show that Svb functions in embryonic epidermal cells *versus* adult stem cells likely rely on distinct sets of effectors.

Conclusions

While OvoL factors mediate conserved regulatory cascades, the ultimate mechanisms of their action is gated by changes in chromatin accessibility between tissues, a fundamental feature likely relevant to their implication in various human diseases.

KEYWORDS

OvoL, ovo/shavenbaby, *Drosophila*, adult stem cells, epidermis, transcription factors, enhancers, histone marks

BACKGROUND

In order to regulate expression of their target genes, transcription factors (TFs) bind to specific DNA sequences located within cis-regulatory elements such as enhancers and promoters [1, 2]. TF binding motifs being short (4-15 base pairs) and degenerate, they are very abundant in the genome. However, the accessibility of TF binding sites is influenced by various parameters, including biophysical constraints, chromatin state and organization, presence/absence of cofactors, etc... Therefore, only a limited subset of binding sites is bound by TFs in a given cell type [3-5], and substantial variations in binding events are often observed across tissues [5-7]. These cell-specific changes in binding site occupancy are expected to have strong impact in transcriptional outputs, but both the underlying mechanisms and functional consequences currently remain to be fully elucidated [8].

The family of Ovo-like (OvoL) transcription factors are specific to metazoans [9] and display key functions during animal development, often conserved across species. The founding member, called Ovo/Shavenbaby (Svb), was initially identified in flies [10] for its dual role in epidermal development and germline. Unlike invertebrates, mammalian species have evolved three paralogs referred to as OvoL1-3, which display partly redundant functions [11] and contribute to genitalia and skin development. OvoL1 and 2 have been involved in various skin disorders and diseases, including atopic march and psoriasis (see [12] for review). Recent work has further established the importance of OvoLs within various epithelial tissues, including kidneys, placenta or mammary glands [13]. The deregulation of OvoL expression is associated to most cancers of epithelial tissues [14-16], and has been also involved in pediatric blastoma [17] and bone sarcoma of mesenchymal origin [18]. Mutations in OvoL2 are a major cause of posterior polymorphous corneal dystrophy [19], a dominant genetic disorder resulting from impaired differentiation of the epithelium [20]. It has been proposed that a main function of OvoL factors, both in cornea [20] as well as in many other tissues [14, 16, 21], is to maintain epithelial integrity by preventing epithelial-mesenchymal transition (EMT). Indeed, OVOL1 and 2 have been shown to antagonize the action of pro-EMT factors such as ZEB1, ZEB2 or Twist [15, 22]. In addition, recent studies have demonstrated the importance of OvoL factors in the control of stem cells. OvoL2 regulates the behavior of hair follicle stem cells and is required for skin repair [21, 23], while OvoL1 has been involved in primordial germ cells [24]. OvoL plays a key role in trophoblast stem cells [25, 26], related to proper placenta function in various mammalian species [27]. OvoL2 has also been involved in embryonic stem cells, to regulate the formation of hematopoietic and endothelial cells [28], and a series of recent studies have

further revealed the importance of OvoL factors in the induction and fate of induced pluripotent stem cells (IPSCs). For example, OvoL1 is a direct target of the so-called OKSM cocktail of TFs that trigger transformation of differentiated cells into IPSCs [29]. Furthermore, OvoL1 can substitute for Myc in reprogramming fibroblasts or hepatocytes into IPSCs [29]. Given the wide functions of OvoL factors in development, epithelial organization, cell differentiation and reprogramming, their implication into a broad spectrum of diseases it not surprising [13]. The molecular mechanisms by which OvoL factors fulfil such different functions, in particular the specific repertoire(s) of target genes they regulate between tissues yet remain poorly understood. In mammals, the situation is complexified by the existence of three paralogs, and their DNA binding domain that comprises four highly-conserved zinc fingers binds to the same core recognition sequence CNGTT [30, 31]. Furthermore, the three OvoL genes in vertebrates produce several protein isoforms, mostly differing from the extent of the N-terminal region (the zinc finger domain is located at the C-terminal extremity), which can exhibit distinct transcriptional activities [16].

The simpler *Drosophila* genome encodes a unique Ovo/Svb gene, providing an attractive paradigm to dissect OvoL functions and investigate the mechanisms underlying tissue specific activities. In the germline, alternative promoters drive the expression of two proteins, OvoA and OvoB, which act as a transcriptional repressor and an activator, respectively [32]. In somatic tissues, a separate promoter leads to the production of the Svb protein, well known for its role in governing differentiation of epidermal cells in the embryo, and in adult derivatives [33]. Svb is translated as a large transcriptional repressor, referred to as Svb-REP. Then, under the joint action of Pri smORF peptides (encoded by the *tarsal-less/polished-rice* locus) and the Ubiquitin ligase Ubr3, Svb-REP undergoes proteasome-dependent processing resulting in the production of a shorter transcriptional activator, Svb-ACT [34, 35]. In the embryonic epidermis, Svb-ACT activates the expression of approximately 150 genes mainly encoding terminal effectors which are collectively responsible for the tridimensional remodeling of epidermal cells [36-38].

Here, taking advantage of the absence of Svb in S2 cells, we specifically expressed Svb-ACT or Svb-REP in these cells and performed ChIP-seq and RNAseq in order to: 1) study their respective DNA binding behavior, 2) analyze the influence of the chromatin environment on their binding, 3) identify their direct target genes. We found that Svb-REP and ACT, constrained by chromatin state, bind to the same genomic regions and antagonistically control a common set of 250 direct target genes. Next, we generated RNAseq in adult intestinal stem cells (ISCs) and identified approximately 200 Svb target genes. Comparison of the three sets

of genes regulated by Svb in S2 cells, embryonic epidermal cells and adult intestinal stem cells shows that they are mostly unrelated, suggesting that the transcriptional outputs of Svb is prominently dependent of the cell type in which it operates.

RESULTS

Svb-REP and Svb-ACT bind to the same array of genomic sites

To map the genome-wide set of sites bound by Svb-ACT and Svb-REP, we performed chromatin immunoprecipitation coupled to next-generation sequencing (ChIP-seq) in cultured S2 cells, which do not express endogenous Svb and have been engineered to express one or the other form of Svb. We made use of two stable lines of S2 cells that allow inducible expression of full-length Svb fused to GFP [35], with or without Pri peptides, thus producing Svb-ACT and Svb-REP, respectively. These cell lines are well adapted to dissect Svb function, since they have allowed disclosing the mechanisms of Svb processing in response to Pri peptides [35] and studying Svb-dependent activity of direct target enhancers [34]. We performed ChIP-Seq experiments in cells expressing Svb-ACT::GFP, Svb-REP::GFP and parental S2 cells for control (Figure 1A), using the same anti-GFP antibody to optimize comparison. ChIP-seq results revealed 5,513 regions that are highly enriched for Svb-ACT binding (Figure 1B,C). Consistent with a binding specified by the Svb zinc finger domain, these regions feature DNA motifs matching Ovo/Svb binding matrixes [38], as indicated by *ab initio* search for enriched oligonucleotide sequences using two different methods [39, 40]. These motifs correspond to the so-called OvoQ6 position weight matrix (PWM) representative of OvoL factors (CNGGT). The most enriched motif matches a slightly enlarged sequence ACNGTTA (hereafter referred to as F7-BS) that was defined by an approach that combines statistical analysis with phylogenetic information [41] and is the best predictor of functional Svb binding sites *in vivo* [38].

Although peak calling detected a markedly smaller number of regions for Svb-REP (1351), they extensively overlapped with Svb-ACT bound regions and there was clear enrichment for Svb-REP across all Svb-ACT peaks, well-illustrated when compared to random genomic regions (Figure 1B, S1A,B). Reciprocally, Svb-REP peaks were also strongly enriched for Svb-ACT in corresponding cells, as opposed to a similar set of random regions or genomic regions that have been shifted by 10kb from the peak center (Figure S1A). The smaller number of Svb-REP peaks correlated with lower ChIP intensity (Figure 1C) suggesting that failure of peak detection might be, at least in part, due to lower signal to noise ratio as seen on genome browser snapshots (Figure 1E). These results therefore supported that Svb-ACT and -REP share similar binding sites. To thoroughly test this conclusion, we quantified Svb-ACT vs Svb-REP signals within common peaks (*i.e.*, in regions detected in both Svb-REP and -ACT cells) and in regions corresponding to peaks only detected in Svb-ACT cells. The ratio between Svb-

ACT vs Svb-REP binding levels was indistinguishable between common peaks and ACT-only peaks (Figure S1D), as expected if the difference in peak numbers mainly relates to threshold effects instead of distinct specificity.

Taken together, these results led us to conclude that Svb-ACT and Svb-REP bind nearly identical sets of genomic regions, Svb-ACT cells allowing peak detection at higher sensitivity.

Svb binds to regions bearing hallmarks of active enhancers

It is well-established that functional TF/DNA interactions in eukaryotes occur within regions proximal or distal to transcription start sites, and that nucleosome post-translational modifications regulate DNA accessibility to TF binding. At a first glance to characterize the chromatin landscape of Svb-bound regions in S2 cells, we interrogated whether Svb peaks displayed recognizable patterns of histone marks using the large set of data available in *Drosophila* cells [42-46].

To maximize sensitivity, we analyzed the whole set of Svb-ACT peaks and quantified levels of histone post-translational modifications (PTMs) on these regions. There is a strong enrichment in Histone3 acetylated on K27 (H3K27ac) across Svb-bound sites, which directly correlates with the intensity of Svb signal (Figure 2A,B). H3K27ac has been shown to mark active chromatin, in particular active enhancers [44, 47, 48] that are also enriched in histone 3 methylated on K4 [49, 50]. Consistently, Svb bound sites are enriched in H3K4me3 and H3K4me1, which accumulate on both sides of the peak (Figure 2A,C). Conversely, H3K27me3 that marks repressive chromatin [51] is depleted, while H4K16ac is mostly seen at background levels. Svb peaks are also depleted in H3K36me3 that often associates with gene bodies [52], reinforcing the notion that Svb preferentially binds to active cis-regulatory elements, *i.e.*, promoters and enhancers.

Additional pieces of evidence were provided by bioinformatic analyses. Previous work has allowed high-throughput identification of genomic regions enhancing transcription in S2 cells, using a method called STARR-seq [53]. Over the 5476 active enhancers detected in STARR-seq, 2079 overlap with Svb peaks, showing strong enrichment of Svb in enhancers (Figure 2D,E).

To functionally test this model, we next performed ChIP-seq experiments, using antibodies specific of the histone PTMs identified *in silico*, in control, Svb-ACT and Svb-REP cells. The pattern of H3K27ac, H3K27me3, H3K4me1 and me3 in control cells nicely correlate

with published data sets (Figure 2A, S2A), showing that genome-wide profiling of chromatin marks is robust against variations in experimental conditions, cell isolates and analyses. Importantly, the results fully validated that the binding of both Svb-ACT and Svb-REP occurs in regions with high H3K27ac, H3K4me1/3, while being depleted in H3K27me3 (Figure 2A-C and S2A). In addition, we did not detect differences in histone PTMs at Svb binding sites between the three conditions (control, Svb-ACT and Svb-REP, Figure 2A and S2A,B), therefore suggesting that Svb isoforms are not sufficient to modify the local patterns of histone PTMs in S2 cells.

In sum, these results indicate that data integration approaches draw a fine model of Svb binding on the genome of S2 cells. As further demonstrated by experimental evidence, we conclude that Svb-ACT and Svb-REP behave as settler transcription factors, binding to opened chromatin regions bearing hallmarks of active enhancers.

Identification of Svb direct target genes

To analyze the influence of the two forms of Svb on gene expression, we next performed RNA-seq in the three cell lines used for ChIP-seq. When compared to control conditions, both Svb-ACT and Svb-REP significantly changed gene expression (Figure 3 A and S3A,B), meaning that Svb is sufficient on its own to modify the transcriptome of differentiated cells. *In vivo* dissection of Svb function in embryonic epidermal cells [33] has shown the antagonistic activity of Svb-ACT and Svb-REP on direct target genes [34, 36, 38], leading to potent upregulation or downregulation of those, respectively [36]. We reasoned that direct target genes of Svb in S2 cells should display a similar behavior. Therefore, we selected all genes showing significant (p -value <0.05) differential expression between Svb-ACT and Svb-REP conditions, and focused on those showing coherent changes ($\log_2(\text{FC})>0$ in Svb-ACT and $\log_2(\text{FC})<0$ in Svb-REP conditions). This supervised approach defined a group of 285 genes (Figure 3B) being upregulated by Svb-ACT and downregulated by Svb-REP when compared to controls. We observed that genes exhibiting highest upregulation in Svb-ACT cells grossly correlate with low expression levels in parental S2, and reciprocally genes with strong downregulation in Svb-REP cells are often highly expressed in S2 (Figure 3B). Unsupervised approaches further confirmed our findings, *e.g.*, hierarchical clustering [54] defined a very similar group of 295 genes, the vast majority of those (279) being common to both supervised and unsupervised approaches (Figure S3 C,D). Consistent with the prediction that these genes likely represent Svb direct target genes, they are in most cases associated with Svb binding peaks, as revealed

by complementary analyses. 247 genes (out of 285) of the supervised set contain one or several peaks of Svb binding within a region comprising the transcribed sequence and 5 kb upstream of TSS (Figure 3C), *i.e.*, where most enhancers are found in *Drosophila*. Similarly strong association with Svb peaks was also obvious for the sets of target genes defined by unsupervised approaches (Pval <10E-10, Figure S3 D). The same was also true for the most conservative conditions, *i.e.* a same group of 195 genes were retrieved by the three approaches (supervised, k-means, hierarchical clustering), 86% of these common genes being associated to Svb binding peaks, as defined above.

Taken together, these data show that Svb-ACT and Svb-REP display antagonistic transcriptional activities acting on a common set of approx. 250 genes in S2 cells.

Different binding behavior of Svb for the regulation of different sets of target genes between cell types

Having defined the set of Svb target genes in S2, we next asked how it compared to genes regulated by Svb in embryonic epidermal cells [38], focusing on direct target genes (*i.e.*, being differentially expressed and displaying a binding peak). The two sets of genes in S2 and embryonic epidermal cells were of comparable size (247 vs 149) and both of them comprise various terminal effectors (enzymes, membrane-associated proteins), as opposed to transcription factors or other regulators. Nevertheless, there was no significant overlap between the two sets of target genes, with only three genes common to S2 and embryonic epidermal cells (Figure 4A). This result was unexpected since previous work had demonstrated that Svb-ACT has the potential to upregulate in S2 cells the transcription of *miniature*, an iconic direct target gene of Svb in the embryonic epidermis [36, 38]. This was deduced from strong activation of a *miniature* enhancer (*Emin*) in transient reporter assays [34] that mimics Svb regulation *in vivo* [36]. Therefore, although Svb is sufficient to directly regulate the *Emin* enhancer in S2 cells, it displays no detectable activity on the expression of the *miniature* gene in similar conditions. This behavior was not restricted to *miniature* and, instead, it was observed for all enhancers targeted by Svb in embryonic epidermal cells that we tested in S2 cells. Each of the individual enhancers was strongly activated by Svb-ACT in reporter experiments (Figure 4B,C), likely as a result of a direct interaction, since mutations of Svb binding sites within the *Dyl2* enhancer (*Dyl2mt*) decreased its Svb-dependent activity (Figure 4C). We noticed that none of the corresponding genes was significantly expressed in S2 cells, in any conditions (control, Svb-ACT, Svb-REP), suggesting a broad inhibitory mechanism to explain the different

behavior observed between individual target enhancers, which are responsive to Svb, and the corresponding genes, which are not. As seen on genome browser snapshots, the prominent peaks of Svb binding to enhancers that occur in embryos were not observed in S2 cells (Figure 4D). Quantification of ChIP signal showed that experimentally-defined embryonic enhancers displayed strongly reduced binding of Svb in S2 cells when compared to embryonic epidermal cells (Figure 4E). On a broader scale, the majority (86%) of Svb-bound regions in embryonic epidermal cells were not bound in S2 cells (Figure 4F). Hence, these data suggested that embryonic target genes cells are insensitive to Svb within the genomic context of S2 cells because it prevents Svb binding. In contrast, individual enhancers within the episomal context of transient transfection assays might display a more relaxed chromatin organization that allows accessibility to Svb binding.

To get a closer look to the chromatin context of Svb binding, we focused on evolutionarily conserved F7-BS motifs that often mediate transcriptional regulation of Svb direct targets *in vivo* [38] and represent the most enriched motif in S2 Svb peaks (Figure 1D). There are 2354 conserved F7-BS in the *Drosophila* genome, 26% of those being bound by Svb in S2 cells (Figure 5 A). We analyzed whether F7-BS overlapped with specific histone PTMs in cultured cells within Svb-bound versus unbound regions. As observed for the whole set of Svb ChIP peaks (see Figure 2), F7-BS bound by Svb in S2 cells were strongly enriched in H3K27 and H3K4me1,3, when compared to unbound F7-BS regions (Figure 5B). We next analyzed the binding of Svb to F7-BS in embryonic epidermal cells. 323 F7-BS were bound both in S2 and embryonic cells, while 581 were bound only in embryonic cells. The set of F7-BS bound in S2 cells and embryos also presented a clear enrichment in H3K27ac, H3K4me1 and me3 (Figure 5 D), as expected for active enhancers within accessible chromatin. In contrast, F7-BS regions that were never bound, showed low levels of H3K27ac, H3K4me1 and me3, while all of them overlapped with the inactive H3K27me3 mark. Whereas F7-BS motifs that were specifically bound in S2 cells also displayed the features of active enhancers, sites bound by Svb only in the embryonic epidermis exhibited closed chromatin configuration in S2 cells (Figure 5 D). These data show that the occupancy of Svb binding sites correlates with histone PTMs, further supporting that Svb-responsive embryonic enhancers are not bound in S2 cells because of chromatin inaccessibility.

Together, these results indicate that the Svb transcription factor regulates strikingly different sets of target genes between S2 cells and embryonic epidermal cells. Specific patterns of histone marks gate the pool of binding sites that are accessible to the Svb transcription factor

in a given cell type, and thereby appear to dominate the choice of which target genes are regulated in response to Svb.

Identification of Svb target genes in intestinal stem cells

While the absence of endogenous expression of Svb in S2 cells was helpful to address the mode of binding and transcriptional effect of each isoform, the extent to which this experimental model was representative of the choice of target genes *in vivo* remained to be established. To further test the hypothesis that, *in vivo*, Svb might regulate different pools of target genes between the tissues in which it has a function, we next investigated adult stem cells.

Recent work has shown that *svb* is required for the homeostasis of both renal [55] and intestinal adult stem cells (ISCs) [56]. Within ISCs, Svb-ACT favors ISC proliferation and renewal, while Svb-REP triggers their differentiation into enterocytes [56], the main population of the intestinal epithelium. To identify genes regulated by Svb in ISCs, we performed RNAseq analysis of stem cells purified by fluorescent assisted cell sorting (FACS) from dissected guts of control animals, and upon RNAi-mediated *svb* depletion, or expression of Svb-ACT and Svb-REP. Manipulation of Svb function was specifically targeted in adult ISCs, using the well-controlled *esg-Gal4; tub-Gal80^{ts}* (*esg^{ts}*) system [57]. Approx. 8,000 genes were found significantly expressed in ISCs in at least one condition. When compared to control, each perturbation of Svb activity was sufficient to alter the repertoire of genes expressed in ISCs (Figure 6A). Building on the results of S2 cell analyses (Figure 3 and S3), we used unsupervised approaches to identify the set of genes whose expression is likely directly regulated by Svb. Hierarchical clustering highlighted a group of 272 genes, exhibiting higher expression levels in Svb-ACT ISCs and lower levels following Svb-REP or *svb*-RNAi expression (Figure 6B). A second unsupervised algorithm, k-means [58] that operates nonhierarchical clustering using vector quantification, also robustly retrieved a similar group of 201 genes (Figure 6B), the vast majority of them (195) being common with genes identified by hierarchical clustering (Figure 6C). For both sets of genes, they were upregulated by Svb-ACT and downregulated by Svb-REP (Figure 6B), showing the antagonistic activity of the two isoforms on gene expression in ISCs. That their expression was also downregulated upon *svb* knockdown further show that these genes are normally expressed and activated by Svb-ACT in ISCs, consistent with previous conclusion that Svb is processed and predominantly behaving as an activator within ISCs [56]. The group of 195 genes common to both unsupervised approaches therefore represents an high confidence set of putatively direct Svb target genes in ISCs. Most genes

encode putative cellular effectors, including 66 enzymes, 43 membrane proteins, and 25 proteins related to cytoskeletal organization, while there are only five transcription factors. In addition, 28 genes are involved in signaling pathways (EGFR, Wnt, Hippo, BMP, Notch and Robo) that are well known for their role in the regulation of ISC behavior (see [59] for recent review), as well as 7 genes related to the cell cycle. Among the latter, Svb-ACT activates the expression of *string* (encoding *Drosophila* Cdc25), *Cyclin-B* and *Cyclin-E*, that could contribute to explain its impact on ISC proliferation [56]. Reciprocally, the downregulation of such regulators by Svb-REP might favor cell cycle exit and commitment towards differentiation.

These data show that the experimental setup and analysis pipeline optimized in cultured cells are directly relevant to address the function of Svb-ACT and Svb-REP *in vivo*, within adult ISCs. They further suggest that these approaches provide a first glimpse into the set of genes which mediate Svb activity for the regulation of adult stem cells.

Comparing the cohorts of Svb target genes

To further explore putative differences in the behavior of Svb between tissues and cell types, we next sought to compare the three cohorts of Svb-regulated genes in S2 cells, the embryonic epidermis and adult ISCs.

The set of Svb target genes in embryonic epidermal cells was previously established by supervised approaches [38]. As a first step, we re-analyzed expression data in the embryo for control conditions, *svb* mutant or *pri* mutant (which at least in part mimics a gain in Svb-REP function), using the same unsupervised approach to allow direct comparison with S2 cells and ISCs. To maximize sensitivity, we used hierarchical clustering that appeared less restrictive in both S2 cells (Figure S3) and ISCs (Figure 5B). When applied to embryonic expression data, this defined a group of 761 Svb target genes (Figure S4 A,B), which contains 142 out of the 149 direct targets defined previously [38]. As also observed in S2 cells, these results in embryos cross-validated supervised and unsupervised approaches, further establishing their suitability as means to identify Svb target genes. Although hierarchical clustering provided a significantly larger number of *svb* target genes (761) in the embryo, we observed very limited overlap between the three sets of genes identified by the same approach in S2, embryonic epidermis, and adult stem cells (Figure 7A). On average, less than ten percent of Svb target genes were common to two conditions, well supporting the notion that cohorts of genes regulated by Svb are predominantly distinct between cell types. Only 4 genes were common to the three sets (Figure 7A) and they do not correspond to genes showing the highest

variations between control and manipulation of Svb activity. For example, *GstD1* displays relatively modest deregulation upon *svb* manipulation ($\log_{FC} < 1$) in any of the three cell types (Figure 7B). Of note, Svb binds to the *GstD1* locus in both S2 and embryonic epidermal cells, even if the observed ChIP-seq peaks only partially overlap. *GstD1* is also expressed in ISCs and upregulated upon Svb-ACT expression (Figure 7B). Consistently, analysis of available ATAC-seq data [60], which allows mapping of open chromatin regions, further suggested that cis-regulatory regions of *GstD1* should be accessible to Svb binding within ISCs (Figure 7B). Additional evidence support that chromatin accessibility is playing a role in selecting which target genes are regulated by Svb in a given tissue. The proportion of common target genes between S2 and ISCs is more significant than any of the two with embryos, and we often observed a better correlation in accessibility to Svb binding sites, as deduced from comparison of ATAC-seq signals between S2 cells and ISCs [60, 61]. For example, *string* was strongly upregulated by Svb-ACT in ISCs ($\log_{FC} > 2.9$) in which Svb binding regions were accessible (Figure 7C). In contrast, Svb did not bind to the *string* (*stg*) locus in embryos, and *stg* levels were not modified in the absence of Svb, or upon blocking it processing into the activator form. In S2 cells, cis-regulatory regions of *stg* were both accessible (ATAC-seq) and bound by Svb (ChIP-seq). However, *stg* expression levels (RNA-seq) were not modulated by Svb-ACT as they were in ISCs (Figure 8C), showing that if accessibility to Svb binding is a prerequisite for regulation, other aspects also contribute to select cell-specific sets of target genes.

In conclusion, our results indicate that the sets of genes regulated by Svb are strikingly different between epidermal cells *versus* intestinal stem cells. The specificity is achieved through both changes in chromatin states (which modulate the accessibility of Svb binding sites) and other cell-type specific features (which allow actual gene regulation upon Svb binding). Although the two sets of Svb targets appear enriched in genes encoding putative cellular effectors, such as membrane proteins or cytoskeletal remodelers, we concluded that Svb functions in embryonic and adult stem cells rely on largely distinct sets of those.

DISCUSSION

OvoL/Svb TFs are emerging as a conserved family of regulators required for the development of many epithelial tissues, playing key roles in the control of adult stem cells across species, and being involved in a growing number of human diseases [12, 13, 62]. How the various isoforms encoded by OvoL/Svb genes can exert such a wide range of functions from embryogenesis to adult stem cell homeostasis yet remains poorly understood. Our studies in *Drosophila* provide first insights into isoform- and tissue-specific activity of Svb. Results from high-throughput profiling of Svb binding sites and downstream effectors reveal that the two isoforms generated by proteasome processing possess antagonistic transcriptional functions, on common sets of target genes which display unexpected diversity between embryonic and adult stem cells. We further show the importance of histone post-translational modifications in the selection of specific sets of genes regulated by Svb between tissues, and discuss the relevance of these findings towards a better understanding of OvoL/Svb TFs in normal and pathological situations.

Svb-REP and Svb-ACT bind to the same genomic set of cis-regulatory regions

TFs are the main determinants of genome expression, thereby establishing the specific sets of genes responsible for the acquisition of different cells fates, differentiation programs and behaviors. How TFs select their target genes is a long-lasting question, and the code that mediates regulatory interactions between TFs and their target cis-regulatory sequences remains to be fully elucidated [4]. Although the DNA-binding domain reads the genome to select cognate recognition motifs, only a relatively small fraction of these sites is bound in a given cell type. Additional domains regulate transcription positively or negatively by recruiting nuclear co-factors, which can contribute to the selection of effective binding sites. The two isoforms of Svb in somatic tissues result from post-translational processing; therefore, they differ from each other only by a N-terminal region of 444 amino acids that contains the repressor region specific to Svb-REP (1351 aa), which is partially degraded by the proteasome to release Svb-ACT (910 aa) [34]. Because both isoforms derive from the same transcript, elucidating the specific role of Svb-ACT and Svb-REP is not easily amenable to genetic analysis. If maturation into Svb-ACT is indispensable for the remodeling of epidermal cells that form trichomes [34, 35, 38, 63], whether transient accumulation of Svb-REP that occurs in presumptive trichome cells at early stages of embryogenesis plays a role in epidermal

differentiation is unknown. In addition, Svb isoforms display strikingly different nuclear patterns: while Svb-ACT exhibits diffuse nucleoplasmic distribution, Svb-REP accumulates in discrete nuclear foci [34, 35], which could for instance represent storage bodies of therefore merely inactive pools of Svb-REP precursor. In contrast, we recently reported that both Svb-ACT and Svb-REP are separately required in the intestinal stem cell lineage: while Svb-ACT promotes renewal and proliferation of early precursors (ISC and enteroblasts), Svb-REP acts at later stages of the lineage to maintain the differentiation of enterocytes [56]. Taking advantage of cultured cell lines engineered to specifically express Svb-ACT or Svb-REP [34, 35], our studies have allowed a detailed molecular profiling of the respective activities of each of the two isoforms. The results demonstrate that Svb-REP is capable of effective DNA binding and transcriptional regulation, refuting the hypothesis that it might behave solely as an inactive precursor. Although the number of Svb-REP ChIP peaks detected by MACS is substantially smaller than that of Svb-ACT, further analyses suggest that both isoforms basically bind to the same array of genomic sites. In addition, the genomic pattern of binding events is likely relevant to Svb transcriptional function, since Svb binding is strongly enriched in regions acting as enhancers in S2 cells [53]. This conclusion is also supported by the antagonistic transcriptional activities of the two isoforms on the same set of target genes in S2 cells. We observed that the respective amplitude of Svb-ACT and Svb-REP effects on transcription of downstream targets depends on their basal expression in parental S2 cells (Figure 3B). It is likely that genes exhibiting barely detectable expression levels in controls are less sensitive to Svb-REP repression, and/or that their changes cannot reach the significance threshold. Reciprocally, genes expressed at very high levels could attain a transcription plateau, which cannot be pushed further up by Svb-ACT, while being efficiently downregulated by Svb-REP. The antagonistic action of Svb-ACT and Svb-REP *in vivo* is further well demonstrated in ISCs, in which both isoforms impact positively or negatively on a same set of downstream targets.

Our cultured cells model has also allowed fine investigation of the chromatin context of Svb activity within the *Drosophila* genome. The results show that the subset of binding sites that are occupied by Svb-ACT and Svb-REP in S2 cells is heavily constrained by histone PTMs. A growing body of evidence support those enhancers bear specific patterns of histone marks. For example, mono- or tri-methylation on lysine 4 of histone 3 is often associated to cis-regulatory regions, promoters and enhancers. Recent results from supervised machine learning suggest that H3K4me3 could help discriminating proximal promoter elements, while H3K4me1 metaprofiles are predictive of enhancers, even if they are inactive in the cells analyzed [50]. In addition, active enhancers are enriched in H3K27ac, while H3K27me3 marks

and/or contribute to silent inactive enhancers. Consistent with the association of Svb with active enhancers, Svb binding sites in S2 cells are enriched in H3K27ac, strongly depleted in H3K27me3, and enriched in H3K4me1,3. Several pieces of evidence are further compatible with an higher enrichment in H3K4me1 compared to H3K4me3 levels at Svb binding sites, albeit collecting additional evidence will be required to reach a firmer conclusion. Of note, we observed virtually identical results when analyzing, on the one hand, published S2 data sets and, on the other hand, results from our experimental profiling of histone PTMs. These findings illustrate the potential of integrating heterogenous data to interrogate the relationships between regulation of gene expression and chromatin organization. Importantly, the binding of either Svb-REP or Svb-ACT does not lead to detectable modification of histone PTMs, the patterns of chromatin marks being indistinguishable at Svb binding sites between cell lines expressing, or not, one or the other isoform. We conclude that Svb binding is dominated by chromatin organization and that, in turn, the binding of Svb isoforms does not modify local chromatin environment.

Studies in *Drosophila* embryos have had strong impact in unraveling the complex bidirectional relationships between sequence-specific TFs and chromatin conformation. They suggest that unlike most TFs, some factors can overcome the nucleosome barrier and are capable of binding to, and increasing accessibility of, closed chromatin - as exemplified by Zelda during zygotic genome activation [64-66]. It is possible that the distinction between the former (settler) and the latter (pioneer) factors is not absolute and the ability for a TF to bind regions in closed chromatin may vary depending of the cell context, *e.g.*, the availability of chromatin remodelers such as histone readers, writers or erasers. For example, one cannot exclude that the absence of influence of Svb on local histone PTMs (observed both for the genomic set of binding sites and when focusing on binding sites associated with regulated genes) is contributed, at least, in part by the absence of specific Svb partners conveying chromatin remodeling activity. An additional parameter to consider is that our experiments relied on short time inducible expression of Svb (16h), and the effect of longer exposure to Svb-ACT and Svb-REP remains unknown.

Distinct sets of target genes between embryonic and adult stem cells

Investigating the repertoire of direct downstream targets, and their evolution across cell types, still remains a challenge. On the one hand, profiling gene expression is a powerful approach, now amenable to single cell profiling, but identifying the set of target genes among direct and

indirect consequences of for instance inactivation of a TF is not straightforward. On the other hand, genome-wide profiling of binding events is well-adapted to assays in cultured cells, but often difficult to apply to a specific population of cells of low abundance within the whole organism. Therefore, there is a paucity in systematic studies assaying which genes are regulated by the same TF across embryonic cells of different functions and/or at different developmental stages. Our work in cultured cells shows that automated analyses of gene expression modifications upon reciprocal manipulations of a TF, here comparing respective consequences of Svb isoforms, allows identifying a high-confidence set of direct target genes. For instance, unsupervised approaches (hierarchical and k-means clustering) identified a robust set of target genes, since there is remarkable overlap between the two gene clusters and with results of our supervised approach. It is worth noting that the two clustering methods are not purely independent, since we used results of hierarchical analysis to infer the optimal number of clusters to consider in k-means (the k number). Nevertheless, the two methods are based on fundamentally distinct algorithms, as we observed only marginal influence of changes in k numbers on the set of Svb target genes identified by k-means analysis (data not shown). Importantly, both clustering approaches allowed very good prediction of Svb direct targets, reaching 98% (242 out of 247) and 68% (167 out of 247) of detection in hierarchical clustering and k-means, respectively. Building on these results, we deployed the same approaches to identify Svb target genes *in vivo*, using published data in embryos [38], and novel data we generated using RNA-seq analyses of ISCs. For the embryo, our analyses predicted a cluster of genes which contain almost the whole set of direct target genes previously identified *in vivo* [38], *i.e.*, 95% (142 out of 149 genes). Nevertheless, hierarchical clustering leads to a relatively large set of candidates, significantly larger than the sets predicted in S2 cells and ISCs (761, *versus* 295 and 272, respectively). Therefore if hierarchical clustering appears pretty efficient in sensitivity (2% of false negative), the extent to which it performs in selectivity (proportion of false positive) is unknown. It is possible that the actual number of Svb targets might be substantially larger in embryos than in S2 or ISCs. There are however possible caveats to consider when trying to estimate the efficiency of unsupervised clustering in predicting the number of Svb embryonic targets. First, transcriptomic assays were performed in whole embryos, while ChIP-seq was focusing on epidermal trichome cells [38]. Therefore, ChIP may have missed at least a proportion of actual targets, if they were expressed in non-epidermal cell populations. Second, unlike S2 and ISCs, we had no access to consequence of Svb-ACT overexpression, which could substantially decrease the quality of expression-based prediction of target genes. Finally, the variability inherent to gene expression profiling on whole embryo samples can also be an issue.

In contrast, we expect that data obtained from FACS-sorted ISCs would allow faithful prediction of the set of genes regulated by *Svb* in adult intestinal stem cells. As also observed in S2 cells, the predictions of both clustering algorithms are markedly convergent (195 common genes out of 201 and 272 genes). Again, k-means seems more conservative and very robust (the analysis was run 15 times, and it predicted an identical cluster of genes in all but one computations). Whereas very few predicted genes are common to ISCs, S2 and embryonic cells, there is a limited but significant ($p\text{-value} < 10E\text{-}5$) overlap between ISCs and S2 cells. This might be related to some aspects of proliferation potential common to S2 cells and ISCs, while embryonic epidermal cells are considered post-mitotic even before they engage their differentiation. In any case, we interpret these common genes as evidence reinforcing the effectiveness of our approaches. These results also suggest that S2 cells might provide an attractive paradigm to further dissect the molecular mechanisms of *Svb* function on the direct regulation of downstream effectors, truthfully relevant to understanding their role within ISCs. Additional pieces of evidence are consistent with the conclusion that unsupervised clustering defined *Svb* target genes directly involved in the control of ISC behavior. Indeed, several members of this gene cluster have been previously demonstrated to regulate ISCs. They include *Piwi*, *string/Cdc25* and components of EGFR, Wnt and integrin pathways EGFR [67-70].

One unexpected result is the observation that seemingly unrelated sets of downstream genes are likely at work to mediate *Svb* function between embryonic and adult stem cells. This strikingly contrasts with the extensive conservation of the mechanisms driving *Svb* expression in response to signaling pathways in both embryonic epidermal cells [33] and ISCs [56]. There is a large array of enhancers (spanning over a large region >100kb) that control *svb* transcription in the embryo [71]. Several embryonic enhancers display pleiotropic function throughout development [72] and are also driving *Svb* expression in adult stem cells, both in the kidney [55] and gut [56]. Furthermore, we found that the dual activity of *Svb* enhancers in the embryo epidermis and adult gut relies on direct control by same regulatory TF, including TCF that mediates response to Wnt signaling, and *Pnt*, a nuclear effector of the EGFR pathway [56]. There is also conservation of the role of *Pdm1*, a POU-homeobox TF [73] that directly controls a second *svb* enhancer, again both in embryos and adult intestinal cells [56].

OvoL and Epithelial plasticity

Epithelial plasticity is a reversible process that allows bidirectional progression from epithelial to mesenchymal states. It plays key roles during development and is also often hijacked in cancers, promoting tumor invasion and metastasis [74, 75]. Epithelial stem cells display an intermediate or hybrid state [76], characterized by mixed epithelial and mesenchymal features. A core of TFs (Twist, Zeb and Snail) is well-known to promote epithelial to mesenchyme transition (EMT) [75]; OvoLs have been described for their ability to counteract pro-EMT TFs and thus to stabilize epithelial state [13, 14]. Although the mutual antagonism between OVOL and ZEB has been documented in several instances [77], the underlying mechanisms are not fully elucidated. It has been proposed that OVOL2/ZEB directly repress each other and, thereby, would provide a tunable control of epithelial *versus* mesenchymal phenotypes [22, 76]. Zfh1, the *Drosophila* homolog of Zeb1 is also acting in the intestinal stem cell lineage, opposing to precursor differentiation into enterocytes [78, 79]. While Zeb1 might antagonize Svb-REP for enterocyte differentiation, Svb-ACT synergizes with Zfh1 within ISCs [56]. Consistently, we did not detect effects of Svb alterations on Zfh1 expression in ISCs, as deduced from gene expression profiling. An additional iconic target of OvoL TFs in mammals is E-cadherin [13], which again was not detected in the different sets of genes regulated by Svb, neither in ISCs, nor S2 or embryonic cells. Whether the sets of OvoL downstream targets have evolved across species, as observed between *Drosophila* tissues is not known. It is also possible that E-cadherin activation is not representative of OvoL functions, across cell types, including in mammals. We hope that our findings will prompt additional studies to determine putative inter- and intra-specific variations in the set of OvoL target genes in the future.

METHODS

S2 cells lines:

Drosophila S2 cells were grown in Schneider medium supplemented with 10% fetal calf serum and 1% penicillin/streptomycin (Invitrogen). We used cell lines stably transformed with the copper-inducible constructs pMT-Svb::GFP and pMT-pri ("SvbACT"), or pMT-Svb::GFP and pMT-pri1-4fs ("SvbREP"), which encodes a frame-shifted variant as a negative control¹. The expression of pMT was induced by CuSO₄ at the final concentration of 1 mM for 24h.

ChIP experiments:

Cells are cross-linked with 0.8% of formaldehyde for 10 min at room temperature. Cross-linking was stopped by adding glycine (2M) and cells were washed twice in cold with 1X PBS 1 mM NaBu. Cells were resuspended in 500 µL of ChIP permeation buffer (1X PBS, 0.2% Triton, 10mM NaBu), and incubated for 20 min at room temperature. Cells were washed twice with Lysis Buffer no SDS (140 mM NaCl, 15 mM HEPES pH 7.6, 1 mM EDTA, 1% Triton, 0.1% NaDOC, 0.5 mM DTT, 10mM NaBU, 25X protease inhibitor) during 5min with gentle rotation at 4°C, then cells are lysed on Lysis Buffer 1% SDS (Lysis Buffer no SDS + 1% SDS, 0.5% N-lauroylsarcosine) during 30min at 4°C with gentle rotation. Chromatin was sonicated in Bioruptor (Diagenode) using high-power settings, intervals of 30s burst/pause for six cycles to obtain fragments of ≈ 250 pb. The sonicated chromatin was diluted 10 times with Lysis Buffer (no SDS) on protein low bind tubes (Eppendorf). The chromatin fraction was cleared by centrifugation (16000 g, 10 °C, 5 min), then pre-clared (incubed with empty beads) O/N at 4°C. Beads were washed and blocked with 0.1 mg/mL BSA in Lysis Buffer 0.1% SDS, 0.5% N-lauroylsarcosyl during 2hours at 4°C with gentle rotation. Beads are then bound to the antibody O/N at 4°C with gentle rotation. For one immunoprecipitation 20µL of blocked beads, 5µL of antibody and 400µL of Lysis Buffer 0.1% SDS (Lysis Buffer no SDS + 0.1% SDS, 0.5% N-lauroylsarcosine). Immunoprecipitation was performed for 4 hours at 4°C with gentle rotation with 5µl antibody, by adding 100µL pre-washed beads to 500µL cleared chromatin in protein low bind tubes. 2% of each sample was kept as input. Chromatin-bead complexes were eluted twice in the same tube by incubating the beads for 20 minutes at 70 °C first in 10 mM EDTA, 1% SDS, 50 mM Tris-Cl pH8, and then with TE, 0.67% SDS. Samples were incubated overnight at 65°C to reverse the cross-link. Decrosslinked samples were treated with 1µL RNase A during 45min at 37°C then with proteinase K solution (35mg/sample glycogen,

100µg/sample proteinase K) during 2h at 55°C. DNA fragments were purified by phenol-chloroform protocol and resuspended in TE.

ChIPseq sequencing and analysis:

Data of ChIPseq for H3K4me1 (GSE63518), H3K4me3 (GSE63518), H3K27ac (GSE36374), H3K27me3 (GSE93100), H3K9me2(GSE47229), H4K16ac (GSE94115) were used to determine Svb binding landscape in S2 cells.

ChIPseq of Control, SvbREP and SvbACT S2 cells (single-end 50 nt) were performed with HiSeq 200 (Illumina) at BGI. Histones ChIPseq (paired-end 50 nt) was performed at Montpellier GenomiX platform on NovaSeq 6000 (Illumina) with NovaSeq Reagent Kits (300 cycles). The histones ChIPseq bank constructions were also done by Montpellier Genomix platform with TruSeq® ChIP Sample Preparation (Illumina).

For all conditions, reads were mapped on dm6 genome (r6.13 flybase) with bwa (v0.7.17-r1188) using default parameters, unique reads were filtered and peak calling was done on replicates merge files with MACS2² for histone marks from databases (--nomodel --broad --f BAM --g dm --B --q 0.00001) or for histone marks generated in this study (--nomodel --f BAMPE --g dm --B --q 1 e-05 --broad) and (-f BAM --B --g dm --to-large --nomodel --q 0.0001 (Sensible condition) or --q 0.1 (Stringent condition)) for Svb forms. Metrics and quality of ChIPseq peaks were calculated with fastQC (v0.11.8) and CHIPQC package³ (v 1.26.0). To compare set of SvbACT and SvbREP peaks ChIPpeakanno⁴ (v3.24.2, option connectedPeaks = "keepAll") package was used. The statistical relevance of the peaks overlap was tested with the ChIPpeakanno permutation test. The theoretic population of peaks was done with the SvbACT characteristics.

ChIPseq signal was normalized by taking into account the genome mapping with bamCoverage⁵ (--binSize 10 --normalizeUsing RPKM --effectiveGenomeSize 125464728), then normalized for the noise due to the experiment with bigwigCompare (-b1 experiment file -b2 input file --scaleFactors 1:1 --operation "log2"). Because SvbREP is less expressed than SvbACT in S2 cells, we scaled peaks signal by the result of the ratio of logFC for *svb* in the two conditions (scaleFactors 1.41:1).

Heatmaps of signal enrichment were performed with computeMatrix (-bS 100 -a 3000 -b 3000) and plotHeatmap (deeptools⁵ v 3.5.0). RSAT^{6,7} was used to determine the motif enrichment in Svb peaks and we compare with the RSAT for non-redundant insect database.

Enrichment analysis for histone marks around SvbF7 motifs (+3kbp upstream and downstream) was made using deeptools fisher⁵ software.

Coverage analysis were done with MultiBamSummary and plotCorrelation of deeptools⁵ software.

Venn diagrams between the Svb peaks and the different region populations (enhancer, svbF7 regions) were made with ChIPpeakanno (with the option connectedPeaks = "keepAll").

RNA Extraction, Sequencing and Analysis on S2 cells:

RNAseq experiments were made on control, SvbACT and SvbREP cells. Cells were harvested and RNA extraction was made with RNeasy Kit (Qiagen). Bank of reads and sequencing were done by HiSeq 200 (Illumina) at BGI to obtain 29-34M of reads (single-end 50nt) per replicate. The quality of sequencing was measured with fastQC (v0.11.5). Mapping was done with STAR⁸ (v 2.5.2b, default parameter) on *Drosophila* genome dm6 (r6.13 flybase), reads were counted with HTseq-count⁹ (v0.6.0, -t gene -r pos -i gene_symbol) and statistical analysis were performed with edgeR¹⁰ using the negative binomial generalized log-linear model to the read counts for each gene (glmLRT) and decideTestDGE to determine differentially expressed genes. Interactive graphs were obtained with Glimma¹¹ (v 2.0.0) and Plotly¹² (v 4.9.3).

Identification of direct target genes was made with GenomicRanges¹³ (v 1.42.0) and superExactTest¹⁴ (v 1.0.7) and GO terms analyses was done with GOrilla¹⁵.

For the clustering analysis cf. Clustering method section.

Assignment of a gene to a peak:

The distance between the coordinate of gene and the gene was calculated with distanceToNearest() of GenomicRanges package (v 1.42.0, options: ignore.strand = T,select = "all") (Lawrence et al, 2013)

Transfection:

S2 cells were plated in six-well plates (1.75×10^6 cells/3ml) and transfected using 100 μ l of Opti-MEM (Invitrogen) with 3 μ l of Fugene^{HD} (Promega) and a mixture of indicated constructs (total

DNA quantity: 1µg). The following plasmids were used: pAc- OvoB::Myc, pAc- parB::mCherry, Actn-eGFP, Dly2-eGFP, Dly2mut-eGFP, Emin-eGFP, Nyo1-eGFP, Sha1-eGFP, Sha1-eGFP, Sha3-eGFP, SnE1-eGFP and Tyn2-eGFP. Cells were fixed for 20 minutes in 4% Paraformaldehyde, washed twice in PBS 1X and permeabilized in PBT 0.1% (PBS1X Triton X100 0,1%). After an incubation in PAT 0.1% (PBS1X TritonX100 0.1%, BSA 0.1%) cells were stained with anti-GFP rabbit (1/250) and anti-myc mouse (1/250). GAR-Alexa 488 nm (1/400) and GAM-Alexa 555 nm (1/400) secondary antibodies were used. Nuclei were stained with Topro 3(1/2000). Confocal microscopy analysis has been done with ImageJ.

Gut dissection and intestinal stem cells dissociation:

esg-Gal4, UAS-GFP; tubGal80^{ts}/SM66B flies were crossed to the following strains: *Canton-S* (BL#64349), *UAS-SvbRNAi* (VDRC# 41584), *UAS-SvbACT::GFP¹⁶*, *UAS-Svb3Kmut::GFP¹⁶* Adult progeny (only virgin females) were dissected after 6 days of induction at 29°C.

Batches of >20 virgin females are CO₂-anaesthetized and put on ice. From then, all solutions are used cold and samples kept on ice whenever possible. Their midgut are dissected in 1XPBS and transferred to a clean PBS dish. Malpighian tubules and hindgut are removed and only the posterior part of the midgut is kept (R3, R4, R5). PBS was replaced by PBS1X-EDTA 1mM. Posterior midguts were transferred on a glass slide in a droplet of PBS, and shrunk into pieces using a razorblade. Samples were transferred into a fresh tube and spun 3-5minutes at 1000rpm 4°C to pellet the tissue and remove all supernatant except 50µL. 200µL TrypLE 10X (8X final with 0.2mM EDTA) was added for enzymatic digestion. Then the guts were incubated at 37°C and rocked vigorously for 10minutes. Digestions were stopped with 750µL SSM (Serum Supplemented Medium = S2 Cell culture medium). Trituration was made by pipetting up and down using low binding, flame-rounded narrow tips (5 times with narrow tip; 2-3 times with very narrow tip). Strain the cell solution (70µM nylon cell strainer, pre-wet with SSM) into a falcon tube. The tube and the strain were rinsed thrice with 1mL PBS-EDTA. The quality of the dissociation and the health of the cells in solution can be checked under a fluorescent microscope. Samples were spun at 700rcf, 4°C, for >10 minutes and supernatant carefully removed except 100/200µL. 200µL PBS + 3µL clean NGS were added to obtain final samples for FACS. For each cell sample a tube was prepared with 300µL Tissue and Cell Lysis solution + 1µg Proteinase K [50µg/µL]. Cells are sorted in this buffer, vortexed and placed on ice just after sorting. RNA extraction was started less than 1 hour after sorting.

RNAseq, Sequencing and Quality Check on ISC:

RNA are purified from each cell sample using MasterPure RNA purification kit (Epicentre), following manufacturer's recommendation. Sorted cells were incubated at 65°C with rocking for 15minutes. Samples were put on ice then total nucleic acid were precipitated. Contaminating DNA was removed using DNase, then nucleic acids were precipitated again. RNAs were finally resuspended in 10uL TE and 1uL RiboGuard (RNases inhibitor) and stored at -20°C.

Construction of RNA banks and sequencing was made at the Montpellier GenomiX Platform. RNA banks were done with the Ovation SoLo RNAseq System kit (Nugen). Bank validation was done with the quantification of the complementary DNA with the Standard Sensitivity NGS kit on Fragment Analyzer and with qPCR (ROCHE Light Cyclers 480). 100nt single-reads sequencing was done on NovaSeq 6000 (Illumina) with NovaSeq Reagent Kits (100cycles).

The quality of sequencing was measured with fastQC (v0.11.5). Mapping was done with STAR⁸ (v 2.5.2b, default parameter) on *Drosophila* genome dm6 (r6.13 flybase), reads were counted with HTseq-count⁹ (v0.6.0, -t gene -r pos -i gene_symbol) and statistical analysis were performed with edgeR¹⁰ using the negative binomial generalized log-linear model to the read counts for each gene (glmLRT) and decideTestDGE to determine differential genes. Interactive graphs were obtained with Glimma¹¹ (v 2.0.0) and Plotly¹² (v 4.9.3).

For the clustering analysis cf. Clustering method section.

Clustering method:

Genes stably expressed in all conditions were filtered (difference of read < 50 between conditions). For each condition, the mean RNAseq count for the replicates was calculated, centered and reduced.

We used two different algorithms for clustering using stats() package of R¹⁷. The function hclust() for the hierarchical ascendant clustering with the Spearman distance and the Ward.D2 option. The function kmeans() for the k-means clustering with default option. We repeat the k-means clustering 15 times (5 times with predefined centroids and 10 with random centroids). The predefined centroids are genes of different clusters defined by hierarchical clustering. Then we compared the different k-means clusters obtained by each iteration with

cluster_similarity (Clusteval (version 0.1))¹⁸. We defined that the clustering is similar if the score equals 1.

From sorted ISC RNAseq data we obtained after filtration a matrix of 817 differentially expressed (DE) genes. We determined by hierarchical ascendant clustering 5 clusters and use this parameter for the k-means clustering k = 5.

From S2 cells RNAseq data we obtained after filtration a matrix of 3556 DE genes.

For the embryo we used the microarray data from a previous study¹⁹ and we obtained after filtration a matrix of 3538 DE genes.

Comparison of gene groups:

To compare three population of genes, the statistical SuperExactTest was used²⁰

Bibliography for Methods

1. Zanet, J. *et al.* Pri sORF peptides induce selective proteasome-mediated protein processing. *Science* **349**, 1356–1358 (2015).
2. Feng, J., Liu, T. & Zhang, Y. Using MACS to identify peaks from ChIP-Seq data. *Current protocols in bioinformatics* **Chapter 2**, Unit 2.14 (2011).
3. Carroll, T. S., Liang, Z., Salama, R., Stark, R. & de Santiago, I. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Frontiers in Genetics* **5**, (2014).
4. Zhu, L. J. *et al.* ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, (2010).
5. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research* **44**, W160-5 (2016).
6. Thomas-Chollier, M. *et al.* RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic acids research* **40**, e31 (2012).
7. Medina-Rivera, A. *et al.* RSAT 2015: Regulatory Sequence Analysis Tools. *Nucleic acids research* **43**, W50-6 (2015).
8. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
9. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

10. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139–40 (2010).

List of abbreviations used

bp: base pair; BS: binding site; CHIP-seq: chromatin immunoprecipitation coupled to next generation sequencing; RNA-seq: RNA sequencing. GFP: green fluorescent protein; GRN: gene regulatory network; TF: transcription factor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AMF performed transfection and CHIP-seq experiments for histone marks in S2 cells, with the help of JA, and most computational analyses with the help of AR (unsupervised clustering) and CI (ISC RNA-seq analyses). CI and CP dissected guts of different genotypes; CI did FASC purification and prepared samples for ISC RNA-seq. PL and SP prepared S2 cell samples for RNA-seq and Svb::GFP CHIP-seq. CP, OC and FP designed and supervised the project. AMF, CI, CP and FP wrote the paper. All the authors have read and approved the final version.

Acknowledgements

We are grateful to M. Gallois and P. Valenti for precious help with plasmid constructs and experimental support. We also thank all members of the Payre lab for critical reading of the manuscript. We thank Marion Aguirrebengoa (BigA platform) for her bioinformatic expertise. This work was supported by ANR (Chrononet), Fondation pour la Recherche Médicale (FRM, grant DEQ20170336739) and CBI. AMF and CI were supported by fellowships from Ligue contre le Cancer and FRM, respectively.

REFERENCES

1. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B: **The role of DNA shape in protein-DNA recognition.** *Nature* 2009, **461**:1248-U1281.
2. Bulyk ML: **Computational prediction of transcription-factor binding site locations.** *Genome Biol* 2003, **5**:201.
3. Guertin MJ, Lis JT: **Chromatin landscape dictates HSF binding to target DNA elements.** *PLoS Genet* 2010, **6**:e1001114.
4. Slattery M, Zhou TY, Yang L, Machado ACD, Gordan R, Rohs R: **Absence of a simple code: how transcription factors read the genome.** *Trends in Biochemical Sciences* 2014, **39**:381-399.
5. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, et al: **Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors.** *Genome Res* 2012, **22**:1798-1812.
6. Arvey A, Agius P, Noble WS, Leslie C: **Sequence and chromatin determinants of cell-type-specific transcription factor binding.** *Genome Res* 2012, **22**:1723-1734.
7. Spitz F, Furlong EE: **Transcription factors: from enhancer binding to developmental control.** *Nat Rev Genet* 2012, **13**:613-626.
8. Sen SQ, Chanchani S, Southall TD, Doe CQ: **Neuroblast-specific open chromatin allows the temporal transcription factor, Hunchback, to bind neuroblast-specific loci.** *Elife* 2019, **8**.
9. Kumar A, Bhandari A, Sinha R, Sardar P, Sushma M, Goyal P, Goswami C, Grapputo A: **Molecular phylogeny of OVOL genes illustrates a conserved C2H2 zinc finger domain coupled by hypervariable unstructured regions.** *PLoS One* 2012, **7**:e39399.
10. Mevelnino M, Terracol R, Salles C, Vincent A, Payre F: **Ovo, a Drosophila Gene Required for Ovarian Development, Is Specifically Expressed in the Germline and Shares Most of Its Coding Sequences with Shavenbaby, a Gene Involved in Embryo Patterning.** *Mechanisms of Development* 1995, **49**:83-95.
11. Teng A, Nair M, Wells J, Segre JA, Dai X: **Strain-dependent perinatal lethality of Ovol1-deficient mice and identification of Ovol2 as a downstream target of Ovol1 in skin epidermis.** *Biochim Biophys Acta* 2007, **1772**:89-95.
12. Tsuji G, Ito T, Chiba T, Mitoma C, Nakahara T, Uchi H, Furue M: **The role of the OVOL1-OVOL2 axis in normal and diseased human skin.** *J Dermatol Sci* 2018, **90**:227-231.
13. Saxena K, Srikrishnan S, Celia-Terrassa T, Jolly MK: **OVOL1/2: Drivers of Epithelial Differentiation in Development, Disease, and Reprogramming.** *Cells Tissues Organs* 2020:1-10.
14. Roca H, Hernandez J, Weidner S, McEachin RC, Fuller D, Sud S, Schumann T, Wilkinson JE, Zaslavsky A, Li H, et al: **Transcription factors OVOL1 and OVOL2 induce the mesenchymal to epithelial transition in human cancer.** *PLoS One* 2013, **8**:e76773.
15. Wang ZH, Li Z, Hu M, Yang QJ, Yan S, Wu RS, Li BA, Guo M: **Ovol2 gene inhibits the Epithelial-to-Mesenchymal Transition in lung adenocarcinoma by transcriptionally repressing Twist1.** *Gene* 2017, **600**:1-8.

16. Watanabe K, Villarreal-Ponce A, Sun P, Salmans ML, Fallahi M, Andersen B, Dai X: **Mammary morphogenesis and regeneration require the inhibition of EMT at terminal end buds by *Ovol2* transcriptional repressor.** *Dev Cell* 2014, **29**:59-74.
17. Chen L, Tian X, Gong W, Sun B, Li G, Liu D, Guo P, He Y, Chen Z, Xia Y, et al: **Periostin mediates epithelial-mesenchymal transition through the MAPK/ERK pathway in hepatoblastoma.** *Cancer Biol Med* 2019, **16**:89-100.
18. Liu J, Wu Q, Wang Y, Wei Y, Wu H, Duan L, Zhang Q, Wu Y: ***Ovol2* induces mesenchymal-epithelial transition via targeting *ZEB1* in osteosarcoma.** *Onco Targets Ther* 2018, **11**:2963-2973.
19. Davidson AE, Liskova P, Evans CJ, Dudakova L, Noskova L, Pontikos N, Hartmannova H, Hodanova K, Stranecky V, Kozmik Z, et al: **Autosomal-Dominant Corneal Endothelial Dystrophies *CHED1* and *PPCD1* Are Allelic Disorders Caused by Non-coding Mutations in the Promoter of *OVOL2*.** *Am J Hum Genet* 2016, **98**:75-89.
20. Kitazawa K, Hikichi T, Nakamura T, Mitsunaga K, Tanaka A, Nakamura M, Yamakawa T, Furukawa S, Takasaka M, Goshima N, et al: ***OVOL2* Maintains the Transcriptional Program of Human Corneal Epithelium by Suppressing Epithelial-to-Mesenchymal Transition.** *Cell Rep* 2016, **15**:1359-1368.
21. Lee B, Villarreal-Ponce A, Fallahi M, Ovadia J, Sun P, Yu QC, Ito S, Sinha S, Nie Q, Dai X: **Transcriptional mechanisms link epithelial plasticity to adhesion and differentiation of epidermal progenitor cells.** *Dev Cell* 2014, **29**:47-58.
22. Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, Dai X: **An *Ovol2*-*Zeb1* Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States.** *PLoS Comput Biol* 2015, **11**:e1004569.
23. Nair M, Teng A, Bilanchone V, Agrawal A, Li B, Dai X: ***Ovol1* regulates the growth arrest of embryonic epidermal progenitor cells and represses *c-myc* transcription.** *J Cell Biol* 2006, **173**:253-264.
24. Hayashi M, Shinozuka Y, Shigenobu S, Sato M, Sugimoto M, Ito S, Abe K, Kobayashi S: **Conserved role of *Ovo* in germline development in mouse and *Drosophila*.** *Sci Rep* 2017, **7**:40056.
25. Bai R, Kusama K, Nakamura K, Sakurai T, Kimura K, Ideta A, Aoyagi Y, Imakawa K: **Down-regulation of transcription factor *OVOL2* contributes to epithelial-mesenchymal transition in a noninvasive type of trophoblast implantation to the maternal endometrium.** *FASEB J* 2018, **32**:3371-3384.
26. Jeyarajah MJ, Jaju Bhattad G, Hillier DM, Renaud SJ: **The Transcription Factor *OVOL2* Represses *ID2* and Drives Differentiation of Trophoblast Stem Cells and Placental Development in Mice.** *Cells* 2020, **9**.
27. Renaud SJ, Chakraborty D, Mason CW, Rumi MA, Vivian JL, Soares MJ: ***OVO*-like 1 regulates progenitor cell fate in human trophoblast development.** *Proc Natl Acad Sci U S A* 2015, **112**:E6175-6184.
28. Kim JY, Lee RH, Kim TM, Kim DW, Jeon YJ, Huh SH, Oh SY, Kyba M, Kataoka H, Choi K, et al: ***OVOL2* is a critical regulator of *ER71/ETV2* in generating *FLK1+*, hematopoietic, and endothelial cells from embryonic stem cells.** *Blood* 2014, **124**:2948-2952.
29. Papathanasiou M, Tsiftoglou SA, Polyzos AP, Papadopoulou D, Valakos D, Klagkou E, Karagianni P, Pliatska M, Talianidis I, Agelopoulos M, Thanos D: **Identification of a dynamic gene regulatory**

- network required for pluripotency factor-induced reprogramming of mouse fibroblasts and hepatocytes.** *EMBO J* 2021, **40**:e102236.
30. Lee S, Garfinkel MD: **Characterization of Drosophila OVO protein DNA binding specificity using random DNA oligomer selection suggests zinc finger degeneration.** *Nucleic Acids Res* 2000, **28**:826-834.
31. Lu J, Oliver B: **Drosophila OVO regulates ovarian tumor transcription by binding unusually near the transcription start site.** *Development* 2001, **128**:1671-1686.
32. Andrews J, Garcia-Estefania D, Delon I, Lu JN, Mevel-Ninio M, Spierer A, Payre F, Pauli D, Oliver B: **OVO transcription factors function antagonistically in the Drosophila female germline.** *Development* 2000, **127**:881-892.
33. Payre F, Vincent A, Carreno S: **Ovo/svb integrates Wingless and DER pathways to control epidermis differentiation.** *Nature* 1999, **400**:271-275.
34. Kondo T, Plaza S, Zanet J, Benrabah E, Valenti P, Hashimoto Y, Kobayashi S, Payre F, Kageyama Y: **Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis.** *Science* 2010, **329**:336-339.
35. Zanet J, Benrabah E, Li T, Pelissier-Monier A, Chanut-Delalande H, Ronsin B, Bellen HJ, Payre F, Plaza S: **Pri sORF peptides induce selective proteasome-mediated protein processing.** *Science* 2015, **349**:1356-1358.
36. Chanut-Delalande H, Fernandes I, Roch F, Payre F, Plaza S: **Shavenbaby couples patterning to epidermal cell shape control.** *PLoS Biol* 2006, **4**:e290.
37. Fernandes I, Chanut-Delalande H, Ferrer P, Latapie Y, Waltzer L, Affolter M, Payre F, Plaza S: **Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes.** *Dev Cell* 2010, **18**:64-76.
38. Menoret D, Santolini M, Fernandes I, Spokony R, Zanet J, Gonzalez I, Latapie Y, Ferrer P, Rouault H, White KP, et al: **Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization.** *Genome Biol* 2013, **14**:R86.
39. Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, Bahin M, Collombet S, Vincens P, Thieffry D, et al: **RSAT 2018: regulatory sequence analysis tools 20th anniversary.** *Nucleic Acids Res* 2018, **46**:W209-W214.
40. Castro-Mondragon JA, Jaeger S, Thieffry D, Thomas-Chollier M, van Helden J: **RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections.** *Nucleic Acids Res* 2017, **45**:e119.
41. Rouault H, Mazouni K, Couturier L, Hakim V, Schweisguth F: **Genome-wide identification of cis-regulatory motifs and modules underlying gene coregulation using statistics and phylogeny.** *Proc Natl Acad Sci U S A* 2010, **107**:14615-14620.
42. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A: **Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian MII3/MI14.** *Genes Dev* 2012, **26**:2604-2620.
43. Huang C, Yang F, Zhang Z, Zhang J, Cai G, Li L, Zheng Y, Chen S, Xi R, Zhu B: **Mrg15 stimulates Ash1 H3K36 methyltransferase activity and facilitates Ash1 Trithorax group protein function in Drosophila.** *Nat Commun* 2017, **8**:1649.

44. Kellner WA, Ramos E, Van Bortle K, Takenaka N, Corces VG: **Genome-wide phosphoacetylation of histone H3 at Drosophila enhancers and promoters.** *Genome Res* 2012, **22**:1081-1088.
45. Li L, Lyu X, Hou C, Takenaka N, Nguyen HQ, Ong CT, Cubenas-Potts C, Hu M, Lei EP, Bosco G, et al: **Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing.** *Mol Cell* 2015, **58**:216-231.
46. Schauer T, Ghavi-Helm Y, Sexton T, Albig C, Regnard C, Cavalli G, Furlong EE, Becker PB: **Chromosome topology guides the Drosophila Dosage Compensation Complex for target gene activation.** *EMBO Rep* 2017.
47. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proc Natl Acad Sci U S A* 2010, **107**:21931-21936.
48. Cubenas-Potts C, Rowley MJ, Lyu X, Li G, Lei EP, Corces VG: **Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture.** *Nucleic Acids Res* 2017, **45**:1714-1730.
49. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**:108-112.
50. Sethi A, Gu M, Gumusgoz E, Chan L, Yan KK, Rozowsky J, Barozzi I, Afzal V, Akiyama JA, Plajzer-Frick I, et al: **Supervised enhancer prediction with epigenetic pattern recognition and targeted validation.** *Nat Methods* 2020, **17**:807-814.
51. Liu J, Ali M, Zhou Q: **Establishment and evolution of heterochromatin.** *Ann N Y Acad Sci* 2020, **1476**:59-77.
52. Rando OJ: **Global patterns of histone modifications.** *Curr Opin Genet Dev* 2007, **17**:94-99.
53. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A: **Genome-wide quantitative enhancer activity maps identified by STARR-seq.** *Science* 2013, **339**:1074-1077.
54. Johnson SC: **Hierarchical clustering schemes.** *Psychometrika* 1967, **32**:241-254.
55. Bohère J, Mancheno-Ferris A, Hayek S, Zanet J, Valenti P, Akino K, Yamabe Y, Inagaki S, Chanut-Delalande H, Plaza S, et al: **Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis.** *Nature Communications* 2018, **9**:5123.
56. Al Hayek S, Alsawadi A, Kambris Z, Boquete JP, Bohere J, Immarigeon C, Ronsin B, Plaza S, Lemaitre B, Payre F, Osman D: **Steroid-dependent switch of OvoL/Shavenbaby controls self-renewal versus differentiation of intestinal stem cells.** *EMBO J* 2021, **40**:e104347.
57. Micchelli CA, Perrimon N: **Evidence that stem cells reside in the adult Drosophila midgut epithelium.** *Nature* 2006, **439**:475-479.
58. MacQueen JB: **Some Methods for classification and Analysis of Multivariate Observations.** In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. pp. 281-297: Berkeley, University of California Press; 1967:281-297
59. Boumard B, Bardin AJ: **An amuse-bouche of stem cell regulation: Underlying principles and mechanisms from adult Drosophila intestinal stem cells.** *Curr Opin Cell Biol* 2021, **73**:58-68.

60. Tauc HM, Rodriguez-Fernandez IA, Hackney JA, Pawlak M, Ronnen Oron T, Korzelius J, Moussa HF, Chaudhuri S, Modrusan Z, Edgar BA, Jasper H: **Age-related changes in polycomb gene regulation disrupt lineage fidelity in intestinal stem cells.** *Elife* 2021, **10**.
61. Ibrahim MM, Karabacak A, Glaes A, Kolundzic E, Hirsekorn A, Carda A, Tursun B, Zinzen RP, Lacadie SA, Ohler U: **Determinants of promoter and enhancer transcription directionality in metazoans.** *Nat Commun* 2018, **9**:4472.
62. Mancheno-Ferris A, Polesello C, Payre F: **[OvoL factors: a family of key regulators of epithelium mesenchyme plasticity and stem cells].** *Med Sci (Paris)* 2020, **36 Hors serie n degrees 1**:61-66.
63. Chanut-Delalande H, Hashimoto Y, Pelissier-Monier A, Spokony R, Dib A, Kondo T, Bohere J, Niimi K, Latapie Y, Inagaki S, et al: **Pri peptides are mediators of ecdysone for the temporal control of development.** *Nat Cell Biol* 2014, **16**:1035-1044.
64. Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C: **The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila.** *Nature* 2008, **456**:400-403.
65. McDaniel SL, Gibson TJ, Schulz KN, Fernandez Garcia M, Nevil M, Jain SU, Lewis PW, Zaret KS, Harrison MM: **Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation.** *Mol Cell* 2019, **74**:185-195 e184.
66. Sun Y, Nien CY, Chen K, Liu HY, Johnston J, Zeitlinger J, Rushlow C: **Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.** *Genome Res* 2015, **25**:1703-1714.
67. Cordero JB, Stefanatos RK, Scopelliti A, Vidal M, Sansom OJ: **Inducible progenitor-derived Wingless regulates adult midgut regeneration in Drosophila.** *EMBO J* 2012, **31**:3901-3917.
68. Jiang H, Edgar BA: **EGFR signaling regulates the proliferation of Drosophila adult midgut progenitors.** *Development* 2009, **136**:483-493.
69. Kohlmaier A, Fassnacht C, Jin Y, Reuter H, Begum J, Dutta D, Edgar BA: **Src kinase function controls progenitor cell pools during regeneration and tumor onset in the Drosophila intestine.** *Oncogene* 2015, **34**:2371-2384.
70. Sousa-Victor P, Ayyaz A, Hayashi R, Qi Y, Madden DT, Lunyak VV, Jasper H: **Piwi Is Required to Limit Exhaustion of Aging Somatic Stem Cells.** *Cell Rep* 2017, **20**:2527-2537.
71. Stern DL, Frankel N: **The structure and evolution of cis-regulatory regions: the shavenbaby story.** *Philos Trans R Soc Lond B Biol Sci* 2013, **368**:20130028.
72. Preger-Ben Noon E, Sabaris G, Ortiz DM, Sager J, Liebowitz A, Stern DL, Frankel N: **Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function.** *Cell Rep* 2018, **22**:3021-3031.
73. Jiang H, Patel PH, Kohlmaier A, Grenley MO, McEwen DG, Edgar BA: **Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the Drosophila midgut.** *Cell* 2009, **137**:1343-1355.
74. Lu W, Kang Y: **Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis.** *Dev Cell* 2019, **49**:361-374.
75. Nieto MA, Huang RY, Jackson RA, Thiery JP: **Emt: 2016.** *Cell* 2016, **166**:21-45.
76. Pasani S, Sahoo S, Jolly MK: **Hybrid E/M Phenotype(s) and Stemness: A Mechanistic Connection Embedded in Network Topology.** *J Clin Med* 2020, **10**.

77. Haensel D, Sun P, MacLean AL, Ma X, Zhou Y, Stemmler MP, Brabletz S, Berx G, Plikus MV, Nie Q, et al: **An Ovol2-Zeb1 transcriptional circuit regulates epithelial directional migration and proliferation.** *EMBO Rep* 2019, **20**.
78. Antonello ZA, Reiff T, Ballesta-Illan E, Dominguez M: **Robust intestinal homeostasis relies on cellular plasticity in enteroblasts mediated by miR-8-Escargot switch.** *EMBO J* 2015, **34**:2025-2041.
79. Antonello ZA, Reiff T, Dominguez M: **Mesenchymal to epithelial transition during tissue homeostasis and regeneration: Patching up the Drosophila midgut epithelium.** *Fly (Austin)* 2015, **9**:132-137.

FIGURE LEGENDS

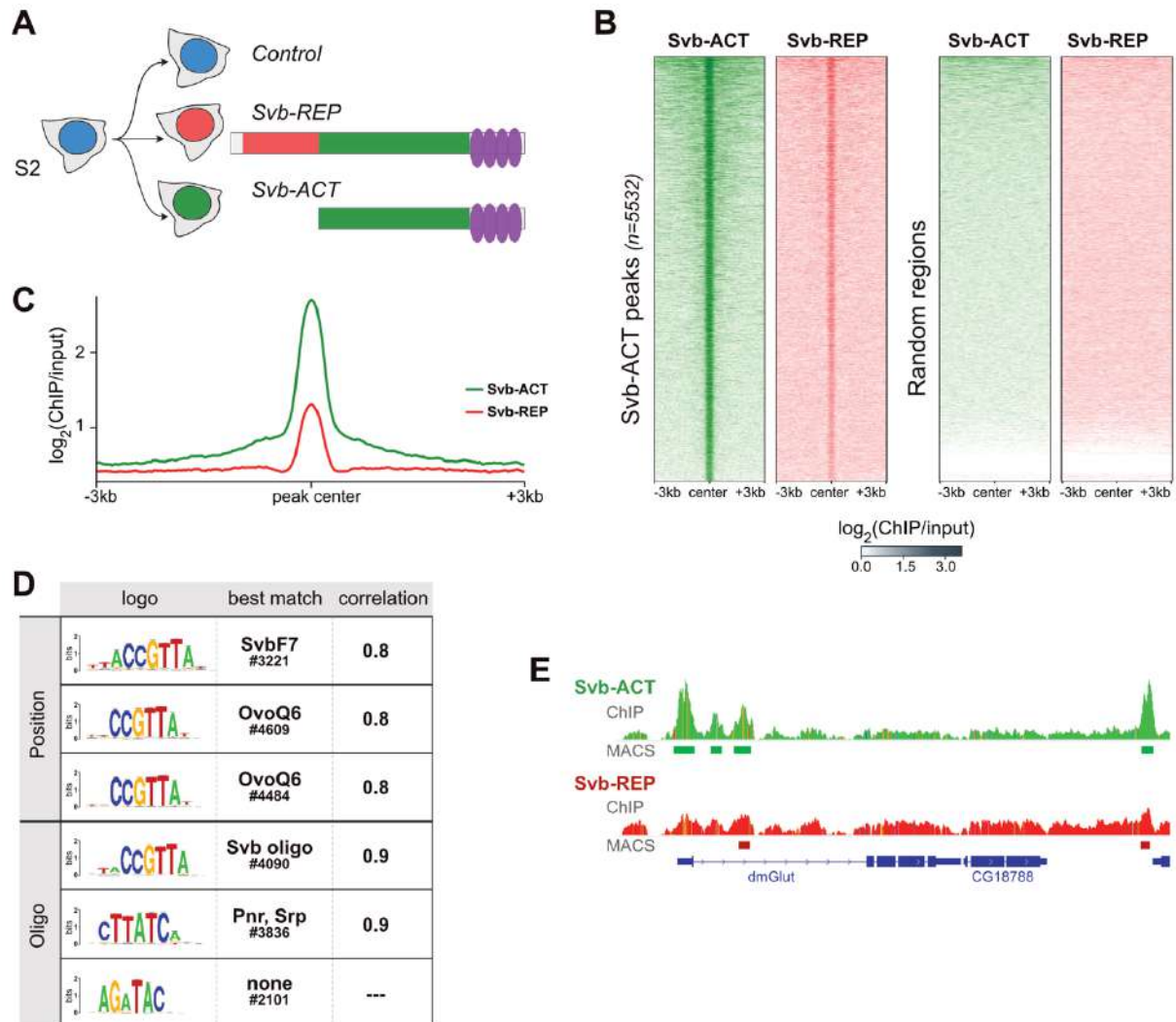


Figure 1: Shavenbaby Activator and Repressor isoforms bind to same genomic regions

(A) Cell lines used for ChIP-seq and RNA-seq experiments. Control correspond to parental S2 cells that lack endogenous *svb* expression; SvB-REP and SvB-ACT are stable cell lines allowing inducible expression of SvB::GFP together with either wild type *pri* mRNA, or a *pri* mRNA with four frameshift mutations preventing the production of Pri peptides, respectively. Green box corresponds to Activator domain, red box to Repressor domain and purple circles corresponds to Zinc-finger DNA-binding domain. (B) Heatmap of SvB-ACT (green) and SvB-REP (red) enrichment across all significant peaks of the *Drosophila* genome detected in SvB-ACT cells ($n=5512$). Each row represents a 6kb window centered on peak midpoint, sorted by the intensity of SvB-ACT ChIP-seq signal. ChIP signal for a similar set of random regions is show on the right. (C) Occupancy of SvB-ACT (green) and SvB-REP (red) isoforms at SvB-bound sites (in SvB-ACT cells, $n=5512$) displayed as metaprofiles integrating signal of all peaks. (D) DNA motifs enriched in SvB-bound peaks, as defined by two algorithms (position or oligo) of the RSAT suite [39]. (E) Genome browser screen shot of the *dmGUT* locus, showing respective intensity of SvB-ACT and SvB-REP ChIP signals.

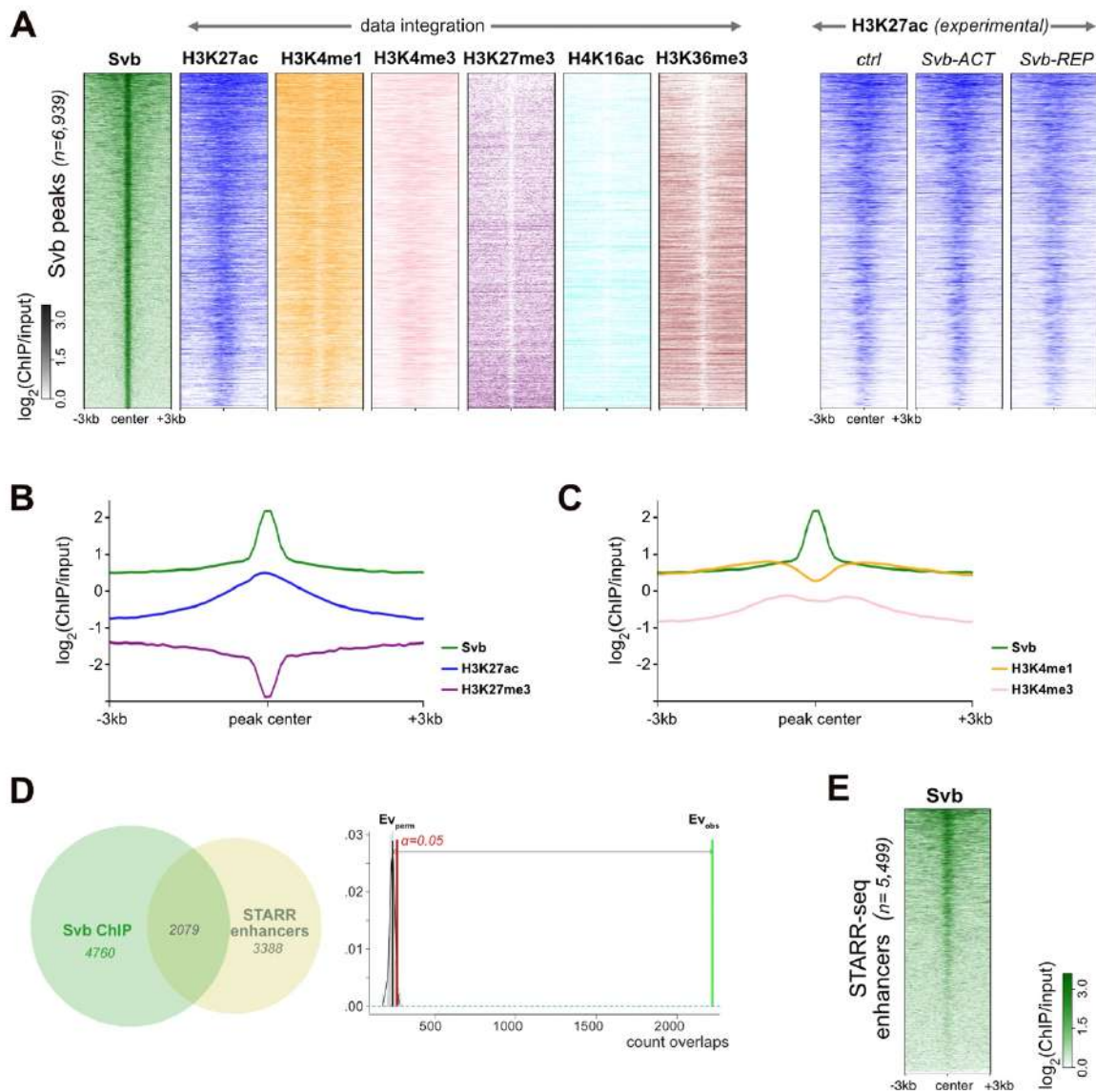


Figure 2: Svb binds in regions presenting enhancer properties

(A) Heatmap of Svb-ACT, H3K27ac, H3K4me1, H3K4me3, H3K27me3, H4K16ac and H3K36me3 ChIP-seq enrichment across all peaks bound by Svb (q -value 10^{-1} , $n=6939$). Raw data for published histone marks (refs) were re-processed using standardized procedures and are shown on the left. The three heatmaps on the right show results of H3K27ac ChIP-seq experiments performed in control (ctrl), Svb-ACT and Svb-REP cell lines. Each row represents a 6kb window centered on Svb peak midpoint, sorted by Svb-ACT enrichment. Average enrichment of H3K27ac and H3K27me3 (B), and of H3K4me1 and H3K4me3 (C), on Svb-bound regions. (D) Venn diagram showing the overlap between Svb-bound regions (green, $n=6939$) and active enhancers (pale yellow, $n=5499$) identified by STARR-seq in S2 cells [53]. For the common population overlapping regions are considered as unique region. The graph shows an hypergeometric test (Z -score=129.05, p -value=0.0099) on the observed data dataset (green) compared to 100 random permutations (grey) of the data; x axis represents the number of overlaps, y axis the observed frequency. (E) The heatmap shows enrichment of Svb-ACT at S2 enhancer regions (5,499); each row represents a 6kb window centered on STARR-seq peak midpoint, sorted by the intensity of Svb-ACT ChIP signal.

Notes explicatives:

Dans les diagrammes de Venn obtenus lors de la comparaison des régions de fixation, l'addition des nombres ne correspond pas au nombre indiqué. En effet la fonction du package ChIPpeakanno permettant de calculer les régions communes entre plusieurs populations de fixation considère pour la population commune que les pics ayant un fort chevauchement ne compte que pour un.

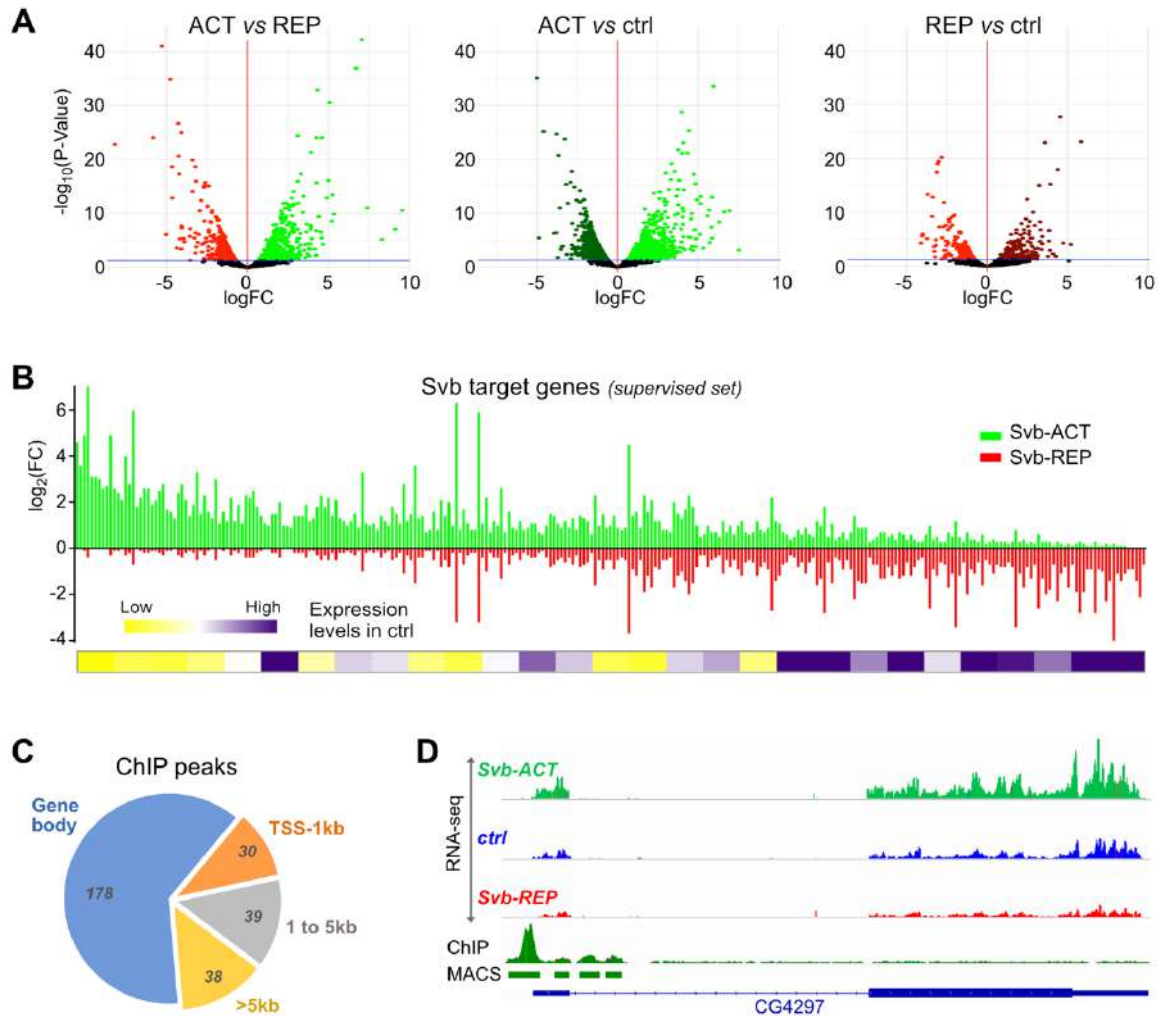


Figure 3: Identification of Svb direct target genes in S2 cells

(A) Volcano plots show differentially expressed (DE) genes upon Svb-ACT or Svb-REP expression in S2 cells. S2 parental cells for ctrl condition. (B) Differential expression in Svb-ACT (green) and Svb-REP (red) cells of a set of 285 target genes, as defined by a supervised approach based on coherent changes in expression levels (p -value $< 0,05$ in Svb-ACT vs Svb-REP, AND $\log_2FC > 0$ in Svb-ACT vs ctrl, AND $\log_2FC < 0$ Svb-REP vs ctrl). The chart below displays averaged expression levels in parental S2 cells of groups of ten genes. Gene expression levels are color coded, white represents the median expression, yellow lower and purple higher levels. (C) Location of Svb peaks within the set of 285 target genes. Gene body represents the whole transcribed region extended from 300bp upstream the TSS down to +1kb downstream the transcriptional stop; TSS-1kb, to 1kb region upstream of the TSS, 1 to 5kb, the region 1-5kb upstream of TSS. Other peaks are located at distances >5 kb upstream of the TSS. (D) Genome browser snapshot of genomic locus encompassing the CG4297 gene showing the effects of Svb isoforms on expression levels (RNAseq), as well as Svb binding (ChIP-seq signal and MACS peaks).

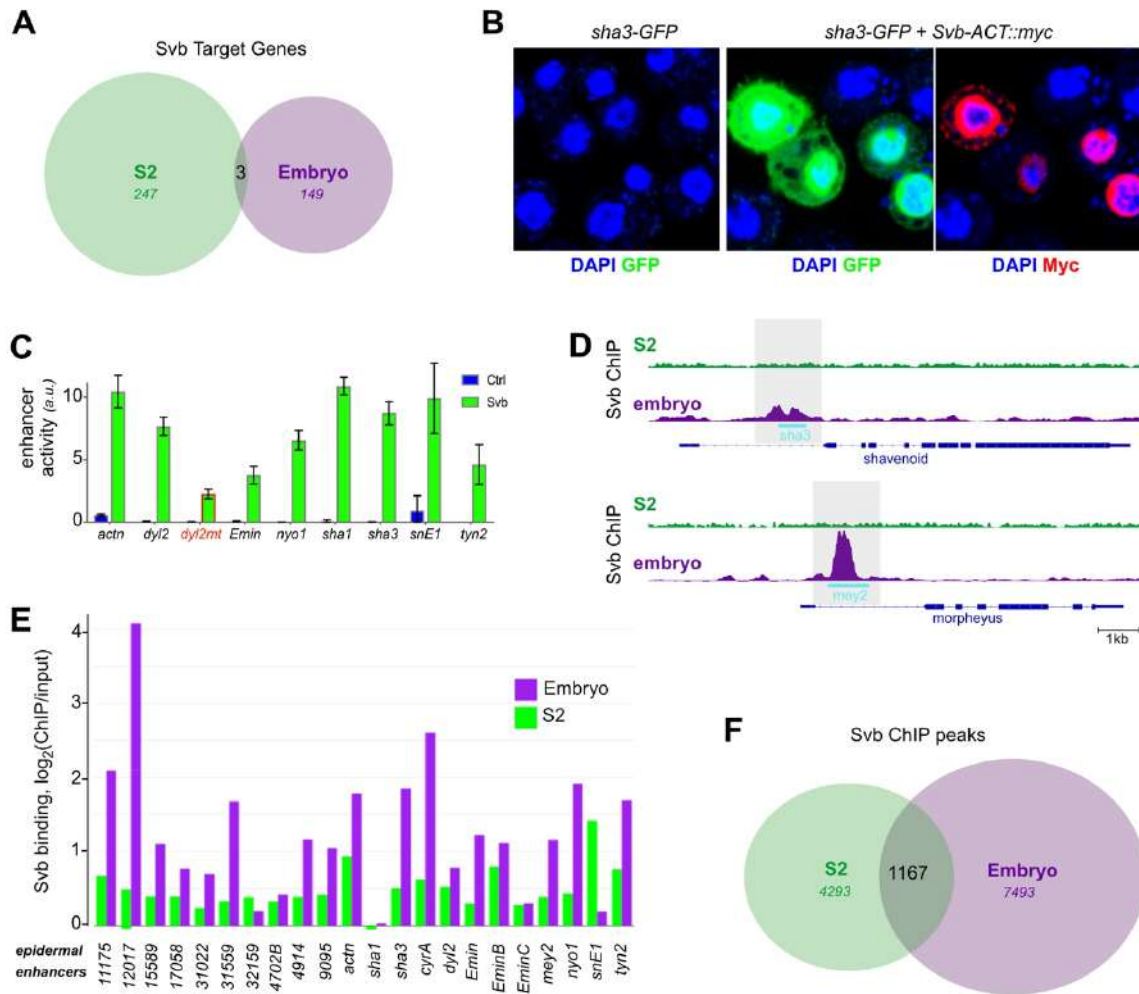


Figure 4: Different binding behavior of Svb regulate different sets of target genes

(A) Venn diagram showing the overlap between the sets of Svb direct target genes in S2 cells (n=247) and in embryonic epidermal cells (n=149) [38]. **(B)** Activity of *sha3* enhancer, as monitored by transient transfection assays of reporter construct (*sha3*-GFP, green), in the absence or presence of Svb-ACT (Svb-ACT::myc, red). Blue is DAPI. **(C)** Quantification of the activity of Svb target enhancers that drive expression in embryonic epidermal cells [36, 38]. GFP levels were used to monitor enhancer activity in control and Svb-ACT cells. **(D)** Genome browser snapshots showing Svb ChIP signal in S2 cells (green) and embryo (purple) at two loci encompassing direct embryonic target, with position of *in vivo* validated enhancers [38]. **(E)** Quantification of Svb binding in S2 and embryonic epidermal cells across all embryonic enhancers previously identified in [38]. **(F)** Comparison of Svb bound regions, as defined by ChIP-seq in S2 cells (green) and in embryonic epidermal cells (purple). For the common population overlapping regions are considered as unique region.

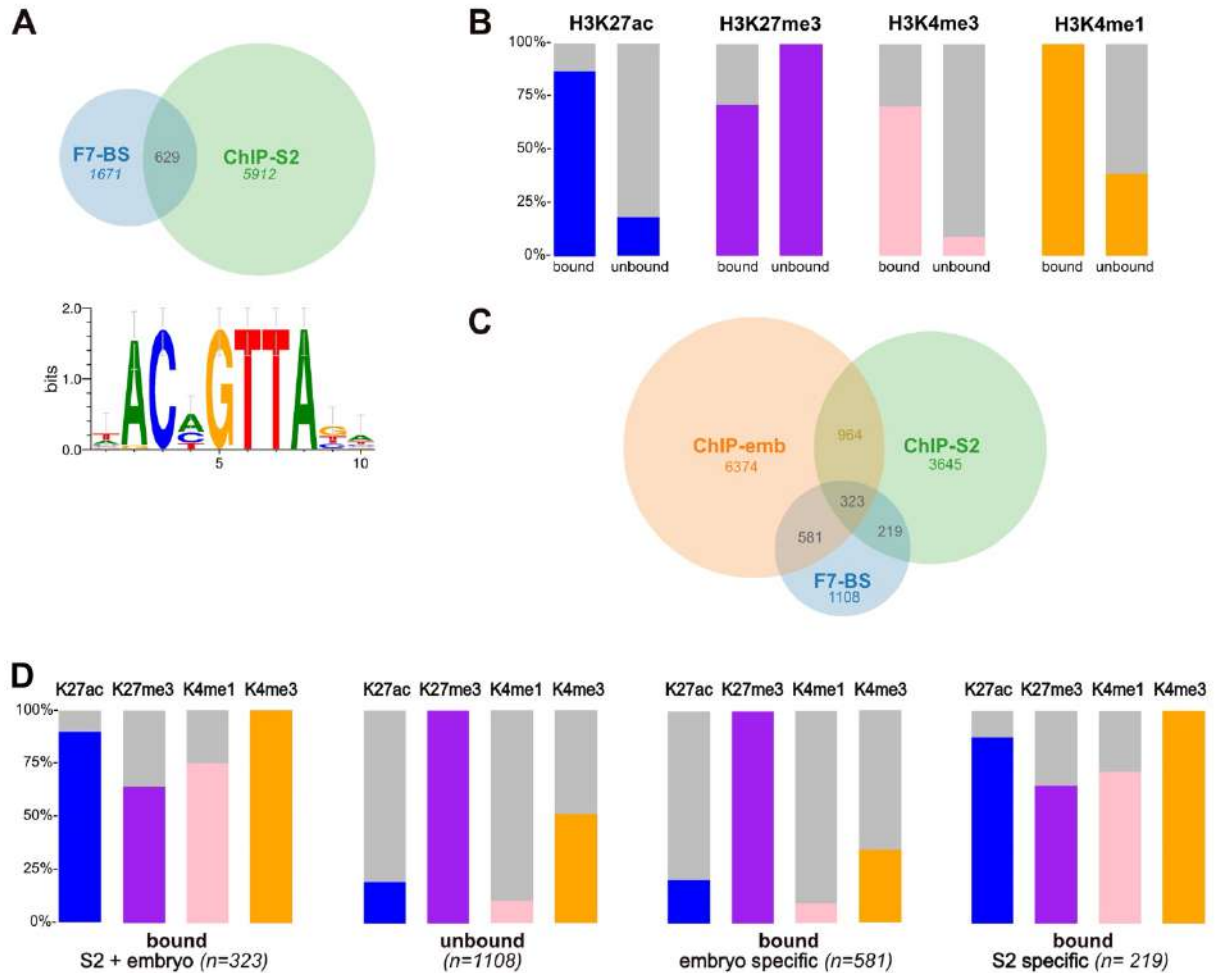


Figure 5: Analysis of Svb environment of fixation in embryo and S2 cells

(A) Venn diagram highlighting the subset of genomic F7-BS motifs bound by Svb in S2 cells and representation of F7-BS position weight matrix [38]. For the common population overlapping regions are considered as unique region. **(B)** Enrichment of histone marks in regions bearing F7-BS, as shown by a Fisher exact test in Svb-bound unbound subsets. **(C)** Comparison of Svb binding behavior between S2 cells (green) and embryo (orange) in genomic regions with F7-BS (blue). For the common population overlapping regions are considered as unique region. **(D)** Enrichment of histone marks around F7-BS motifs in the different sets defined in B.

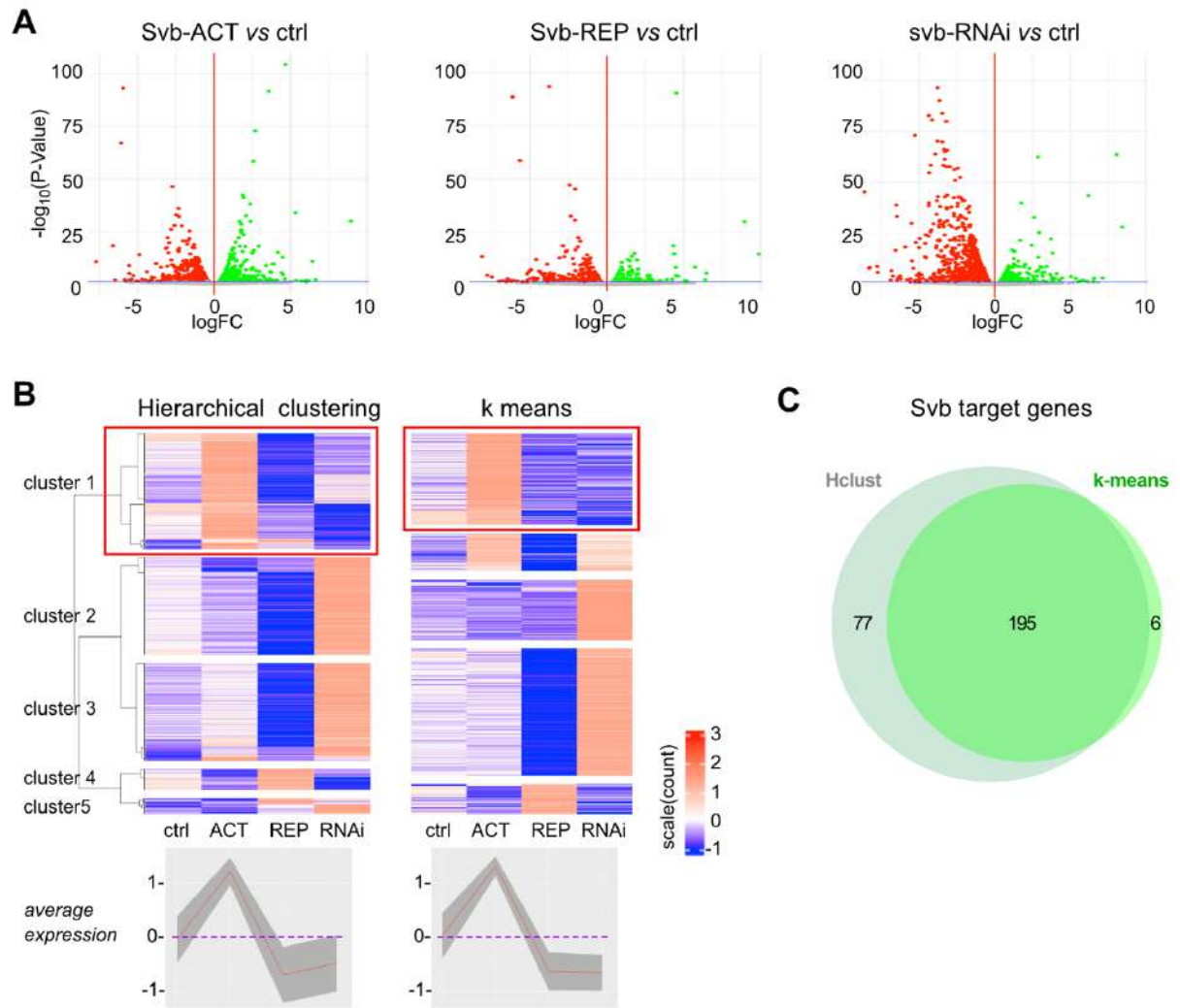


Figure 6: Identification of Svb target genes in adult intestinal stem cells

(A) DE genes upon targeted expression of Svb-ACT, Svb-REP or *svb-RNAi* in intestinal stem cells, shown as volcano plots. Each condition is compared to control conditions. Upregulated genes are in green, downregulated genes are in red. **(B)** Heatmaps represent clusters of DE genes in ISCs, using hierarchical clustering (left) and k-means (right) unsupervised approaches, based on mRNA expression levels in control, Svb-ACT, Svb-REP and *svb-RNAi* conditions. Each horizontal line represents a gene, with three replicates in each of the four conditions. Expression levels (scale count) are color coded. Cluster representative of Svb target genes are boxed ($n=272$ genes for hierarchical clustering, $n=201$ genes for k-means). The charts below represent cumulative expression of all genes of the cluster in the four conditions; the red line shows average expression levels. **(F)** Comparison between the two sets of genes identified by unsupervised approaches; hierarchical clustering (Hclust) and k-means.

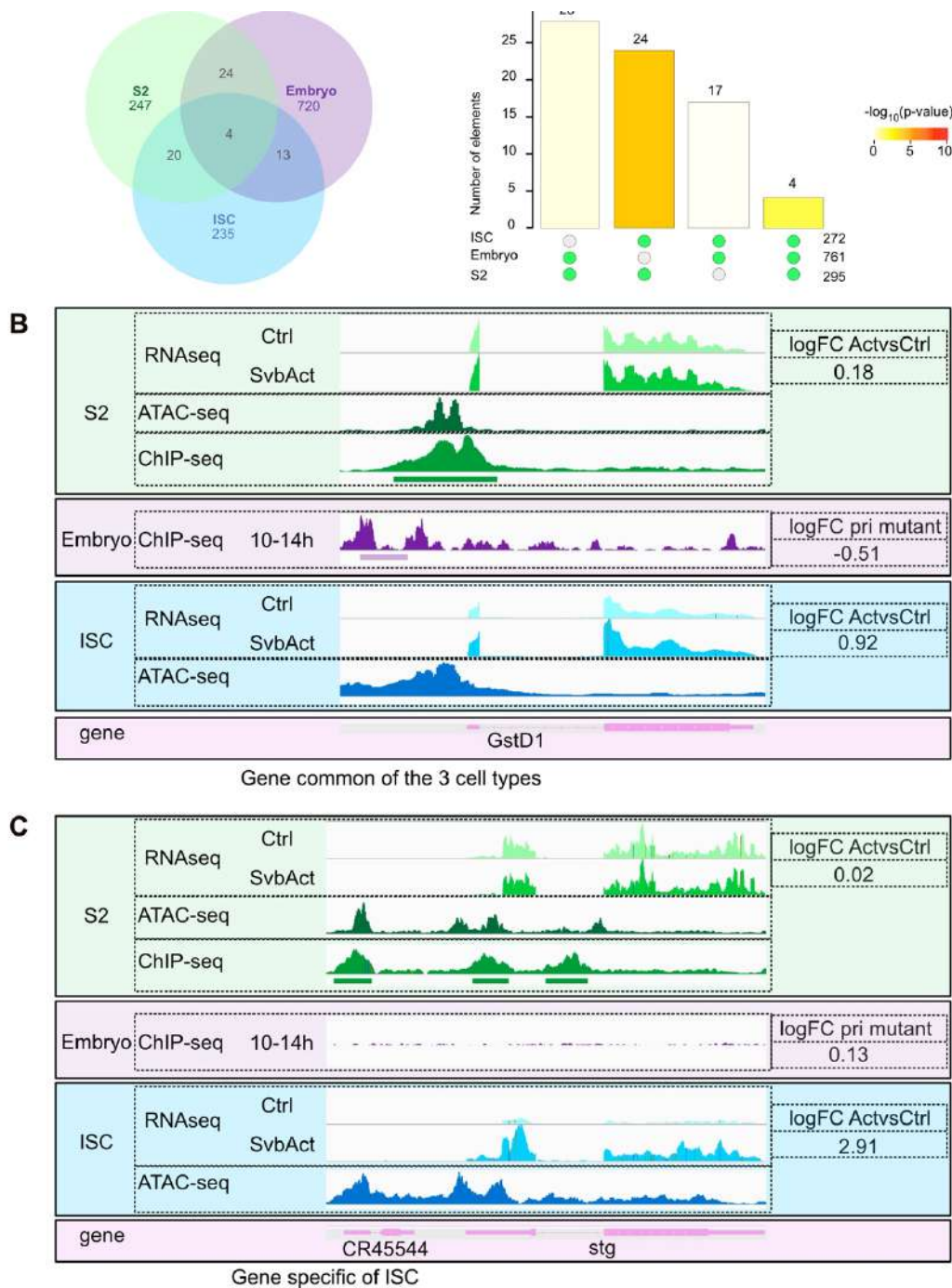


Figure 7: Comparing Svb target genes across several cell types

(A) Comparison of the sets of Svb target genes identified by hierarchical clustering in embryo (n=761), S2 cells (n=295) and intestinal stem cells (n=272). P-values estimating the significance of the observed intersections between the different sets of genes were calculated by super exact tests and represented by the color code. **(B-C)** Genome browser snapshots of the genomic loci of *GstD1* (B) and *stg* (C), with showing mRNA levels (RNAseq), Svb binding (ChIP-seq signal and MACS regions) and chromatin accessibility (ATAC-seq) in S2 cells, embryonic epidermal cells and ISCs.

SUPPLEMENTARY MATERIAL

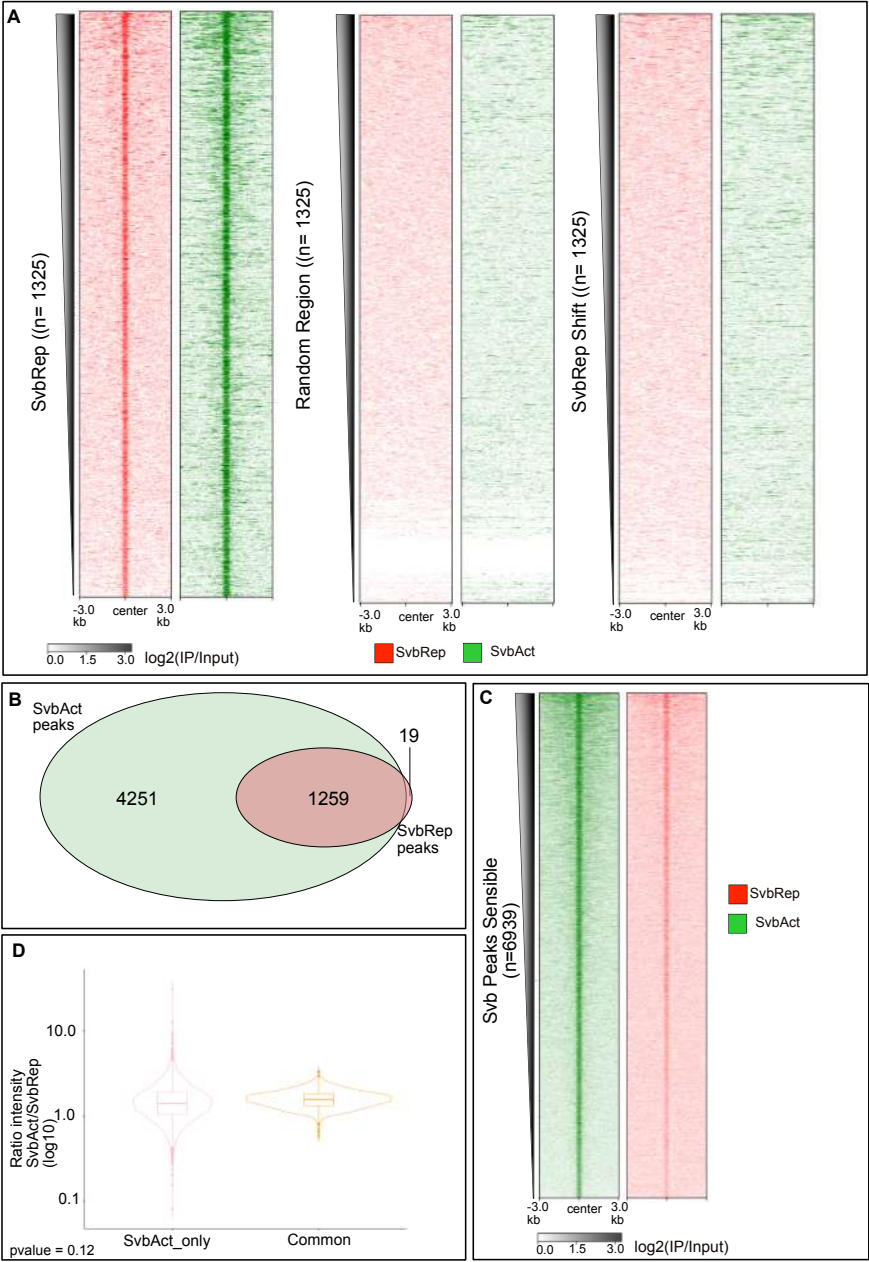


Figure S1: Shavenbaby Activator and Repressor bind on the same genomic regions

(A) Heatmap showing Svb ChIP signal at Svb-REP-bound regions (n =1325, left), random regions (middle) and Svb-REP-bound regions shifted by 10kb (right). Each row represents a 6kb window centered on peak midpoint, sorted by the intensity of Svb-REP ChIP signal. **(B)** Venn diagram comparing the set of MACS peaks detected for Svb bound regions in Svb-REP and Svb-ACT cells. For the common population overlapping regions are considered as unique region. **(C)** Heatmap showing ChIP signal in Svb-ACT and Svb-REP cells, across all peaks detected by MACS using more relaxed parameter (q value=10⁻⁴, n=6939). Regions are ranked by the intensity of Svb-ACT signal. **(D)** Quantification of the ratio of Svb-REP over Svb-ACT signals, in regions corresponding to MACS peaks

common to Svb-ACT and Svb-REP cells (n=1259, orange), and MACS peaks only detected in Svb-ACT cells (n=4251, pink). The p-value was calculated by a Student test.

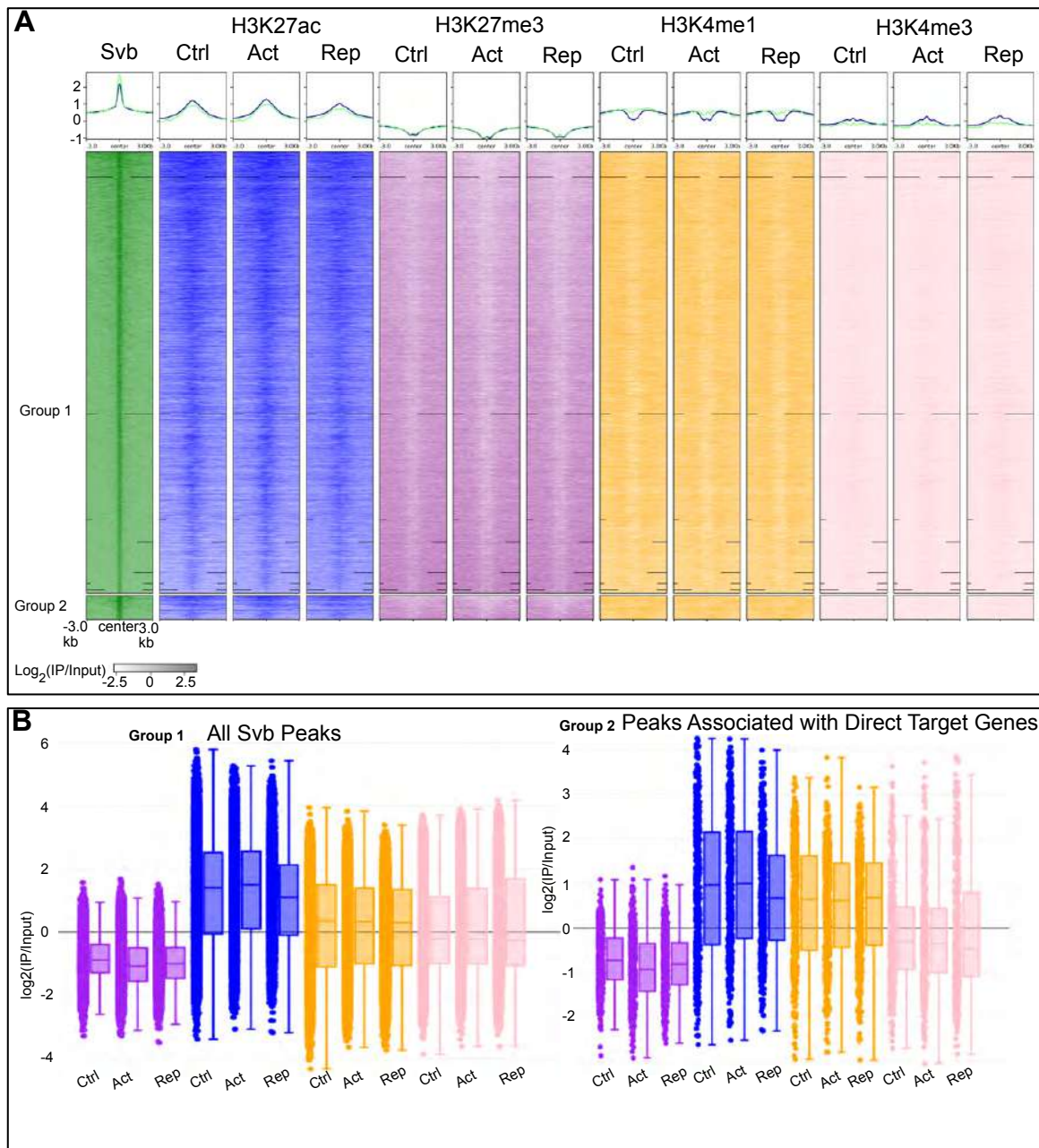


Figure S2: Effect of Svb-ACT and Svb-REP on histone marks

(A) Heatmap showing ChIP-seq signal for Svb-ACT and histone marks, as experimentally assayed in control, Svb-ACT and Svb-REP cells. Each row represents a 6kb window centered on peak midpoint, sorted by the intensity of Svb-ACT ChIP signal across all regions (n =6939, group 1). Group 2 represents a subset of Svb bound regions that overlap with Svb direct target genes in S2 cells (n= 423). **(B)** Quantification of histone marks ChIP signal at MACS regions corresponding to all peaks (n=6939) and peaks associated with Svb target genes (n=423).

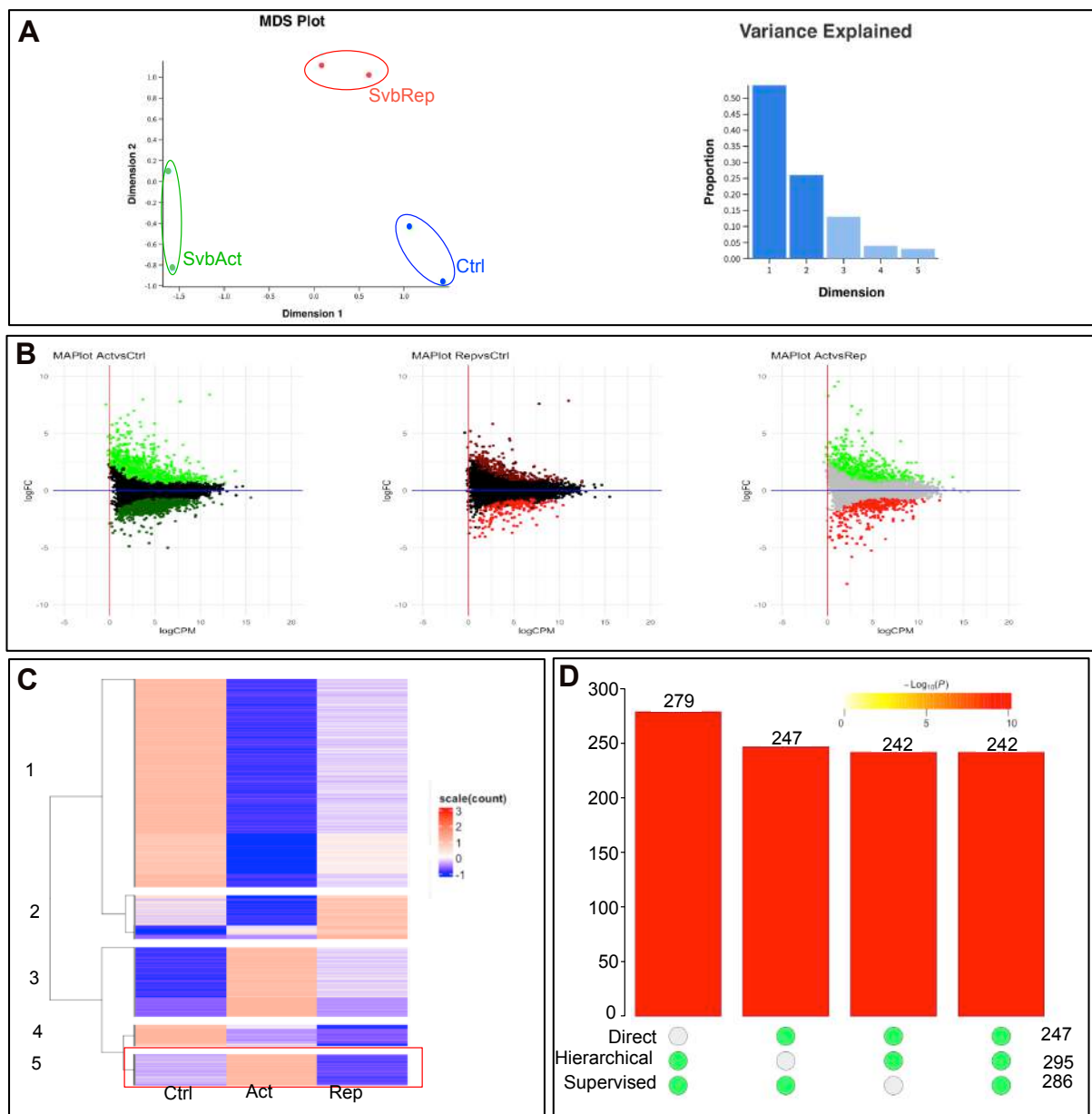


Figure S3: Identification of Svb target genes in S2 cells

(A) MDS plot representing global variation in gene expression observed between control, Svb-ACT and Svb-REP RNAseq data set. Each experiment was made in duplicate. >70 percent of the variance is explained by the two first dimensions (right panel). **(B)** MA plots comparing individual gene expression in Svb-ACT vs control (Ctrl), Svb-REP vs control, and Svb-ACT vs Svb-REP conditions. Deregulated are colored in green for Svb-ACT and red for Svb-REP. **(C)** Heatmap showing hierarchical clustering of all genes deregulated between control, Svb-ACT and Svb-REP cells, with color coded expression levels (scale counts). Cluster 5 groups 295 genes, 279 (out of 285 genes) being common to the set of genes identified by supervised approach. **(D)** The graph represents the number of common genes between the sets defined by hierarchical clustering (n=295), supervised approach (n=285), and the subset of those that are bound by Svb (n=247). P-values are color coded.

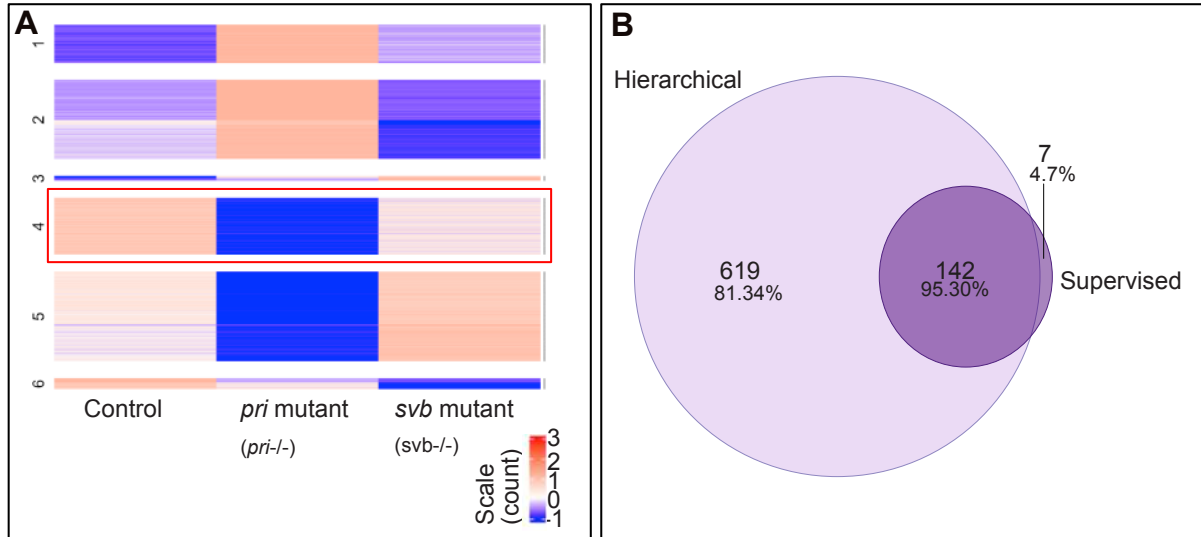


Figure S4: Identification of Svb target genes in embryos

(A) Heatmap showing hierarchical clustering of deregulated genes in control, *pri*-mutant and *svb*-mutant embryos. Raw data are from [38]. Expression levels are color coded; cluster #3 corresponds to 761 genes, downregulated in *svb* and *pri* mutants, which represent candidate Svb target genes. **(B)** Venn diagram comparing the two groups of Svb target genes identified by supervised (n=149) and unsupervised (hierarchical clustering, n=761) approaches.

DATA REFERENCE

Dataset	SRA ID	GEO ID	URL	Tissue	References	PMID
Svb Embryo 1	SRR931826	GSE48791	https://www.ebi.ac.uk/biosamples/samples/SAMN02231205	embryo	Menoret D et al. Genome Biol 2013	23972280
Svb Embryo 2	SRR931827	GSE48791	https://www.ebi.ac.uk/biosamples/samples/SAMN02231213	embryo	Menoret D et al. Genome Biol 2013	23972280
Svb Embryo input	SRR931828	GSE48791	https://www.ebi.ac.uk/biosamples/samples/SAMN02231206	embryo	Menoret D et al. Genome Biol 2013	23972280
Input K27ac/ K4me1/K4me3 Rep 1	SRX193308	GSE41440	https://www.ebi.ac.uk/ena/data/view/SRX193308	S2	Herz Hmet al. Genes Dev 2012	23166019
Input K27ac/ K4me1/K4me3 Rep 2	SRX193309	GSE41440	https://www.ebi.ac.uk/ena/data/view/SRX193309	S2	Herz Hmet al. Genes Dev 2012	23166019
H3K4me3	SRR1636780	GSE63518	https://www.ebi.ac.uk/ena/browser/view/SRX749052	S2	Li L et al. Mol Cell 2015	25818644
H3K27ac	SRX128170	GSE36374	https://www.ebi.ac.uk/ena/data/view/SRX128170	S2	Kellner et al. Genome Res 2012	22508764
H3K4me1	SRX749047	GSE63518	https://www.ebi.ac.uk/ena/data/view/SRX749047	S2	Li L et al. Mol Cell 2015	25818644
H3K27me3	SRX3170975	GSE93100	https://www.ebi.ac.uk/ena/data/view/SRX3170975	S2	Huang C,et al. Nat Commun 2017	29158494
Input H3K27me3	SRX2458935	GSE93100	https://www.ebi.ac.uk/ena/data/view/SRR5140284	S2	Huang C,et al. Nat Commun 2017	29158494
H4K16ac	SRX2520633	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRX2520633	S2	Schauer T et al. EMBO Rep	28794204
Input H4K16ac Rep1	SRX2520631	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRX2520631	S2	Schauer T et al. EMBO Rep	28794204
Input H4K16ac Rep2	SRX2520634	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRX2520634	S2	Schauer T et al. EMBO Rep	28794204
Input H3K36me3 Rep1	SRR5206763	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRR5206763	S2	Schauer T et al. EMBO Rep	28794204
Input H3K36me3 Rep2	SRX2520646	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRR5206765	S2	Schauer T et al. EMBO Rep	28794204
H3K36me3 Rep1	SRR5206764	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRR5206764	S2	Schauer T et al. EMBO Rep	28794204
H3K36me3 Rep2	SRR5206766	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRR5206766	S2	Schauer T et al. EMBO Rep	28794204
H4K16ac Rep1	SRX2520641	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRX2520641	KC	Schauer T et al. EMBO Rep	28794204
H4K16ac Rep2	SRX2520643	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRX2520643	KC	Schauer T et al. EMBO Rep	28794204
Input H4K16ac Rep1	SRX2520640	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRX2520640	KC	Schauer T et al. EMBO Rep	28794204
Input H4K16ac Rep2	SRX2520642	GSE94115	https://www.ebi.ac.uk/ena/data/view/SRX2520642	KC	Schauer T et al. EMBO Rep	28794204
enhancer STARRseq		GSE40739	https://data.starklab.org/publications/arnold_science_2013/processed.tar.gz	S2	Arnold CD et al. Science 2013	23328393
ATAC-seq S2	SRX314107	GSE103177	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE103177&format=file	S2	Ibrahim MM et al. Nat Commun 2018	30367057
ATAC-seq ISC	SRX9786715	GSE164317	https://www.ncbi.nlm.nih.gov/geo/download/?acc=GSE164317&format=file	ISC	Tauc HM et al. Elife 2021	33724181

Discussion

Mon travail de thèse a porté sur la caractérisation des modes de fonctionnement des deux formes du facteur de transcription Shavenbaby (Svb). Ce travail, essentiellement fondé sur des approches génomiques a permis de montrer que la forme activatrice de Svb (SvbAct) et la forme répressive (SvbRep) se lient à l'ADN sur les mêmes régions enrichies en motifs Ovo spécifiques, OvoQ6 et SvbF7. Ces régions présentent des caractéristiques d'éléments cis-régulateurs et peuvent être considérées comme des enhancers. Ces travaux, ainsi que ceux de Delphine Ménoret (Menoret et al., 2013), ont permis de définir les gènes cibles directs de Svb dans trois types cellulaires différents : les cellules S2, les cellules épidermiques embryonnaires et les cellules souches intestinales (CSI). Ces trois groupes de gènes cibles sont globalement différents, suggérant que Svb **contrôle des ensembles de gènes cibles distincts en fonction du contexte cellulaire**. En ce qui concerne les cellules S2 et les cellules de l'épiderme, cette disparité peut s'expliquer par une organisation/accessibilité chromatinienne différente. En fonction de cette organisation, des sites de fixation svbF7/ovoQ6 peuvent être fixables ou inaccessibles. Lorsque l'on compare les cellules S2 aux CSI, il ne semble pas que l'environnement chromatinien puisse expliquer un comportement de fixation de Svb différent entre ces deux types cellulaires.

Ces observations nous amènent à formuler les hypothèses suivantes :

- L'accessibilité de Svb pour ses sites de fixation nécessite la coopération d'un autre facteur de transcription de type pionnier.
- L'activité de Svb est dépendante de partenaire(s) protéique(s) présent(s) dans les cellules dans lequel il opère.

Lors de notre analyse de « clustering » visant à identifier les ensembles de gènes régulés par Svb au sein des CSI, nous avons pu observer qu'un **groupe de gènes semble être spécifiquement inhibé par la forme répressive de Svb**, (SvbRep). Cette découverte vient quelque peu bousculer notre vision du mode de régulation de Svb car jusqu'ici, les gènes cibles de Svb étaient régulés positivement par SvbAct et négativement par SvbRep. Il est donc possible que SvbRep puisse avoir des fonctions de répression spécifiques.

Enfin, les formes SvbAct et SvbRep ne se distribuent pas de la même façon au sein du noyau : SvbAct est diffus alors que SvbRep est localisé en foci nous laissant supposer que SvbRep puisse agir sur l'organisation globale de la chromatine. Mes études montrent que cette hypothèse est peu probable et des résultats préliminaires issus d'analyse de l'organisation chromatinienne (Hi-C : high throughput 3C) suggèrent que **SvbRep aurait un effet plus localisé et serait impliqué dans des phénomènes de formation de boucles** dont les fonctions restent à définir.

Chapitre 7 Les acteurs régulant la fixation de Svb

L'identité cellulaire est le reflet d'une expression génique spécifique dépendante de l'activité de facteur(s) de transcription qui se fixent sur des régions cis-régulatrices spécifiques et accessibles. Pour Svb - comme pour de nombreux facteurs de transcription - l'accessibilité de ces régions de fixation joue un rôle déterminant dans l'activité transcriptionnelle de ces facteurs. Il existe différentes façons de modifier cet environnement pour favoriser l'activité des facteurs de transcription et nous allons discuter ici des régulations envisageables pour expliquer le comportement de Svb dans les différents types cellulaires.

7.1 Le rôle des facteurs de transcription pionniers.

Les facteurs de transcription (TF) sont des protéines constituées d'un domaine de fixation à l'ADN et de domaine d'activation et/ou de répression de la transcription d'un gène. Ces protéines voient leur séquence fortement conservée entre les espèces et servent de relais entre les signaux intra et extracellulaires (Levine et al., 2005). Chez la drosophile, il est établi que 700 facteurs de transcriptions sont présents. La majorité de ces TF sont des facteurs à doigts de zincs qui régulent l'expression des 13 000 gènes (Hammonds et al., 2013). Grâce à des approches *in vitro* (protection à la DNAase I, de retard sur gel) et le développement des matrices poids position (PMW), il a été montré que les facteurs de transcription, pour se fixer à l'ADN, utilisent des suites de bases nucléotidiques d'environ 6bp, appelées motif ou site de fixation, spécifiques à chaque TF (Small et al., 2020; Zeitlinger, 2020) Du fait de la présence de sites de fixation multiples (Zeitlinger, 2020) et de formation de complexes protéiques, il est possible que la région de fixation atteigne jusqu'à 20bp (Noyes et al., 2008; Crocker et al., 2015). Le nombre de sites de fixation du TF peut également contribuer à la vitesse d'activation du promoteur par l'enhancer. Cela a été montré dans le cas de Zelda, où l'ajout de deux sites CAGGTAG dans l'enhancer *snaE* proches du promoteur diminue la vitesse d'activation de moitié (Dufourt et al., 2018). Pour que les TF se fixent sur leurs sites de fixation putatifs, il est nécessaire que ces derniers soient accessibles, c'est-à-dire présent dans une région de chromatine ouverte. Pour cela, plusieurs stratégies peuvent être mises en place (cf. Figure 36b-e), comme le recrutement de protéines remodelleuses de la chromatine (Mannervik, 2013; Gómez-Díaz et al., 2014 a) ou la coopération entre facteurs de transcription avec l'intervention de TF particuliers appelés « facteurs pionniers » (Small et al., 2020; Zeitlinger, 2020). Ces facteurs pionniers, dont font partie Zelda et Grainyhead, ont la capacité d'ouvrir la chromatine, rendant ainsi accessible les régions cis-régulatrices aux autres facteurs de transcription alors qualifiés de colons (« settlers ») dont Svb (cf. Figure 36-g) (Schulz et al., 2015; Jacobs et al., 2018).

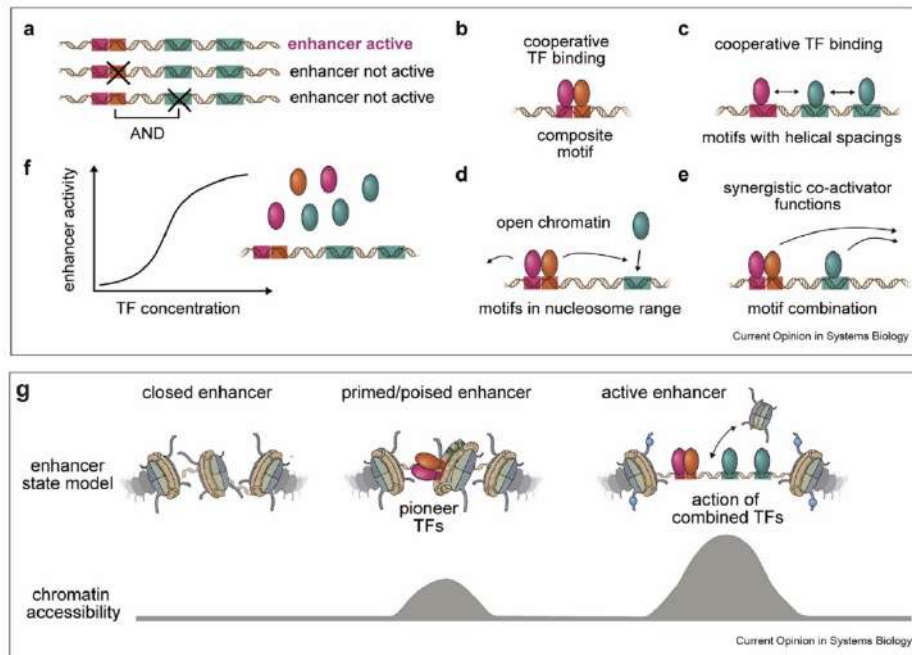


Figure 36 : Modèles de fonctionnements des régions enhancers

a) La mutation dans un ou plusieurs motifs de fixation d'un TF dans un enhancer peut conduire à son inactivation. Les motifs de fixation dans les enhancers peuvent fonctionner soit : b) de façon coopérative par liaison de TF sur des motifs composites proches c) de façon coopérative par liaison de TF sur des motifs espaces d'un tour d'hélice d'ADN d) un TF (facteur pionnier) rend accessible les sites de fixation aux autres TF en ouvrant la chromatine e) en synergie entre plusieurs cofacteurs de transcription par combinaison de plusieurs motifs de fixation différents. f) L'activité de l'enhancer est corrélée de façon sigmoïdale avec la concentration en TF fixé. g) L'accessibilité à l'ADN contraint l'activité des enhancers et corrèle avec la non-fixation des TF à leur motif de fixation. Lorsque la chromatine est fermée les enhancers sont inaccessibles aux TF (à gauche), des facteurs de transcription peuvent ouvrir la chromatine, les facteurs pionniers et rendre accessibles les enhancers (au milieu), les enhancers ainsi préparés peuvent être fixés par d'autres TF et être activés (à droite) (adapté de (Zeitlinger, 2020)

Grainyhead (Grh) qui se comporte comme un facteur pionnier (Jacobs et al., 2018), est le membre fondateur de la famille des facteurs de transcription Grainyhead-like (GRHL). Tout comme Svb, Grainyhead chez la Drosophile est essentiel au développement et au maintien de l'épiderme (Bray et al., 1991; Mace et al., 2005). Il a été montré que les homologues de Grainyhead chez les vertébrés sont impliqués dans les phénomènes de transition Epithélio-Mésenchymateuse (EMT) ainsi que dans le maintien du caractère épithélial des cellules en inhibant les facteurs pro-EMT, ZEB1,2 (Ting et al., 2005; Puisieux et al., 2014; Chung et al., 2016, 2019 b; Nieto et al., 2016; Deng et al., 2021). De plus, comme les facteurs OvoL, la forme GHRL-2 est impliquée dans une forme de dystrophie cornéenne polymorphique postérieur (PPCD). Dans cette forme de PPCD, l'expression de GRHL-2 est augmentée, entraînant une dérégulation de l'EMT conduisant à une déformation de la barrière épidermique de la cornée (Chung et al., 2019 b). Grainyhead est également impliqué dans la maintenance des cellules souches intestinales et leur différenciation (Dominado et al., 2020). Grh, tout comme Svb, possède deux formes Grh-O et Grh-N. Dominado et al., dans une publication sur *BioRxiv*, ont observé que lorsque Grh-O est surexprimé dans les entéroblastes (EB) celui-ci semblerait promouvoir une sur-prolifération des EB. L'expression de Grh-N dans les cellules souches intestinales (CSI) activerait la différenciation de ceux-ci en entérocytes (Dominado et al., 2020). Il est important de noter que Grainyhead est absent tout comme Svb des cellules S2²¹. Dans ces cellules S2, Svb n'est pas capable d'induire la transcription de gènes cibles épidermiques canoniques (*miniature*, *shavenoid*, etc.) (Menoret et al., 2013) qui présentent un

²¹ <https://flybase.org/reports/FBgn0259211>

contexte chromatinien fermé dans les cellules S2. Au sein de l'embryon, cet environnement chromatinien est ouvert. La présence de Svb dans ces cellules n'est pas suffisante pour modifier cet environnement et permettre l'activation de ces gènes. Nous en avons déduit que Svb n'était pas un facteur pionnier et que de ce fait la fixation de Svb dans les cellules S2 est contrainte par l'organisation préexistante de la chromatine.

Aux vues des observations mentionnées ici, il nous semble possible que Grainyhead puisse être le facteur pionnier nécessaire à l'action de Shavenbaby.

Pour tester cette hypothèse, j'ai utilisé les données de ChIP-seq et ChIPmentation de Grh au sein des cellules œil-antenne et de disques d'aile en larve L3 de l'étude de *Jacobs et al.* J'ai comparé les données issues du disque œil-antenne (n=10578 régions uniques) et du disque d'aile (n= 13090 régions) avec les données de fixation de Svb en S2 (n=5343 régions) et cellules embryonnaires (n= 8626 régions). Ces régions de fixation ont été nettoyées des régions de haute fixation. J'ai pu observer que Svb et Grh partagent de façon significative :

- 2362 régions (Svb S2 vs Grh Oeil-antenne)
- 1865 régions (Svb S2 vs Grh Disque d'ailes)
- 1399 régions (Svb Embryo vs Grh Œil-antenne)

Les cellules embryonnaires et les disques d'ailes partagent 778 régions communes mais on ne peut pas dire si ces régions ne sont pas dues au hasard (test de permutation non significatif) (cf. Figure 37). Ces expériences laissent penser que Grh et Svb puissent coopérer afin de mettre en place des programmes transcriptionnels spécifiques. En recherchant la présence du motif de fixation de Grainyhead (aAACcaGTTt) dans les séquences des régions de fixation de Svb j'ai noté que ce motif est faiblement présent dans ces régions, pondérant l'argumentation précédente (cf. Figure 38). Afin de pouvoir conclure sur le lien entre Grh et Svb, nous avons envisagé d'exprimer Grh dans nos lignées de cellules S2 exprimant les différentes formes de Svb puis de réaliser des expériences de RNA-seq. Cela permettrait de tester si la présence de Grh modifie le répertoire des gènes cibles de Svb et permet notamment l'expression des gènes cibles épidermiques. Des expériences de ChIP-seq dirigées contre Grh dans ces lignées cellulaires permettraient également de comparer les sites de fixation de Grh et Svb et de tester si la présence de Grh facilite la fixation de Svb.

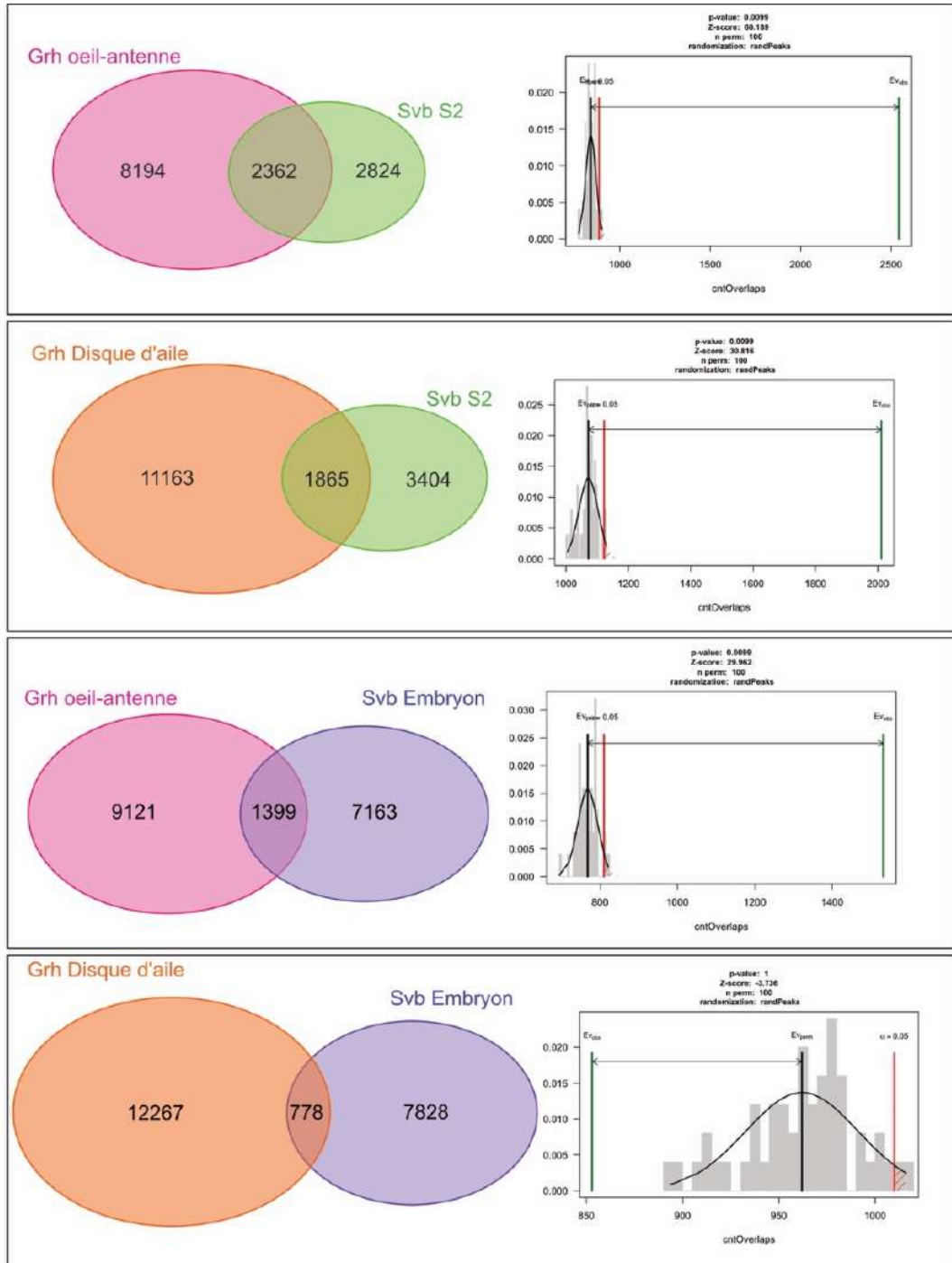


Figure 37 : Diagramme de Venn de comparaison des sites de fixation de Grh (dans les disques d'aile et dans l'œil-antenne) avec Svb (S2 et embryon)

Les régions fixées par Grh ou par Svb sont considérés comme chevauchantes dès qu'elles présentent 1bp de commun. La significativité des chevauchements a été testée par des tests de permutation. Ces tests consistent à recréer une distribution de régions fixées à partir des régions obtenues en ChIP-seq. Cette nouvelle distribution est générée en réattribuant (permutant) de façon aléatoire les caractéristiques de chaque région à une autre région. Ici on effectue 100 permutations. Pour chaque répétition les chevauchements entre les populations Grh et Svb sont comptés et l'espérance V_{perm} est calculée. L'espérance V_{obs} de la distribution des chevauchements pour les régions obtenues en ChIP-seq est également calculée. Ces deux espérances sont comparées, si celle-ci sont distantes et que V_{obs} ne se trouvent pas dans la gaussienne de la distribution des régions obtenues par permutation on dit alors que le chevauchement entre les deux populations de régions de ChIP-seq comparées n'est pas dû au hasard.

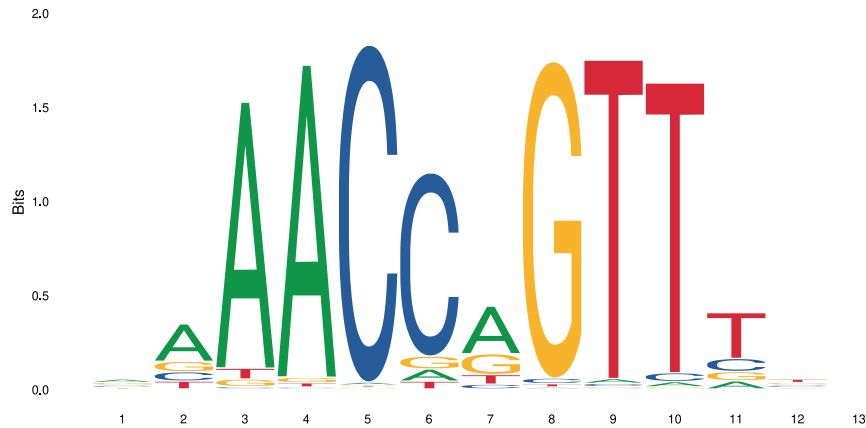


Figure 38 : Site de fixation de Grainyhead
(source : base de données Jaspar)

7.2 Polycomb et Shavenbaby ont une localisation similaire au sein du noyau : peuvent-ils être partenaires ?

Chez la *Drosophila* un complexe majeur de la répression formant des foci au sein des cellules est le complexe Polycomb (Bantignies et al., 2011). Comme mentionné SvbRep présente également une localisation subnucléaire sous forme de foci (Kondo et al., 2007 b). Nous nous sommes donc intéressés à une possible interaction entre Polycomb et Shavenbaby. Avec l'aide de Philippe Valenti nous avons réalisé des immunomarquages de Polycomb et Shavenbaby au sein des cellules S2 exprimant SvbRep. Notre hypothèse était que si Svb et Polycomb étaient partenaires, les deux protéines seraient localisées de façon identique au sein des cellules. Ce n'est pas le cas. Nous n'observons pas de co-localisation de SvbRep et Polycomb au sein des cellules S2 (cf. Figure 39).

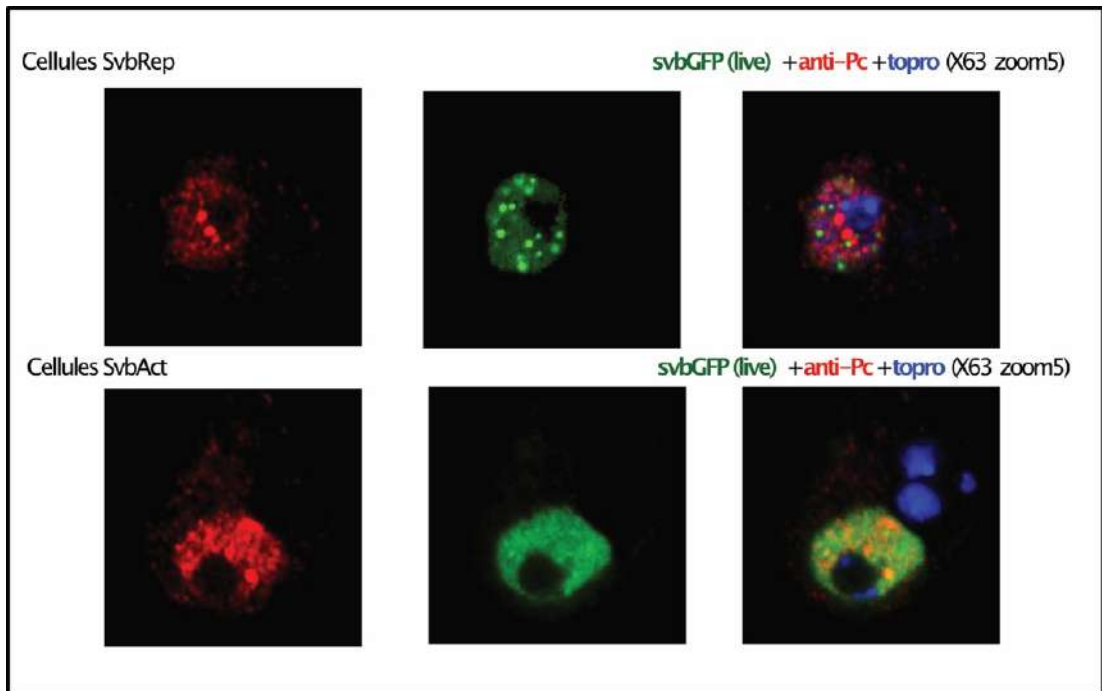


Figure 39 : Observation par immunomarquage de Svb et Polycomb dans les cellules exprimant les formes de Svb. Les foci Polycomb et Svb ne localisent pas ensemble. Svb et Polycomb ne semblent pas être partenaires.

En parallèle des expériences de Philippe, j'ai procédé à une comparaison bio-informatique des sites de fixation de Polycomb et Shavenbaby. J'ai pu observer que le signal de Polycomb anticorrèle avec celui de Shavenbaby. Il semblerait donc que Shavenbaby et Polycomb ne soient pas partenaires directs l'un de l'autre (cf. Figure 40).

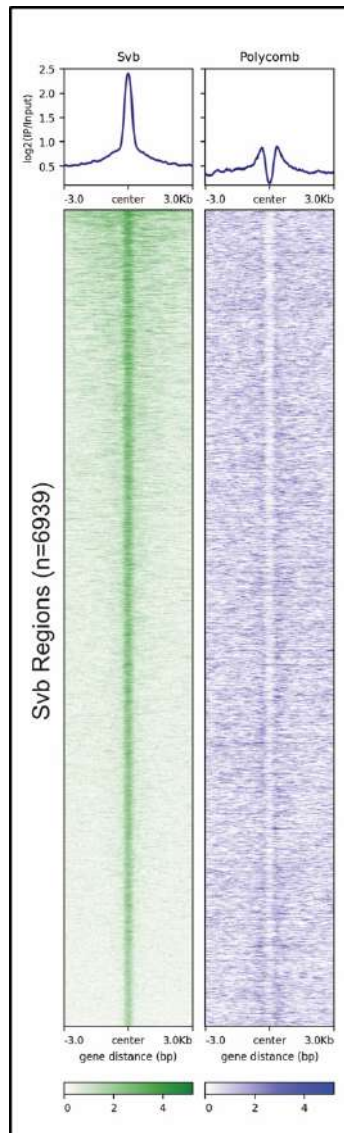


Figure 40 : Svb et Polycomb anticorrèlent sur les régions de fixation de Svb

7.3 Rôle des effecteurs du remodelage de la chromatine ?

D'autres protéines sont impliquées dans la communication enhancer-promoteur : les protéines architecturales ou remodeleurs de la chromatine. Ces protéines architecturales jouent un rôle essentiel dans la formation de structure spatiale de l'ADN et en particulier lors de formation de boucle (Phillips-Cremins et al., 2013; van Bortle et al., 2013; Gómez-Díaz et al., 2014 a). Pour tester si les régions de fixation de Svb présentaient un enrichissement en protéines architecturale, j'ai intégré les données de CHIP-seq disponibles en S2 pour divers remodeleurs (GAF, CBP, BEAF-32, Dref, CAPH2, Rad21/Cohésine, Chromator, Z4, L3mbt, CP190, Mod2.2, ISWI et Nup98) avec les données de CHIP-seq de Svb. Pour ce faire j'ai calculé l'enrichissement du signal de fixation des protéines architecturales sur les régions de fixation de Svb. J'ai pu observer que les régions de fixation de Svb étaient fortement enrichies en GAF, Rad21/Cohésine, ISWI (cf. Figure 41).

Les facteurs GAF ou GAGA Factor, codé par le gène *trithorax-like* chez la Drosophile, ont été décrit comme étant impliqué dans des phénomènes d'activation de la transcription

(Adkins et al., 2006; Gutierrez-Perez et al., 2019; Chetverina et al., 2021; Gaskill et al., 2021). Leur rôle d'activateur vient du fait que ceux-ci rendent accessibles les promoteurs des gènes cibles en déplaçant les nucléosomes (Chetverina et al., 2021). Les facteurs GAFs peuvent aussi être considérés comme des insulateurs et bloquer le déplacement des nucléosomes pour maintenir les régions promotrices accessibles (Adkins et al., 2006; Chetverina et al., 2021).

Le facteur ISWI a été décrit comme faisant parti de plusieurs complexes remodeleurs de la chromatine : NURF, CHRAC, ACF (Deuring et al., 2000). Par exemple au sein du complexe NURF, ISWI en interagissant avec GAF permet l'activation des gènes au cours du développement en modifiant localement la chromatine par modification de l'histone H4 et H2A(262). L'activité de formation de boucles par ISWI est ATP-dépendante et est essentielle chez la Drosophile. En effet, il a été montré que des mutations d'ISWI entraînent une réduction de la viabilité cellulaire, une diminution d'expression des gènes homéotiques et de la segmentation dans les disques imaginaux d'aile et une altération globale de la structure chromatinienne (Deuring et al., 2000).

La Cohésine est un complexe protéique dont la structure forme un anneau (Dorsett, 2019). Cet anneau est recruté par CTCF (ou dCTCF chez la Drosophile) et est impliqué dans la formation des boucles par le rapprochement des régions cis-régulatrices à leur promoteur cibles. Une déplétion de la cohésine réduit les interactions à longue-distance (Dorsett, 2009; Schaaf et al., 2013; Gómez-Díaz et al., 2014 a; Schoenfelder et al., 2019). Cette protéine a été décrite comme étant impliquée dans de nombreux processus dont la différenciation des lymphocytes mais aussi dans la régulation des gènes Hox durant le développement (Gómez-Díaz et al., 2014 b; Ong et al., 2014).

Les facteurs ISWI, GAF et Cohésine enrichies dans les régions de fixation pourraient être impliqués dans la préparation de l'accessibilité des enhanceurs pour le facteur de transcription de Svb. Cependant ces protéines architecturales sont également connues pour être impliquées dans la formation de structure tridimensionnelle de l'ADN.

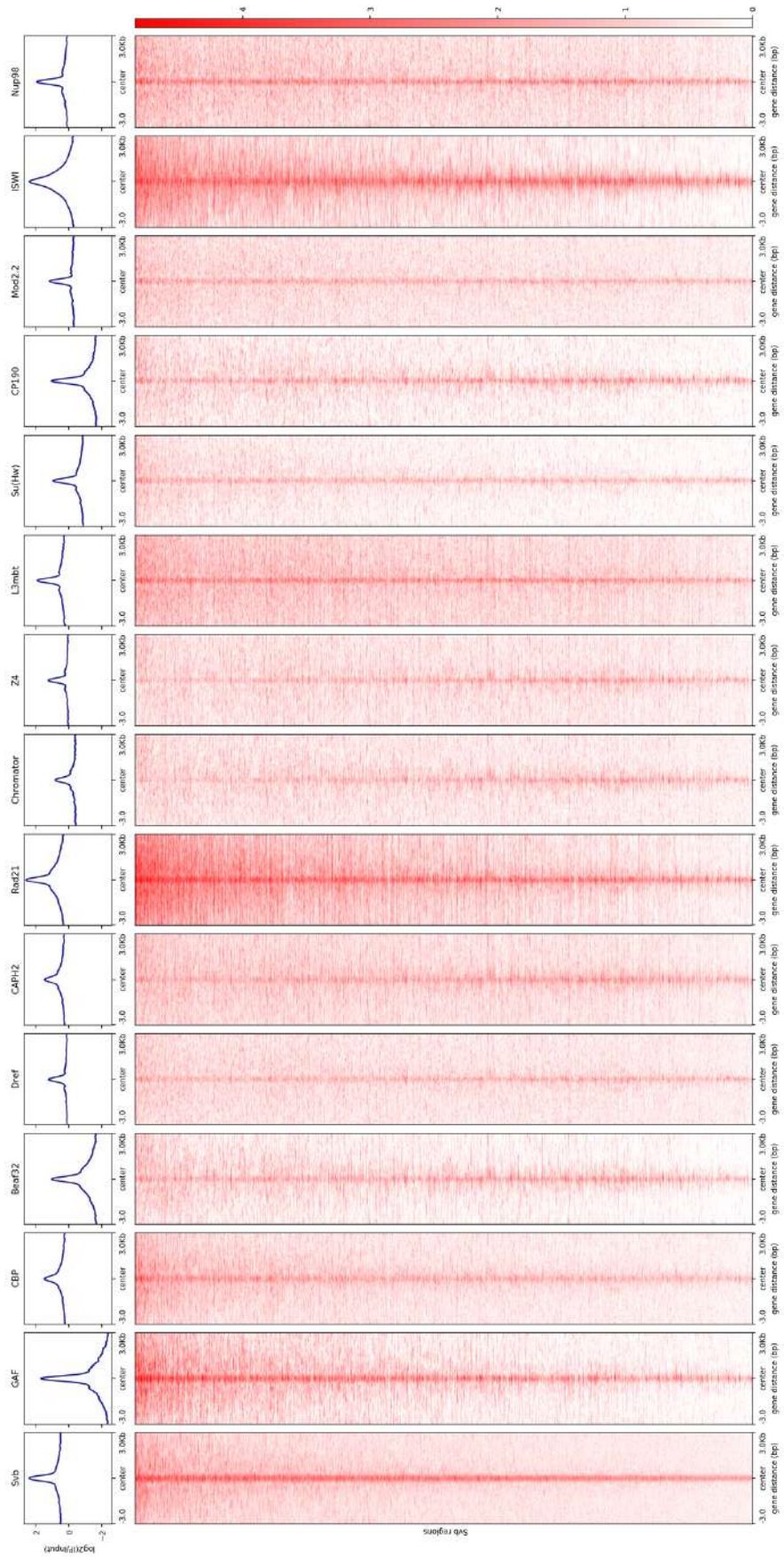


Figure 41 : Enrichissement de différentes protéines architecturales sur les régions de fixation de Svb
On observe un enrichissement en GAF, Rad21/Cohésine et ISWI (résultats préliminaires)

Une autre protéine est connue pour interagir avec GAF et ISWI, il s'agit de Pipsqueak (Gutierrez-Perez et al., 2019). La forme courte (PsqS) participe à la formation de certaines boucles entre les enhancers et les promoteurs. La formation de ces boucles est dépendante de GAF et Pc. La forme longue (PsqL) quant à elle interagit avec ISWI et est impliquée dans des phénomènes de surprolifération (Gutierrez-Perez et al., 2019). Ces observations font que PsqL apparaît être un interacteur potentiel de Svb. Pour tester cette hypothèse, j'ai entrepris de calculer l'enrichissement du signal de PsqL sur les régions de fixation de Svb et inversement. J'ai pu observer un enrichissement du signal de Svb sur les régions fixées par PsqL. J'observe également un enrichissement du signal de PsqL sur les régions de fixation de Svb (cf. Figure 42).

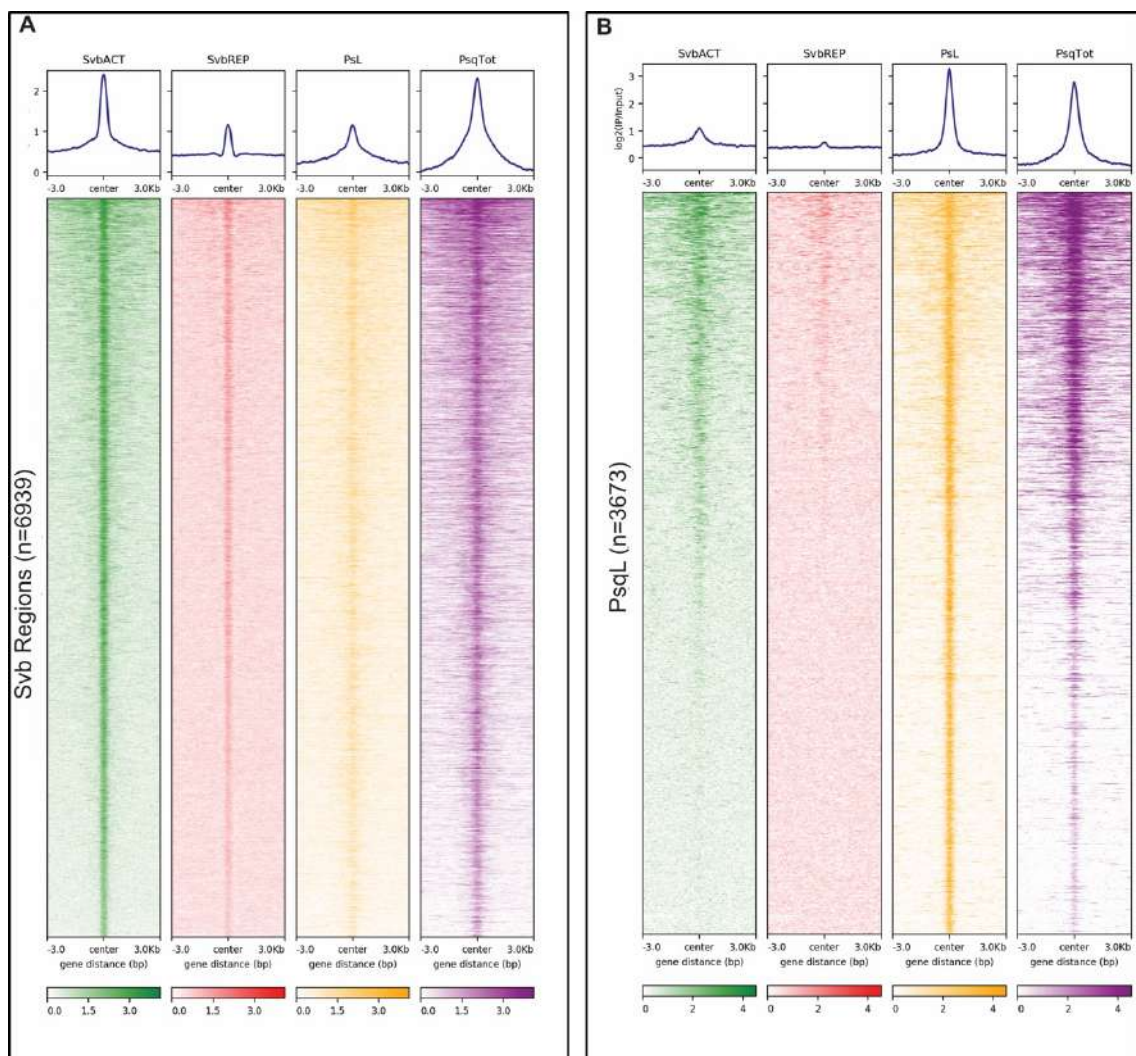


Figure 42 : Pipsqueak pourrait être partenaire de Svb

A) Enrichissement des signaux de fixation de Svb (Act/Rep) et Pipsqueak Long (PsqL) et Pipsqueak Total (PsqL + PsqS) sur les régions Svb B) Enrichissement des signaux de fixation de Svb (Act/Rep) et Pipsqueak Long (PsqL) et Pipsqueak Total (PsqL + PsqS) sur les régions Pipsqueak

De plus lors de la recherche de motifs de fixation au sein des régions fixées par Svb, j'ai pu observer que le motif BTB/iov, motifs sur lequel se fixent les protéines Pipsqueak étaient présents au sein de ces régions (cf. Figure 43). L'ensemble de ces analyses suggère une collaboration possible entre Pipsqueak et Shavenbaby dans la formation de boucles grâce au remodelleur ISWI. Il serait donc intéressant de regarder si les deux protéines ont des

localisations subcellulaires similaires et le cas échéant de tester l'impact de Pipsqueak sur la fixation et l'activité transcriptionnelle de Svb.

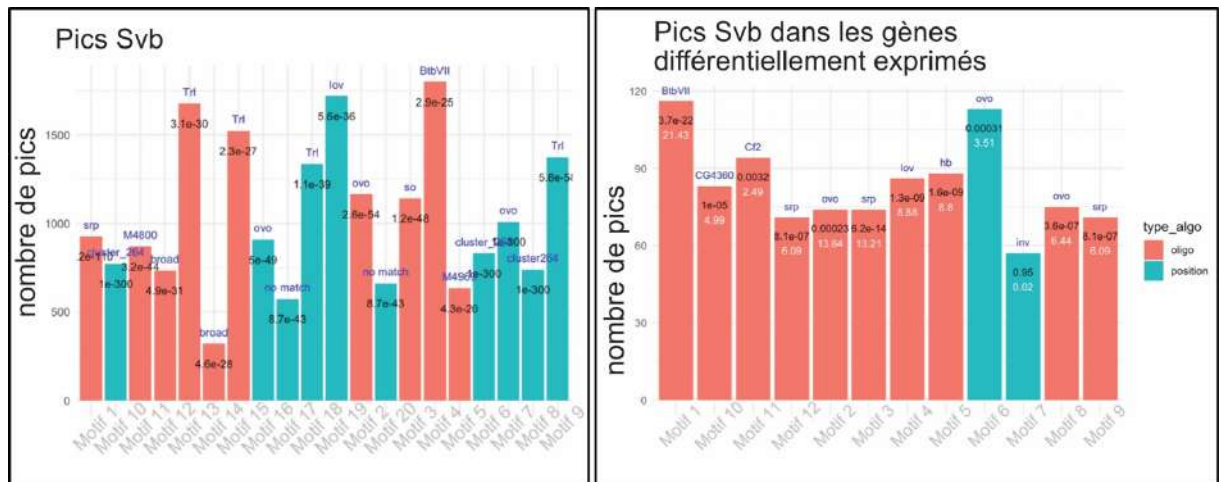


Figure 43 : Les pics Svb (totaux et associés aux gènes dérégulés) présentent outre les motifs ovo des motifs lov/BTB, Trf et srp

Chapitre 8 Implication de SvbRep dans la pause des enhancers

Shavenbaby Activateur et Shavenbaby Répresseur ont été observés comme étant localisés différemment au sein du noyau. Dans le noyau, SvbAct est diffus alors que SvbRep forme des foci. Cette différence de localisation nous a permis de supposer que Svb aurait un rôle dans la compaction de la chromatine et/ou serait lié à des phénomènes de boucles. Nous avons tout d'abord cherché à identifier la nature des foci composé de SvbRep.

Afin de mieux appréhender le rôle de Svb sur l'organisation de la chromatine et la formation de boucles, nous avons mis en place en collaboration avec l'équipe d'Olivier Cuvier, des expériences de capture de la chromatine (Hi-C). En plus de l'organisation globale de la chromatine, il existe des régions de chromatine qui s'associent pour former des structures appelées « domaines associés à la topologie » ou TAD (Furlong et al., 2018; Szabo et al., 2019). Ces TADs sont des régions qui montrent des interactions préférentielles et sont séparés les uns des autres par des frontières ou barrières. Les interactions intra-TADs, par définition, sont fortement présentes, les interactions inter-TADs sont quant à elle peu probables. Parmi les interactions intra-TADs se trouvent les contacts enhancer-promoteur via des boucles (Furlong et al., 2018). Des expériences de suppression ou de modification des frontières montrent une dérégulation dans l'expression des gènes (269). Cependant en s'appuyant sur l'organisation génomiques des chromosomes balanciers *Ghavi-Helm et al.* ont démontré que cette dérégulation ne touche que très peu de gènes (environ 10%) et que c'est une modification plus locale d'interaction enhancer-promoteur qui en serait la cause (Ghavi-Helm et al., 2019). Bien que touchant peu de gènes ce phénomène est associé à des pathologies comme certains cancers (Lupiáñez et al., 2016).

La technique de Hi-C (high throughput 3C) permet d'identifier à l'échelle du génome entier les interactions entre différentes régions et donc de définir les TADs. Son principe est simple. Toutes les interactions sont fixées avec du formaldéhyde. Puis, il y a digestion et ligation des fragments d'ADN avec insertion d'un résidu biotynilé pour purifier sélectivement

les fragments. Ensuite, viennent les étapes classiques d'amplification par PCR et séquençage paired-end (cf. Figure 44) (271, 272).

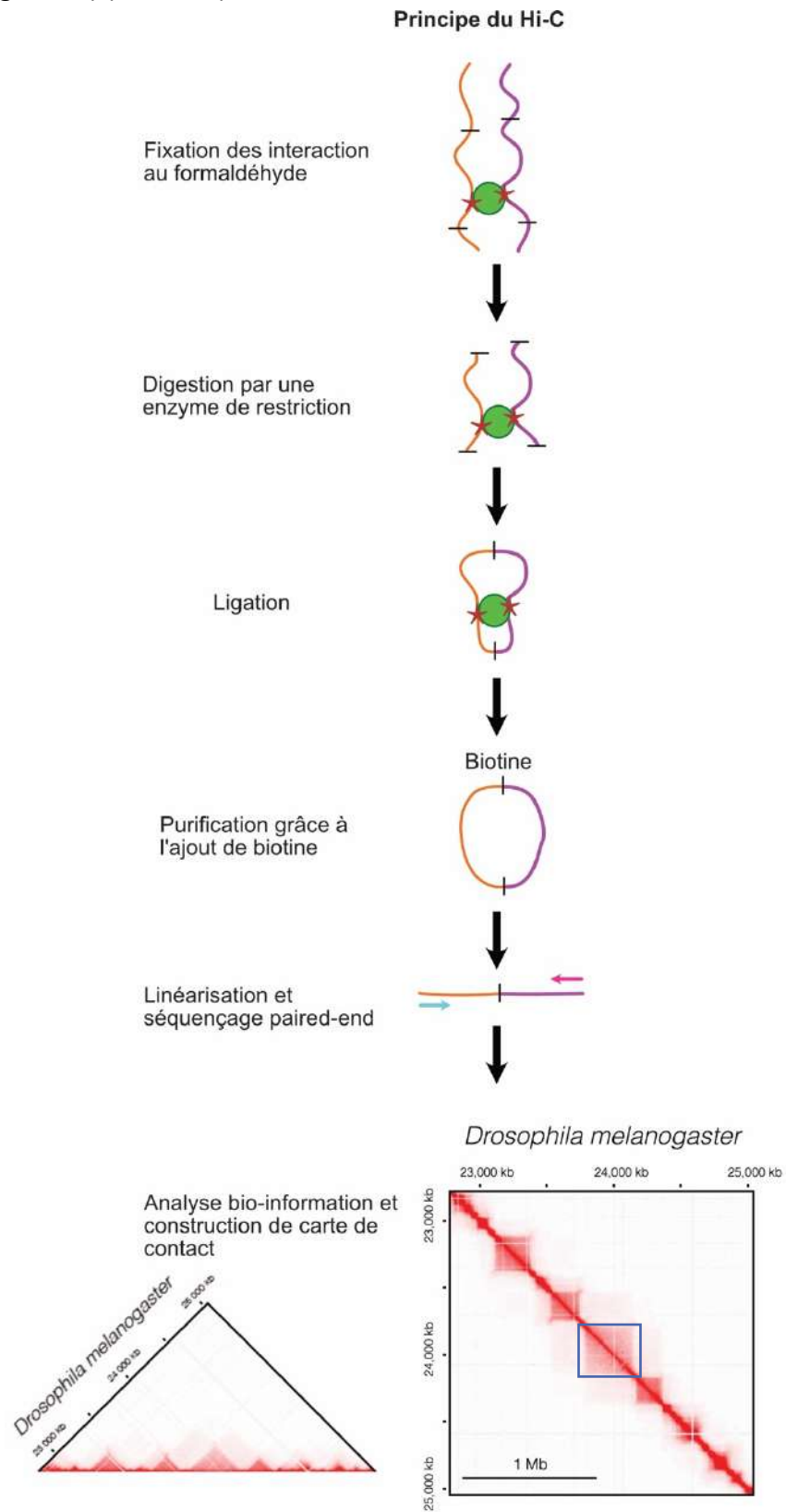


Figure 44 : Principe du Hi-C

Les cartes de contacts présentent les TADs (exemple d'un TAD encadré) au sein chez la *Drosophila* (adapté de (Szabo et al., 2019))

La *Drosophila* possède entre 2000 et 4000 TADs, en fonction de l'échelle à laquelle on place notre observation, d'en moyenne 100kb (Szabo et al., 2019). Les TADs chez la mouche ont été identifiés pour la première fois par *Sexton et al.* par des techniques de Hi-C dans des embryons entiers (Sexton et al., 2012). Lors de cette étude, il a également été observé que ces TADs corrèlent fortement avec les marques épigénétiques. Cela a permis de classer les TADs en quatre catégories :

- Les TADs transcriptionnellement actifs ou TADs actifs associés avec des signatures des modifications des histones actives comme H3K4me3 et H3K36me3
- Les TADs réprimés par Polycomb, ou TADs PcG, sont enrichis en H3K27me3 et en protéines du groupe Polycomb
- Les TADs ne présentant aucune signature connue spécifique, TADs « null » ou « void »,
- Les TADs de l'hétérochromatine associé à un enrichissement en H3K9me2, HP1 et Su(var)3-9 (Sexton et al., 2012)

Les TADs associés à de la répression (PcG, null, associés à l'hétérochromatine) sont majoritaires au sein du génome. Les bordures des TADs chez la *Drosophila* présentent des protéines insultrices comme BEAF-32, Chromator, CP190, M1BP mais aussi à une combinaison de ces différentes protéines architecturales BEAF-32/ Chromator ; BEAF32/CP190 ; Cohesin/ TFIIC (Szabo et al., 2019). Ces bordures sont moins condensées et contiennent des gènes exprimés, de la chromatine plus relâchée et des marques d'histones actives (Hou et al., 2012).

Les TADs se mettent en place durant l'activation du génome zygotique (ZGA). Avant la ZGA le génome est déstructuré même si des semblants de bordures enrichis en gènes de ménage sont présents. Par des approches novatrices combinant observation de la transcription et de la conformation de la chromatine dans des embryons intacts appelé Hi-M(275), *Martin Espinola et al.* ont observé que la formation des boucles enhancer-promoteur (mais aussi enhancer-enhancer sous forme de hub) est antécédente à la formation des TADs. La formation de hubs d'enhancers est dépendant de la protéine Zelda (Espinola et al., 2021). Les TADs se forment ensuite et il est possible d'observer des réarrangements à longue distance au sein des TADs dû à la communication enhancer-promoteur (Szabo et al., 2019).

Les résultats préliminaires générés à partir des expériences de Hi-C réalisées en présence de SvbAct ou SvbRep suggèrent que Svb n'agit pas de façon globale sur la structure de la chromatine et que SvbRep est impliqué dans la formation de boucles à longue distance au sein des TADs (cf. Figure 45).

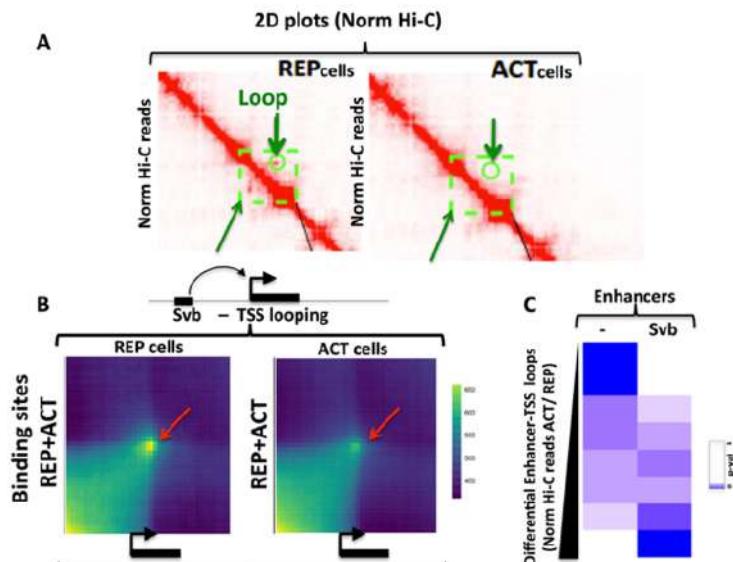


Figure 45 : La présence de SvbRep induit la formation de boucle enhancer-promoteur

A) Cartes de probabilité de contact issus des Hi-C dans les cellules présentant SvbAct ou SvbRep. Le signal est normalisé. On peut observer la présence de TADs dans les lignées cellulaires. En présence de SvbRep, il y a formation de boucles (flèche verte). La boucle disparaît avec le changement de SvbRep en SvbAct. B) Estimation des contacts à longue distance entre les enhancers et les promoteurs par agrégation des reads de Hi-C autour du TSS. Cette estimation a été réalisée pour 561 enhancers fixés par Svb et distants du TSS (>20kb). On observe un enrichissement significatif du signal de formation d'une boucle entre les enhancers fixés par Svb (Wilcoxon : p value < $10e-93$ contre p value 0.99 pour ceux non fixés). C) Les enhancers fixés sont significativement retrouvés dans des boucles (test de Fisher). La liste (-) est une liste aléatoire d'enhancers ayant les mêmes caractéristiques de distances au TSS que les enhancers fixés par Svb (résultats préliminaires issus de la collaboration avec l'équipe d'Olivier Cuvier)

Ces observations suggèrent que SvbRep puisse être impliqué dans des phénomènes de préparation des promoteurs par rapprochement des enhancers pour une activation rapide. Ce phénomène, appelé mise en pause des enhancers, permet de garantir une activation rapide et précise des gènes développementaux en réponse à un signal donné. Cette régulation précise est essentielle lors du développement (Lagha et al., 2012) et lorsque les régions cis-régulatrices sont fixées par leurs facteurs de transcription, la machinerie de transcription et l'environnement chromatinien avoisinant commencent à être modifiés. Puis seulement en présence d'un signal externe ces enhancers perdent les dernières marques de la répression et activent le promoteur du gène cible. Pour Svb, le signal d'activation serait donné par l'ecdysone et transmis par les peptides Pri. L'exemple du facteur de transcription Zelda est couramment utilisé pour parler ce phénomène chez la Drosophile. Zelda est essentiel lors de l'embryogénèse précoce et en particulier lors de la transition materno-zygotique. En effet, ce facteur pionnier régule l'activation d'une centaine de gènes de façon rapide, en 2-3h après la fertilisation de l'œuf (Gutierrez-Perez et al., 2019 ; Lagha et al., 2012). Il a été montré que les enhancers en pause présentaient un taux faible en H3K27ac et un taux élevé de H3K4me1 (Koenecke et al., 2017). De ce fait pour confirmer le rôle de Svb dans le phénomène de pause il serait intéressant de calculer le taux de H3K27ac et de H3K4me1 dans les régions observées pour être impliquées dans des boucles à longue distance en Hi-C.

Chapitre 9 Shavenbaby dans les cellules intestinales

Lors de ma dernière année de doctorat, j'ai eu la possibilité de co-encadrer Alexia Rivero, étudiante de Master 2 Bio-informatique et Biologie des Systèmes pour son stage de fin d'études. Lors de ce stage, nous avons pour objectif de caractériser les gènes cibles de Shavenbaby au sein des cellules souches intestinales (CSI) par approche de clustering. Ces analyses ont été intégrées à la publication 3 de la partie Résultats. Alexia a développé une méthode impliquant deux approches de regroupement : la **classification hiérarchique ascendante** et le **clustering par k-moyennes**. Ces deux méthodes de classification non supervisées s'appuient sur deux algorithmes différents. L'objectif de cela consistait à trouver les gènes cibles de Svb robustes avec deux méthodes de clustering indépendantes. Il faut noter que l'on considère un gène cible de Svb si celui-ci est activé par SvbAct et réprimé par SvbRep et/ou réprimé en absence de Svb (RNAi-svb).

La classification hiérarchique ascendante s'appuie sur un test de corrélation de Spearman (Spearman, 1904). Ce test, non paramétrique, est dit robuste et se base sur le rang des données. Il ne dépend donc pas de la distribution des données et est ainsi adapté à l'analyse des données RNA-seq (de Siqueira Santos et al., 2014). Ces données doivent toutefois être triées, filtrées, centrées et réduites. La filtration permet de réduire le bruit dû aux gènes peu ou pas exprimés. Le centrage et la réduction permet de minimiser l'impact des gènes ayant un comportement aberrant tout en permettant de normaliser le signal et ainsi de les comparer. L'algorithme de clustering hiérarchique regroupe les éléments les plus proches en une seule entité (classe) (cf. Figure 46). Le rapprochement des données s'est fait en suivant la méthode de Ward (Ward, 1963). Cette méthode a pour objectif de réduire l'inertie intraclasse ou de maximiser l'inertie interclasse. Ce clustering est itératif et s'arrête quand les individus sont regroupés en une seule et même classe (cf. Figure 46) (Johnson, 1967).

Algorithme du clustering hiérarchique ascendant

Initialisation

Tous les individus a ... n sont considérés comme des classes soit n classe

Objectif

Obtenir un nombre de classe < n soit n = 1

Itération

- Calcul de la distance de dissimilarité entre deux classes C_x et C_y avec x appartenant à C_x et y appartenant de C_y
- Aggrégation de x et y en une seule classe C pour lesquelles la valeur de distance est minimale $dx;y = \min(\text{dissimilarité})$
- Mise à jour de la matrice de distance avec les nouvelles classes $C_x;y$

Fin

Tous les individus sont regroupés en une seule grande classe C_n

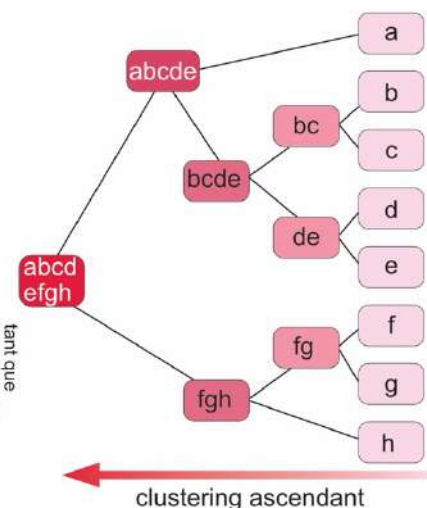


Figure 46 : Algorithme du clustering hiérarchique ascendant

Grâce à cette approche, 5 groupes permettant d'expliquer les données de RNA-seq ont été définis et une première liste de 272 gènes cibles de Svb est ressortie. Ce sont ces gènes que nous utilisons lors de la comparaison des gènes cibles dans les trois types cellulaires.

Suite à cela, Alexia a exécuté un clustering par k-moyennes sur données filtrées, réduites et centrées. La philosophie de l'algorithme de k-moyennes est de regrouper les individus en k groupes, k étant défini par l'utilisateur et correspondant aux k centres des groupes. Puis les données sont regroupées en fonction de leur distance aux centres du groupe. Les centres étant mobiles, leur situation en fonction des données change à chaque itération. L'algorithme prend fin quand les centres sont immobiles : on dit que la distance des individus aux centres a atteint un minimum que l'on espère global. Pour s'assurer d'atteindre ce minimum global, le nombre d'itérations à effectuer doit être fortement supérieur au nombre d'individus constituant le jeu de données (cf. Figure 47) (281).

Algorithme du clustering k-moyennes

Initialisation

Les individus sont dispersés sur une grille

Les k centres mobiles sont définis sur cette grille => partition initiale

k est défini par l'utilisateur

Objectif

Obtention d'une partition stable

Itération

- La distance des individus aux centres est calculée
- Les individus sont attribués au centre le plus proche
- Nouveaux centres sont générés en fonction des regroupements => nouvelle partition

tant que partition non stable

Fin

Convergence vers une partition stable, plus aucun point ne bouge. On atteint un minimum que l'on suppose global. Le nombre d'itération doit être largement supérieur aux nombres d'individu à classer.

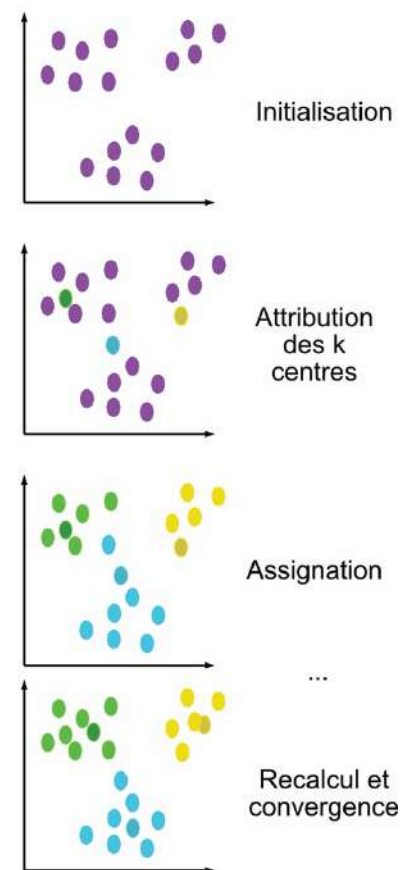


Figure 47 : Algorithme du clustering de k-moyennes

Suite aux traitements des données par le clustering k-moyennes (répétés plusieurs fois pour s'affranchir du biais dû aux choix des centres), un groupe robuste composé de 201 gènes cibles de Svb a été défini. Il est important de noter que ce groupe, malgré les nombreuses itérations, est toujours ressorti. En comparant les deux listes de gènes obtenues, une liste de 195 gènes cibles de Svb a été définie qui constitue le socle pour décrypter plus avant le rôle de Svb dans les cellules souches.

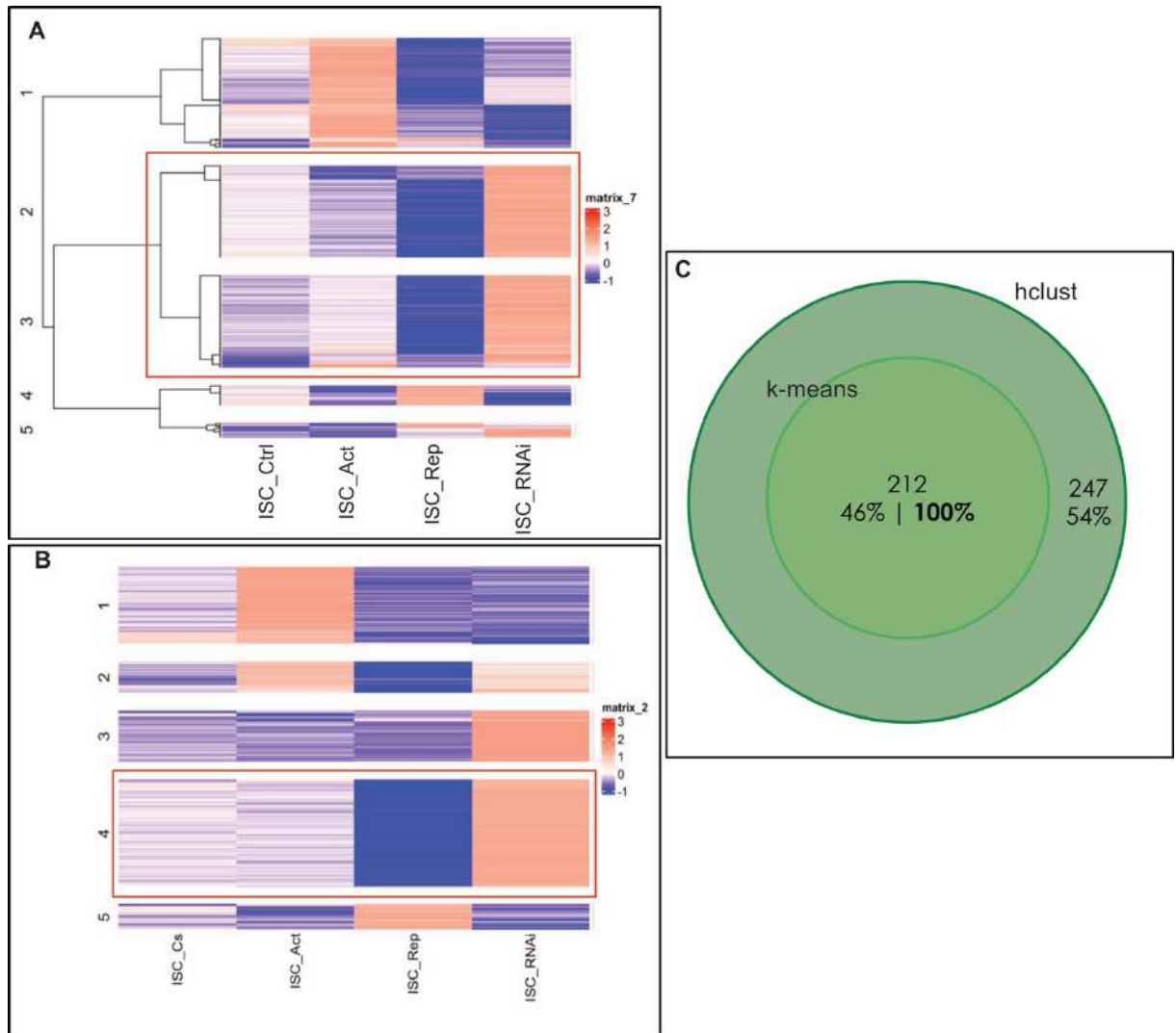


Figure 48 : Détermination du groupe de gènes cibles spécifiques de *SvbRep* par clustering
 A) Analyse par clustering hiérarchique B) Analyse par k-moyennes (k-means) C) Comparaison des deux groupes de gènes déterminés en A et B, 212 gènes cibles de *SvbRep* sont robustes aux deux méthodes de détermination

9.1 Vers une fonction spécifique de *SvbRep* dans le lignage intestinal

Dans les études que nous avons menées avec Alexia un second groupe est apparu comme étant robuste lors des analyses de k-moyennes. Ce groupe, bien que divisé en deux est également présent dans les analyses de clustering hiérarchique. Ce groupe de gènes composé de 212 gènes semblent être **spécifiquement régulés par *SvbRep***. En effet, la présence de *SvbRep* réduit l'expression de ces gènes, la perte de fonction de *Svb* (*RNAi-svb*) augmente au contraire leur expression alors que la forme *SvbAct* ne semble pas avoir d'effet (cf. Figure 48). C'est la première fois qu'on observe *in vivo* un rôle spécifique de *SvbRep*, indépendant de *SvbAct*. Il serait donc intéressant de pousser l'étude de ce groupe de gènes pour appréhender la fonction de *SvbRep* dans les cellules souches intestinales mais aussi au sein du lignage intestinal puisque seul *SvbRep* est présent dans les entéroblastes matures issus de la différenciation des cellules souches.

Une manière d'appréhender la fonction de ce groupe de gènes est de procéder à une analyse de leur ontologie (Gene Ontology). Une ontologie par définition est « *une représentation formelle d'un ensemble de connaissances dans un domaine donné. Les ontologies consistent généralement en un ensemble de classes (ou termes ou concepts) avec des relations qui opèrent entre elles*²² ». L'ontologie des gènes (GO) décrit les connaissances biologiques sous trois grands thèmes :

- Fonction moléculaire : ces termes désignent les fonctions moléculaires (catalyse, transport) des protéines codées par les gènes. Aucune information n'est donnée sur le lieu ni le processus.
- Composants cellulaires : localisation subcellulaire où la fonction du gène a lieu (mitochondrie, noyau, etc.). Aucune information n'est donnée sur la fonction ou le processus.
- Processus Biologique : Processus dans lequel est impliqué le gène (réparation de l'ADN, formation du cytosquelette, etc.). Aucune information sur le lieu ni la fonction n'est donnée.

Les termes permettant de décrire les gènes en fonction de ces trois domaines sont appelés GOTerms. Bien que développée dans l'optique de faciliter et homogénéiser la classification des gènes sous trois grands aspects, l'utilisation des GOTerms peut être fastidieuse et peu informative. En utilisant les logiciels de recherche de GOTerm GOrilla²³ (282) et DAVID²⁴ (283), nous obtenons pour la liste de gènes cibles de SvbRep pour les processus biologiques les plus significativement représentés : **organisation des composants cellulaires, métabolismes des ARNs, régulation de la transcription, différenciation cellulaire/développement, organisation de la chromatine, morphogénèse** (cf. Figure 49). Il est intéressant de noter que plus de 10 pourcents de ces gènes sont liés à l'organisation de la chromatine. Sont notamment présents *E(Pc)*, *Pcl*, *ph-p*, *upSET* et *Spps* pour le complexe Polycomb, *Bap60* et *Bap170* pour le complexe Brahma, *Atxn7* pour le complexe SAGA et *Ncoa6* pour le complexe Trithorax-related. Plusieurs de ces complexes comme le complexe Brahma et Polycomb sont connus pour réguler l'activité des cellules souches intestinales (199) et au moins deux sont reliés à la voie de signalisation Hippo, Brahma et Trithorax-related qui interagit avec Svb. Il sera donc important dans le futur de mieux appréhender la relation entre SvbRep et ces complexes de remodelage de la chromatine dans le processus de différenciation des cellules souches. SvbRep est aussi spécifiquement exprimé dans les entérocytes pour lesquelles nous disposons aussi des données de RNA-seq et dont nous devons finir l'analyse. Il sera donc utile pour la compréhension des fonctions de SvbRep de comparer les gènes cibles de SvbRep dans les EC avec les gènes cibles spécifiques de SvbRep identifiés dans les cellules souches. Cette analyse est d'autant plus importante que chez les vertèbres, les facteurs OvoL agissent principalement comme des répresseurs transcriptionnels (Ye et al., 2016; Wang et al., 2017; Wu et al., 2017).

²² <http://geneontology.org/docs/ontology-documentation/>

²³ <http://cbl-gorilla.cs.technion.ac.il/>

²⁴ <https://david.ncifcrf.gov/>

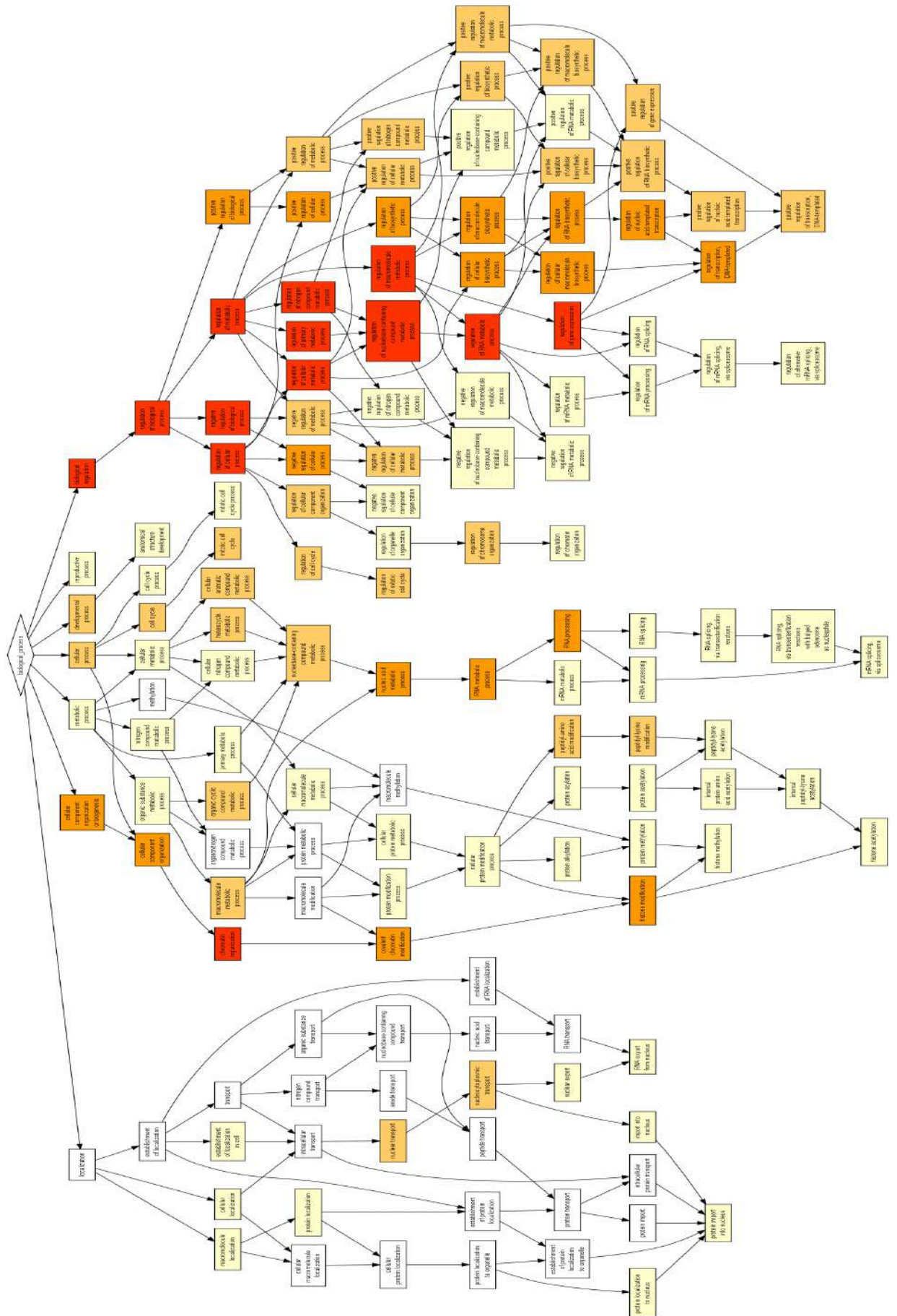


Figure 49 : Gene Ontology (Processus Biologique) des gènes cibles de SvbRep (GORilla)

Conclusion

Pour conclure, mes travaux de thèse, ont permis de mieux comprendre le mode de fonctionnement du facteur de transcription Shavenbaby (Svb) à travers l'identification de ses sites de fixation sur le génome et l'identification de ses gènes cibles. Nous avons ainsi mis en évidence que la fixation de SvbAct et REP est dépendante non seulement de l'accessibilité à la chromatine mais aussi de co-facteur(s) restant à identifier. Une des découvertes inattendues à été la mise en lumière de fonction spécifique de la forme SvbRep dont la fonction reste à définir. SvbRep qui s'accumule en foci semble induire la formation de boucle à longue-distance entre les enhancers et les promoteurs des gènes cibles. Nous avons pu observer que les régions de fixation de Svb étaient enrichies en protéines architecturales (GAF, ISWI et Cohésine). Ces protéines sont également impliquées dans la formation de boucle et une hypothèse possible serait que SvbRep capture les enhancers en formant des boucles grâce à GAF, ISWI et la Cohésine. Ces enhancers ainsi en pause pourraient être rapidement activés en réponse à un signal ecdysone.

Références

Adams, M. D.; Celniker, S. E.; Holt, R. A.; Evans, C. A.; Gocayne, J. D.; Amanatides, P. G.; Scherer, S. E.; Li, P. W.; Hoskins, R. A.; Galle, R. F.; George, R. A.; Lewis, S. E.; Richards, S.; Ashburner, M.; Henderson, S. N.; Sutton, G. G.; Wortman, J. R.; Yandell, M. D.; Zhang, Q. et al., 2000: The genome sequence of *Drosophila melanogaster*. *Science (New York, N.Y.)*, **287**, 2185–2195.

Adkins, N. L.; Hagerman, T. A.; Georgel, P., 2006: GAGA protein: a multi-faceted transcription factor. This paper is one of a selection of papers published in this Special Issue, entitled 27th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. *Biochemistry and Cell Biology*, **84**, 559–558.

Ahmed, S. M. H.; Maldera, J. A.; Kronic, D.; Paiva-Silva, G. O.; Pénalva, C.; Teleman, A. A.; Edgar, B. A., 2020: Fitness trade-offs incurred by ovary-to-gut steroid signalling in *Drosophila*. *Nature*, **584**, 415–419.

Akam, M., 1987: The molecular basis for metamerism in the *Drosophila* embryo. *Development*, **101**, 1–22.

al Hayek, S.; Alsawadi, A.; Kambris, Z.; Boquete, J.; Bohère, J.; Immarigeon, C.; Ronsin, B.; Plaza, S.; Lemaitre, B.; Payre, F.; Osman, D., 2021: Steroid-dependent switch of OvoL/Shavenbaby controls self-renewal versus differentiation of intestinal stem cells. *The EMBO Journal*, **40**, e104347.

Alkhatib, S. G.; Landry, J. W., 2011: The nucleosome remodeling factor. *FEBS letters*, **585**, 3197–3207.

ALLFREY, V. G.; FAULKNER, R.; MIRSKY, A. E., 1964: ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America*, **51**, 786–794.

Andrews, J.; Garcia-Estefania, D.; Delon, I.; Lü, J.; Mével-Ninio, M.; Spierer, A.; Payre, F.; Pauli, D.; Oliver, B., 2000: OVO transcription factors function antagonistically in the *Drosophila* female germline. *Development (Cambridge, England)*, **127**, 881–892.

Antonello, Z. A.; Reiff, T.; Ballesta-Illan, E.; Dominguez, M., 2015: Robust intestinal homeostasis relies on cellular plasticity in enteroblasts mediated by miR-8–Escargot switch. *The EMBO Journal*, **34**, 2025–2041.

Arif, S.; Kittelmann, S.; McGregor, A. P., 2015: From shavenbaby to the naked valley: trichome formation as a model for evolutionary developmental biology. *Evolution & Development*, **17**, 120–126.

Arnold, C. D.; Gerlach, D.; Stelzer, C.; Boryn, L. M.; Rath, M.; Stark, A., 2013: Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science*, **339**, 1074–1077.

Arnold, C. D.; Zabidi, M. A.; Pagani, M.; Rath, M.; Schernhuber, K.; Kazmar, T.; Stark, A., 2017: Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nature Biotechnology*, **35**, 136–144.

Arnold, P. R.; Wells, A. D.; Li, X. C., 2020: Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Frontiers in Cell and Developmental Biology*, **7**, 377.

Aughey, G. N.; Southall, T. D., 2016: Dam it's good! DamID profiling of protein-DNA interactions. *Wiley Interdisciplinary Reviews: Developmental Biology*, **5**, 25–37.

Aulner, N.; Monod, C.; Mandicourt, G.; Jullien, D.; Cuvier, O.; Sall, A.; Janssen, S.; Laemmlli, U. K.; Käs, E., 2002: The AT-Hook Protein D1 Is Essential for *Drosophila melanogaster* Development and Is Implicated in Position-Effect Variegation. *Molecular and Cellular Biology*, **22**, 1218–1232.

Banerji, J.; Rusconi, S.; Schaffner, W., 1981: Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell*, **27**, 299–308.

Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T., 2001: Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, **410**, 120–124.

Bannister, A. J.; Kouzarides, T., 2011: Regulation of chromatin by histone modifications. *Cell research*, **21**, 381–395.

Bantignies, F.; Cavalli, G., 2011: Polycomb group proteins: repression in 3D. *Trends in Genetics*, **27**, 454–464.

Bardin, A. J.; Perdigoto, C. N.; Southall, T. D.; Brand, A. H.; Schweisguth, F., 2010: Transcriptional control of stem cell maintenance in the *Drosophila* intestine. *Development*, **137**, 705–714.

Beyenbach, K. W.; Skaer, H.; Dow, J. A. T., 2010: The developmental, molecular, and transport biology of Malpighian tubules. *Annual review of entomology*, **55**, 351–374.

Bi, H.; Xu, X.; Li, X.; Zhang, Y.; Huang, Y.; Li, K.; Xu, J., 2019: CRISPR Disruption of BmOvo Resulted in the Failure of Emergence and Affected the Wing and Gonad Development in the Silkworm *Bombyx mori*.

Biran, A.; Meshorer, E., 2020: Chromatin plasticity in pluripotent and cancer stem cells. *Stem Cell Epigenetics*. Elsevier, pp. 207–230.

Biteau, B.; Jasper, H., 2014: Slit/Robo Signaling Regulates Cell Fate Decisions in the Intestinal Stem Cell Lineage of *Drosophila*. *Cell Reports*, **7**, 1867–1875.

Bohère, J.; Mancheno-Ferris, A.; al Hayek, S.; Zanet, J.; Valenti, P.; Akino, K.; Yamabe, Y.; Inagaki, S.; Chanut-Delalande, H.; Plaza, S.; Kageyama, Y.; Osman, D.; Polesello, C.; Payre, F., 2018: Shavenbaby and Yorkie mediate Hippo signaling to protect adult stem cells from apoptosis. *Nature Communications*, **9**, 5123.

Boros, I. M., 2012: Histone modification in *Drosophila*. *Briefings in Functional Genomics*, **11**, 319–331.

Boumard, B.; Bardin, A. J., 2021: An amuse-bouche of stem cell regulation: Underlying principles and mechanisms from adult *Drosophila* intestinal stem cells. *Current Opinion in Cell Biology*, **2021**, 58–68.

Bray, S. J.; Kafatos, F. C., 1991: Developmental function of Elf-1: an essential transcription factor during embryogenesis in *Drosophila*. *Genes & Development*, **5**, 1672–1683.

Buchon, N.; Osman, D., 2015: All for one and one for all: Regionalization of the *Drosophila* intestine. *Insect biochemistry and molecular biology*, **67**, 2–8.

Bulger, M.; Groudine, M., 1999: *Looping versus linking: toward a model for long-distance gene activation*.

Busson, D.; Gans, M.; Komitopoulou, K.; Masson, M., 1983: Genetic Analysis of Three Dominant Female-Sterile Mutations Located on the X Chromosome of *DROSOPHILA MELANOGASTER*. *Genetics*, **105**, 309–325.

Cardozo Gizzi, A. M.; Cattoni, D. I.; Fiche, J. B. B.; Espinola, S. M.; Gurgo, J.; Messina, O.; Houbron, C.; Ogiyama, Y.; Papadopoulos, G. L.; Cavalli, G.; Lagha, M.; Nollmann, M., 2019:

Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell.*, **74**, 212-222.e5.

Celniker, S. E.; Rubin, G. M., 2003, November 28: The *Drosophila Melanogaster* Genome. *Annual Review of Genomics and Human Genetics*.

Chang, B.; Chen, Y.; Zhao, Y.; Bruick, R. K., 2007: JMJD6 Is a Histone Arginine Demethylase. *Science.*, **318**, 444–447.

Chanut-Delalande, H.; Fernandes, I.; Roch, F.; Payre, F.; Plaza, S., 2006: Shavenbaby couples patterning to epidermal cell shape control. (Martinez Arias, A., Ed.) *PLoS Biology.*, **4**, 1549–1561.

Chanut-Delalande, H.; Hashimoto, Y.; Pelissier-Monier, A.; Spokony, R.; Dib, A.; Kondo, T.; Bohère, J.; Niimi, K.; Latapie, Y.; Inagaki, S.; Dubois, L.; Valenti, P.; Polesello, C.; Kobayashi, S.; Moussian, B.; White, K. P.; Plaza, S.; Kageyama, Y.; Payre, F., 2014: Pri peptides are mediators of ecdysone for the temporal control of development. *Nature Cell Biology.*, **16**, 1035–1044.

Chen, H.; Xu, Z.; Mei, C.; Yu, D.; Small, S., 2012: A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. *Cell.*, **149**, 618–629.

Chetverina, D.; Erokhin, M.; Schedl, P., 2021: GAGA factor: a multifunctional pioneering chromatin protein. *Cellular and Molecular Life Sciences.*, **78**, 4125–4141.

Choubey, P. K.; Nandy, N.; Pandey, A.; Roy, J. K., 2020: Rab11 plays a key role in stellate cell differentiation via non-canonical Notch pathway in Malpighian tubules of *Drosophila melanogaster*. *Developmental Biology.*, **461**, 19–30.

Chung, D. D.; Frausto, R. F.; Cervantes, A. E.; Gee, K. M.; Zakharevich, M.; Hanser, E. M.; Stone, E. M.; Heon, E.; Aldave, A. J., 2017: Confirmation of the OVOL2 Promoter Mutation c.-307T>C in Posterior Polymorphous Corneal Dystrophy 1. (Janecke, A. R., Ed.) *PLOS ONE.*, **12**, e0169215.

Chung, D. D.; Zhang, W.; Jatavallabhula, K.; Barrington, A.; Jung, J. Y.; Aldave, A. J., 2019a: Alterations in GRHL2-OVOL2-ZEB1 axis and aberrant activation of Wnt signaling lead to altered gene transcription in posterior polymorphous corneal dystrophy. *Experimental Eye Research.*

Chung, D. D.; Zhang, W.; Jatavallabhula, K.; Barrington, A.; Jung, J. Y.; Aldave, A. J., 2019b: Alterations in GRHL2-OVOL2-ZEB1 axis and aberrant activation of Wnt signaling lead to altered gene transcription in posterior polymorphous corneal dystrophy. *Experimental Eye Research.*, **188**, 107696.

Chung, V. Y.; Tan, T. Z.; Tan, M.; Wong, M. K.; Kuay, K. T.; Yang, Z.; Ye, J.; Muller, J.; Koh, C. M.; Guccione, E.; Thiery, J. P.; Huang, R. Y. J., 2016: GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Scientific Reports.*, **6**, 19943.

Clark, A. G.; Eisen, M. B.; Smith, D. R.; Bergman, C. M.; Oliver, B.; Markow, T. A.; Kaufman, T. C.; Kellis, M.; Gelbart, W.; Iyer, V. N.; Pollard, D. A.; Sackton, T. B.; Larracuente, A. M.; Singh, N. D.; Abad, J. P.; Abt, D. N.; Adryan, B.; Aguade, M.; Akashi, H. et al., 2007: Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.*, **450**, 203–218.

Cobb, M., 2017: 60 years ago, Francis Crick changed the logic of biology. *PLOS Biology.*, **15**, e2003243.

Cohen, E.; Sawyer, J. K.; Peterson, N. G.; Dow, J. A. T.; Fox, D. T., 2020: Physiology, Development, and Disease Modeling in the *Drosophila* Excretory System. *Genetics.*, **214**, 235–264.

Colombani, J.; Bianchini, L.; Layalle, S.; Pondeville, E.; Dauphin-Villemant, C.; Antoniewski, C.; Carré, C.; Noselli, S.; Léopold, P., 2005: Antagonistic Actions of Ecdysone and Insulins Determine Final Size in *Drosophila*. *Science*, **310**, 667–670.

Colombani, J.; Andersen, D. S., 2020: The *Drosophila* gut: A gatekeeper and coordinator of organism fitness and physiology. *Wiley Interdisciplinary Reviews: Developmental Biology*, **9**.

Craig Venter, J.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L. et al., 2001: The sequence of the human genome. *Science*, **291**, 1304–1351.

Crick, F., 1970: Central Dogma of Molecular Biology. *Nature*, **227**, 561–563.

Crocker, J.; Abe, N.; Rinaldi, L.; McGregor, A. P. P.; Frankel, N.; Wang, S.; Alsawadi, A.; Valenti, P.; Plaza, S.; Payre, F.; Mann, R. S. S.; Stern, D. L. L., 2015: Low affinity binding site clusters confer HOX specificity and regulatory robustness. *Cell*, **160**, 191–203.

Cross, S. D.; Johnson, A. A.; Gilles, B. J.; Bachman, L. A.; Inoue, T.; Agata, K.; Marmorstein, L. Y.; Marmorstein, A. D., 2015a: Control of maintenance and regeneration of planarian eyes by *ovo*. *Investigative Ophthalmology and Visual Science*, **56**, 7604–7610.

Cross, S. D.; Johnson, A. A.; Gilles, B. J.; Bachman, L. A.; Inoue, T.; Agata, K.; Marmorstein, L. Y.; Marmorstein, A. D., 2015b: Control of Maintenance and Regeneration of Planarian Eyes by *ovo*. *Investigative Ophthalmology & Visual Science*, **56**, 7604–7610.

Cubeñas-Potts, C.; Rowley, M. J.; Lyu, X.; Li, G.; Lei, E. P.; Corces, V. G., 2017: Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic acids research*, **45**, 1714–1730.

Dai, X.; Schonbaum, C.; Degenstein, L.; Bai, W.; Mahowald, A.; Fuchs, E., 1998: The *ovo* gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. *Genes & Development*, **12**, 3452–3463.

Davidson, A. E. E. E.; Liskova, P.; Evans, C. J. J. J.; Dudakova, L.; Nosková, L.; Pontikos, N.; Hartmannová, H.; Hodaňová, K.; Stránecký, V.; Kozmík, Z.; Levis, H. J. J. J.; Idigo, N.; Sasai, N.; Maher, G. J. J. J.; Bellingham, J.; Veli, N.; Ebenezer, N. D. D. D.; Cheetham, M. E. E. E.; Daniels, J. T. T. T. et al., 2016: Autosomal-Dominant Corneal Endothelial Dystrophies CHED1 and PPCD1 Are Allelic Disorders Caused by Non-coding Mutations in the Promoter of OVOL2. *American Journal of Human Genetics*, **98**, 75–89.

Dayhoff, M.; Ledley, R. S.; Eck, R. v., 1965: *Atlas of Protein Sequence and Structure*.

de Siqueira Santos, S.; Takahashi, D. Y.; Nakata, A.; Fujita, A., 2014: A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in Bioinformatics*, **15**, 906–918.

Delandre, C.; Marshall, O. J., 2019: *United colours of chromatin? Developmental genome organisation in flies*. *Biochemical Society Transactions*, Vol. 47pp. 691–700.

Delanoue, R.; Slaidina, M.; Léopold, P., 2010: The Steroid Hormone Ecdysone Controls Systemic Growth by Repressing dMyc Function in *Drosophila* Fat Cells. *Developmental Cell*, **18**, 1012–1021.

Delon, I.; Chanut-Delalande, H.; Payre, F., 2003: The *Ovo/Shavenbaby* transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*. *Mechanisms of Development*, **120**, 747–758.

Delon, I.; Payre, F., 2004, July 1: Evolution of larval morphology in flies: Get in shape with shavenbaby. *Trends in Genetics*.

Deng, W.; Lee, J.; Wang, H.; Miller, J.; Reik, A.; Gregory, P. D.; Dean, A.; Blobel, G. A., 2012: Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell.*, **149**, 1233–1244.

Deng, Z.; Cangkrama, M.; Butt, T.; Jane, S. M.; Carpinelli, M. R., 2021: Grainyhead-like transcription factors: guardians of the skin barrier. *Veterinary dermatology*.

Denholm, B.; Sudarsan, V.; Pasalodos-Sanchez, S.; Artero, R.; Lawrence, P.; Maddrell, S.; Baylies, M.; Skaer, H., 2003: Dual Origin of the Renal Tubules in *Drosophila*: Mesodermal Cells Integrate and Polarize to Establish Secretory Function. *Current Biology.*, **13**, 1052–1057.

Deuring, R.; Fanti, L.; Armstrong, J. A.; Sarte, M.; Papoulas, O.; Prestel, M.; Daubresse, G.; Verardo, M.; Moseley, S. L.; Berloco, M.; Tsukiyama, T.; Wu, C.; Pimpinelli, S.; Tamkun, J. W., 2000: The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Molecular cell.*, **5**, 355–365.

Dominado, N.; Casagrande, F.; Siddall, N. A.; Abud, H. E.; Hime, G. R., 2020: Low levels of Grainy head isoforms regulate *Drosophila* midgut intestinal stem cell differentiation. *bioRxiv.*, 2020.12.20.423699.

Dorsett, D., 2009, February: Cohesin, gene expression and development: Lessons from *Drosophila*. *Chromosome Research*.

Dorsett, D., 2019: The Many Roles of Cohesin in *Drosophila* Gene Transcription. *Trends in Genetics.*, **35**, 542–551.

dos Santos, G.; Schroeder, A. J.; Goodman, J. L.; Strelets, V. B.; Crosby, M. A.; Thurmond, J.; Emmert, D. B.; Gelbart, W. M.; Brown, N. H.; Kaufman, T.; Werner-Washburne, M.; Cripps, R.; Broll, K.; Gramates, L. S.; Falls, K.; Matthews, B. B.; Russo, S.; Zhou, P.; Zytkevich, M. et al., 2015: FlyBase: Introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Research.*, **43**, D690–D697.

Dow, J. A. T.; Krause, S. A.; Herzyk, P., 2021: Updates on ion and water transport by the Malpighian tubule. *Current Opinion in Insect Science.*, **47**, 31–37.

Dubrovsky, E. B., 2005: Hormonal cross talk in insect development. *Trends in endocrinology and metabolism: TEM.*, **16**, 6–11.

Dufourt, J.; Trullo, A.; Hunter, J.; Fernandez, C.; Lazaro, J.; Dejean, M.; Morales, L.; Nait-Amer, S.; Schulz, K. N.; Harrison, M. M.; Favard, C.; Radulescu, O.; Lagha, M., 2018: Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nature Communications.*, **9**.

Ebert, A.; Lein, S.; Schotta, G.; Reuter, G., 2006: Histone modification and the control of heterochromatic gene silencing in *Drosophila*. *Chromosome Research.*, **14**, 377–392.

Eden, E.; Navon, R.; Steinfeld, I.; Lipson, D.; Yakhini, Z., 2009: GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics.*, **10**, 48.

Entrevan, M.; Schuettengruber, B.; Cavalli, G., 2016: Regulation of Genome Architecture and Function by Polycomb Proteins. *Trends in Cell Biology.*, **26**, 511–525.

Espinola, S. M.; Götz, M.; Bellec, M.; Messina, O.; Fiche, J. B.; Houbon, C.; Dejean, M.; Reim, I.; Cardozo Gizzi, A. M.; Lagha, M.; Nollmann, M., 2021: Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early *Drosophila* development. *Nature Genetics.*, **53**, 477–486.

Feng, J.; Liu, T.; Zhang, Y., 2011: Using MACS to identify peaks from ChIP-Seq data. *Current protocols in bioinformatics.*, **Chapter 2**, Unit 2.14.

Fernandes, I.; Chanut-Delalande, H.; Ferrer, P.; Latapie, Y.; Waltzer, L.; Affolter, M.; Payre, F.; Plaza, S., 2010: Zona Pellucida Domain Proteins Remodel the Apical Compartment for Localized Cell Shape Changes. *Developmental Cell.*, **18**, 64–76.

Field, D.; Garrity, G.; Gray, T.; Morrison, N.; Selengut, J.; Sterk, P.; Tatusova, T.; Thomson, N.; Allen, M. J.; Angiuoli, S. v.; Ashburner, M.; Axelrod, N.; Baldauf, S.; Ballard, S.; Boore, J.; Cochrane, G.; Cole, J.; Dawyndt, P.; Vos, P. de et al., 2008: The minimum information about a genome sequence (MIGS) specification. *Nature Biotechnology 2008 26:5.*, **26**, 541–547.

Field, D.; Amaral-Zettler, L.; Cochrane, G.; Cole, J. R.; Dawyndt, P.; Garrity, G. M.; Gilbert, J.; Glöckner, F. O.; Hirschman, L.; Karsch-Mizrachi, I.; Klenk, H. P.; Knight, R.; Kottmann, R.; Kyrpides, N.; Meyer, F.; Gil, I. S.; Sansone, S. A.; Schriml, L. M.; Sterk, P. et al., 2011: The Genomic Standards Consortium. *PLOS Biology.*, **9**, e1001088.

Filion, G., 2011: Les cinq couleurs de la chromatine de drosophile. *médecine/sciences.*, **27**, 246–248.

Filion, G. J.; van Bommel, J. G.; Braunschweig, U.; Talhout, W.; Kind, J.; Ward, L. D.; Brugman, W.; de Castro, I. J.; Kerkhoven, R. M.; Bussemaker, H. J.; van Steensel, B., 2010: Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell.*, **143**, 212–224.

Fischer, T.; Cui, B.; Dhakshnamoorthy, J.; Zhou, M.; Rubin, C.; Zofall, M.; Veenstra, T. D.; Grewal, S. I. S., 2009: Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast. *Proceedings of the National Academy of Sciences.*, **106**, 8998–9003.

Fleischmann, R.; Adams, M.; White, O.; Clayton, R.; Kirkness, E.; Kerlavage, A.; Bult, C.; Tomb, J.; Dougherty, B.; Merrick, J.; al., et, 1995: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science.*, **269**, 496–512.

Francou, A.; Anderson, K. v., 2020: The Epithelial-to-Mesenchymal Transition in Development and Cancer. *Annual Review of Cancer Biology.*, **4**, 197–220.

Frankel, N.; Davis, G. K. G. K.; Vargas, D.; Wang, S.; Payre, F.; Stern, D. L. D. L., 2010: Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature.*, **466**, 490–493.

Frankel, N.; Wang, S.; Stern, D. L.; N, F.; S, W.; DL, S., 2012: Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America.*, **109**, 20975–20979.

Fujioka, M.; Emi-Sarker, Y.; Yusibova, G. L.; Goto, T.; Jaynes, J. B., 1999: Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development (Cambridge, England).*, **126**, 2527.

Furlong, E. E. M.; Levine, M., 2018, September 28: Developmental enhancers and chromosome topology. *Science.*

Galindo, M. I.; Pueyo, J. I.; Fouix, S.; Bishop, S. A.; Couso, J. P., 2007a: Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS biology.*, **5**, e106.

Galindo, M. I.; Pueyo, J. I.; Fouix, S.; Bishop, S. A.; Couso, J. P., 2007b: Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS biology.*, **5**, e106.

Gallant, P.; Shiiio, Y.; Cheng, P. F.; Parkhurst, S. M.; Eisenman, R. N., 1996: Myc and Max homologs in Drosophila. *Science (New York, N.Y.).*, **274**, 1523–1527.

- Garfinkel, M. D.; Lohe, A. R.; Mahowald, A. P., 1992: Molecular genetics of the *Drosophila melanogaster* ovo locus, a gene required for sex determination of germline cells. *Genetics.*, **130**, 791–803.
- Gaskill, M. M.; Gibson, T. J.; Larson, E. D.; Harrison, M. M., 2021: GAF is essential for zygotic genome activation and chromatin accessibility in the early *Drosophila* embryo. *eLife.*, **10**.
- Gauthier, J.; Vincent, A. T.; Charette, S. J.; Derome, N., 2019: A brief history of bioinformatics. *Briefings in Bioinformatics.*, **20**, 1981–1996.
- Ghavi-Helm, Y.; Klein, F. A.; Pakozdi, T.; Ciglar, L.; Noordermeer, D.; Huber, W.; Furlong, E. E. M., 2014: Enhancer loops appear stable during development and are associated with paused polymerase. *Nature.*, **512**, 96–100.
- Ghavi-Helm, Y.; Jankowski, A.; Meiers, S.; Viales, R. R.; Korb, J. O.; Furlong, E. E. M., 2019: Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature Genetics.*, **51**, 1272–1282.
- GLENN, T. C., 2011: Field guide to next-generation DNA sequencers. *Molecular Ecology Resources.*, **11**, 759–769.
- Gómez-Díaz, E.; Corces, V. G., 2014a: Architectural proteins: Regulators of 3D genome organization in cell fate. *Trends in Cell Biology.*, **24**, 703–711.
- Gómez-Díaz, E.; Corces, V. G., 2014b: Architectural proteins: Regulators of 3D genome organization in cell fate. *Trends in Cell Biology.*
- Guerra-Almeida, D.; Tschoeke, D. A.; Da-Fonseca, R. N., 2021: Understanding small ORF diversity through a comprehensive transcription feature classification. *DNA Research.*
- Jones, P. A.; Issa, J. P. J.; Baylin, S., 2016a, October 1: Targeting the cancer epigenome for therapy. *Nature Reviews Genetics.*
- Guo, M.; Yan, W., 2015b: Epigenetics of Gastric Cancer. *Methods in Molecular Biology.*, Vol. 1238pp. 783–799.
- Baylin, S. B.; Jones, P. A., 2016c: Epigenetic determinants of cancer. *Cold Spring Harbor Perspectives in Biology.*, **8**, 783–799.
- Guo, Z.; Ohlstein, B., 2015d: Bidirectional Notch signaling regulates *Drosophila* intestinal stem cell multipotency. *Science.*, **350**.
- Guo, Z.; Lucchetta, E.; Rafel, N.; Ohlstein, B., 2016a: Maintenance of the adult *Drosophila* intestine: all roads lead to homeostasis. *Current Opinion in Genetics & Development.*, **40**, 81–86.
- Guo, Z.; Lucchetta, E.; Rafel, N.; Ohlstein, B., 2016b: Maintenance of the adult *Drosophila* intestine: all roads lead to homeostasis. *Current Opinion in Genetics & Development.*, **40**, 81–86.
- Gupta, P. B.; Pastushenko, I.; Skibinski, A.; Blanpain, C.; Kuperwasser, C., 2019: Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell.*, **24**, 65–78.
- Gutierrez-Perez, I.; Rowley, M. J.; Lyu, X.; Valadez-Graham, V.; Vallejo, D. M.; Ballesta-Illan, E.; Lopez-Atalaya, J. P.; Kremsky, I.; Caparros, E.; Corces, V. G.; Dominguez, M., 2019: Ecdysone-Induced 3D Chromatin Reorganization Involves Active Enhancers Bound by Pipsqueak and Polycomb. *Cell Reports.*
- Hagen, J. B., 2000: The origins of bioinformatics. *Nature Reviews Genetics 2000 1:3.*, **1**, 231–236.

Hammonds, A. S.; Bristow, C. A.; Fisher, W. W.; Weiszmann, R.; Wu, S.; Hartenstein, V.; Kellis, M.; Yu, B.; Frise, E.; Celniker, S. E., 2013: Spatial expression of transcription factors in *Drosophila* embryonic organ development. *Genome Biology.*, **14**.

Hayashi, M.; Shinozuka, Y.; Shigenobu, S.; Sato, M.; Sugimoto, M.; Ito, S.; Abe, K.; Kobayashi, S., 2017: Conserved role of *Ovo* in germline development in mouse and *Drosophila*. *Scientific Reports.*, **7**, 1–10.

Hergeth, S. P.; Schneider, R., 2015: The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO reports.*, **16**, 1439–1453.

Hersh, R. T.; Eck, R. v.; Dayhoff, M. O., 1967: Atlas of Protein Sequence and Structure, 1966. *Systematic Zoology.*, **16**, 262.

Hesper, B.; Hogeweg, P., 1970: Bioinformatica: een werkconcept. *Kameleon.*, **1**, 28–29.

Hillyar, C.; Rallis, K. S.; Varghese, J., 2020: Advances in Epigenetic Cancer Therapeutics. *Cureus.*

Hino, S.; Kohrogi, K.; Nakao, M., 2016: Histone demethylase LSD1 controls the phenotypic plasticity of cancer cells. *Cancer Science.*

Hogeweg, P., 2011: The Roots of Bioinformatics in Theoretical Biology. *PLoS Computational Biology.*, **7**.

Hong, T.; Watanabe, K.; Ta, C. H.; Villarreal-Ponce, A.; Nie, Q.; Dai, X., 2015: An *Ovol2-Zeb1* Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. *PLoS Computational Biology.*, **11**.

Hoskins, R. A.; Smith, C. D.; Carlson, J. W.; Carvalho, A. B.; Halpern, A.; Kaminker, J. S.; Kennedy, C.; Mungall, C. J.; Sullivan, B. A.; Sutton, G. G.; Yasuhara, J. C.; Wakimoto, B. T.; Myers, E. W.; Celniker, S. E.; Rubin, G. M.; Karpen, G. H., 2002: Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome biology.*, **3**, RESEARCH0085.

Hoskins, R. A.; Carlson, J. W.; Wan, K. H.; Park, S.; Mendez, I.; Galle, S. E.; Booth, B. W.; Pfeiffer, B. D.; George, R. A.; Svirskas, R.; Krzywinski, M.; Schein, J.; Accardo, M. C.; Damia, E.; Messina, G.; Méndez-Lago, M.; de Pablos, B.; Demakova, O. v.; Andreyeva, E. N. et al., 2015: The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Research.*, **25**, 445–458.

Hou, C.; Li, L.; Qin, Z. S.; Corces, V. G., 2012: Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Molecular cell.*, **48**, 471–484.

Huang, D. W.; Sherman, B. T.; Lempicki, R. A., 2009: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols.*, **4**, 44–57.

Hudry, B.; Khadayate, S.; Miguel-Aliaga, I., 2016: The sexual identity of adult intestinal stem cells controls organ size and plasticity. *Nature 2016 530:7590.*, **530**, 344–348.

Hung, R. J.; Hu, Y.; Kirchner, R.; Liu, Y.; Xu, C.; Comjean, A.; Tattikota, S. G.; Li, F.; Song, W.; Sui, S. H.; Perrimon, N., 2020: A cell atlas of the adult *Drosophila* midgut. *Proceedings of the National Academy of Sciences of the United States of America.*, **117**, 1514–1523.

Husmann, D.; Gozani, O., 2019: Histone lysine methyltransferases in biology and disease. *Nature Structural & Molecular Biology.*, **26**, 880–889.

Hyun, K.; Jeon, J.; Park, K.; Kim, J., 2017: Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine.*, **49**, e324–e324.

Ingham, P. W., 1988: The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature 1988 335:6185.*, **335**, 25–34.

Ito, T.; Tsuji, G.; Ohno, F.; Nakahara, T.; Uchi, H.; Furue, M., 2017: Potential role of the OVOL1–OVOL2 axis and c-Myc in the progression of cutaneous squamous cell carcinoma. *Modern Pathology*, **30**, 919–927.

Jacobs, J.; Atkins, M.; Davie, K.; Imrichova, H.; Romanelli, L.; Christiaens, V.; Hulselmans, G.; Potier, D.; Wouters, J.; Taskiran, I. I.; Paciello, G.; González-Blas, C. B.; Koldere, D.; Aibar, S.; Halder, G.; Aerts, S., 2018: The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nature Genetics*, **50**, 1011–1020.

Jia, Y.; Jin, S.; Hu, K.; Geng, L.; Han, C.; Kang, R.; Pang, Y.; Ling, E.; Tan, E. K.; Pan, Y.; Liu, W., 2021: Gut microbiome modulates *Drosophila* aggression through octopamine signaling. *Nature Communications*, **12**, 2698.

Jiang, H.; Edgar, B. A., 2012: Intestinal stem cell function in *Drosophila* and mice. *Current opinion in genetics & development*, **22**, 354–360.

Jiang, S.; Mortazavi, A., 2018a: Integrating ChIP-seq with other functional genomics data. *Briefings in Functional Genomics*, **17**, 104–115.

Jiang, S.; Mortazavi, A., 2018b: Integrating ChIP-seq with other functional genomics data. *Briefings in Functional Genomics*, **17**, 104–115.

Johnson, A. D.; Fitzsimmons, D.; Hagman, J.; Chamberlin, H. M., 2001: EGL-38 Pax regulates the ovo-related gene *lin-48* during *Caenorhabditis elegans* organ development. *Development (Cambridge, England)*, **128**, 2857–2865.

Johnson, S. C., 1967: Hierarchical clustering schemes. *Psychometrika*, **32**, 241–254.

Jolly, M. K.; Ware, K. E.; Gilja, S.; Somarelli, J. A.; Levine, H., 2017, July 12: EMT and MET: necessary or permissive for metastasis? *Molecular Oncology*.

Kharchenko, P. v.; Alekseyenko, A. A.; Schwartz, Y. B.; Minoda, A.; Riddle, N. C.; Ernst, J.; Sabo, P. J.; Larschan, E.; Gorchakov, A. A.; Gu, T.; Linder-Basso, D.; Plachetka, A.; Shanower, G.; Tolstorukov, M. Y.; Luquette, L. J.; Xi, R.; Jung, Y. L.; Park, R. W.; Bishop, E. P. et al., 2011a: Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*, **471**, 480–485.

Kharchenko, P. v.; Alekseyenko, A. A.; Schwartz, Y. B.; Minoda, A.; Riddle, N. C.; Ernst, J.; Sabo, P. J.; Larschan, E.; Gorchakov, A. A.; Gu, T.; Linder-Basso, D.; Plachetka, A.; Shanower, G.; Tolstorukov, M. Y.; Luquette, L. J.; Xi, R.; Jung, Y. L.; Park, R. W.; Bishop, E. P. et al., 2011b: Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*, **471**, 480–486.

Kitazawa, K.; Hikichi, T.; Nakamura, T.; Mitsunaga, K.; Tanaka, A.; Nakamura, M.; Yamakawa, T.; Furukawa, S.; Takasaka, M.; Goshima, N.; Watanabe, A.; Okita, K.; Kawasaki, S.; Ueno, M.; Kinoshita, S.; Masui, S., 2016: OVOL2 Maintains the Transcriptional Program of Human Corneal Epithelium by Suppressing Epithelial-to-Mesenchymal Transition. *Cell Reports*, **15**, 1359–1368.

Kittelman, S.; Preger-Ben Noon, E.; McGregor, A. P.; Frankel, N., 2021: *A complex gene regulatory architecture underlies the development and evolution of cuticle morphology in Drosophila*, Vol. 69pp. 21–27.

Klose, R. J.; Kallin, E. M.; Zhang, Y., 2006: JmjC-domain-containing proteins and histone demethylation. *Nature Reviews Genetics*, **7**, 715–727.

Koenecke, N.; Johnston, J.; He, Q.; Meier, S.; Zeitlinger, J., 2017: *Drosophila* poised enhancers are generated during tissue patterning with the help of repression. *Genome Research*, **27**, 64–74.

Kondo, T.; Hashimoto, Y.; Kato, K.; Inagaki, S.; Hayashi, S.; Kageyama, Y., 2007a: Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nature Cell Biology.*, **9**, 660–665.

Kondo, T.; Hashimoto, Y.; Kato, K.; Inagaki, S.; Hayashi, S.; Kageyama, Y., 2007b: Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nature Cell Biology.*, **9**, 660–665.

Kondo, T.; Plaza, S.; Zanet, J.; Benrabah, E.; Valenti, P.; Hashimoto, Y.; Kobayashi, S.; Payre, F.; Kageyama, Y., 2010: Small peptides switch the transcriptional activity of shavenbaby during drosophila embryogenesis. *Science.*, **329**, 336–339.

Korzelius, J.; Naumann, S. K.; Loza-Coll, M. A.; Chan, J. S.; Dutta, D.; Oberheim, J.; Gläßer, C.; Southall, T. D.; Brand, A. H.; Jones, D. L.; Edgar, B. A., 2014: Escargot maintains stemness and suppresses differentiation in Drosophila intestinal stem cells. *The EMBO Journal.*, **33**, 2967–2982.

Kouzarides, T., 2007, February 23: Chromatin Modifications and Their Function. *Cell*.

Kumar, A.; Bhandari, A.; Sinha, R.; Sardar, P.; Sushma, M.; Goyal, P.; Goswami, C.; Grapputo, A., 2012: Molecular phylogeny of OVOL genes illustrates a conserved C2H2 zinc finger domain coupled by hypervariable unstructured regions. *PloS one.*, **7**, e39399.

Kvon, E. Z.; Kazmar, T.; Stampfel, G.; Omar Yáñez-Cuna, J.; Pagani, M.; Schernhuber, K.; Dickson, B. J.; Stark, A., 2014: Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature.*, **512**, 91–95.

Kvon, E. Z.; Kamneva, O. K.; Melo, U. S.; Barozzi, I.; Osterwalder, M.; Mannion, B. J.; Tissières, V.; Pickle, C. S.; Plajzer-Frick, I.; Lee, E. A.; Kato, M.; Garvin, T. H.; Akiyama, J. A.; Afzal, V.; Lopez-Rios, J.; Rubin, E. M.; Dickel, D. E.; Pennacchio, L. A.; Visel, A., 2016: Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell.*, **167**, 633–642.e11.

Lagha, M.; Bothma, J. P.; Levine, M., 2012: Mechanisms of transcriptional precision in animal development. *Trends in Genetics.*, **28**, 409–416.

Lapan, S. W.; Reddien, P. W., 2012: Transcriptome analysis of the planarian eye identifies ovo as a specific regulator of eye regeneration. *Cell reports.*, **2**, 294–307.

Le manifeste GNU - Projet GNU - Free Software Foundation, (n.d.).

Lee, S.; Garfinkel, M. D., 2000: Characterization of Drosophila OVO protein DNA binding specificity using random DNA oligomer selection suggests zinc finger degeneration. *Nucleic acids research.*, **28**, 826–834.

Levine, M.; Davidson, E. H., 2005: Gene regulatory networks for development. *Proceedings of the National Academy of Sciences.*, **102**, 4936–4942.

Li, B.; Dai, Q.; Li, L.; Nair, M.; Mackay, D. R.; Dai, X., 2002: Ovol2, a mammalian homolog of Drosophila ovo: Gene structure, chromosomal mapping, and aberrant expression in blind-sterile mice. *Genomics.*, **80**, 319–325.

Li, B.; Nair, M.; Mackay, D. R.; Bilanchone, V.; Hu, M.; Fallahi, M.; Song, H.; Dai, Q.; Cohen, P. E.; Dai, X., 2005: Ovol1 regulates meiotic pachytene progression during spermatogenesis by repressing Id2 expression. *Development (Cambridge, England).*, **132**, 1463–1473.

Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R., 2009: The Sequence Alignment/Map format and SAMtools. *Bioinformatics.*, **25**, 2078–2079.

Li, H.; Janssens, J.; de Waegeneer, M.; Kolluru, S. S.; Davie, K.; Gardeux, V.; Saelens, W.; David, F.; Brbić, M.; Leskovec, J.; McLaughlin, C. N.; Xie, Q.; Jones, R. C.; Brueckner, K.; Shim,

J.; Tattikota, S. G.; Schnorrer, F.; Rust, K.; Nystul, T. G. et al., 2021: Fly Cell Atlas: a single-cell transcriptomic atlas of the adult fruit fly. *bioRxiv*.

Li, S.; Yang, J., 2014a, April 14: Ovol Proteins: Guardians against EMT during Epithelial Differentiation. *Developmental Cell*.

Li, Z.; Liu, S.; Cai, Y., 2014b: Differential Notch Activity Is Required for Homeostasis of Malpighian Tubules in Adult *Drosophila*. *Journal of Genetics and Genomics.*, **41**, 649–652.

Lorton, B. M.; Shechter, D., 2019: Cellular consequences of arginine methylation. *Cellular and Molecular Life Sciences.*, **76**, 2933–2956.

Love, M. I.; Huber, W.; Anders, S., 2014: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology.*, **15**, 550.

Lupiáñez, D. G.; Spielmann, M.; Mundlos, S., 2016: Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends in Genetics.*, **32**, 225–237.

Lutz, M., 1996: *Programming Python 1st edition*. O'Reilly.

Ma, S.; Meng, Z.; Chen, R.; Guan, K. L., 2019: The Hippo Pathway: Biology and Pathophysiology. <https://doi.org/10.1146/annurev-biochem-013118-111829>., **88**, 577–604.

Mace, K. A.; Pearson, J. C.; McGinnis, W., 2005: An epidermal barrier wound repair pathway in *Drosophila* is mediated by grainy head. *Science (New York, N.Y.)*., **308**, 381–385.

Mackay, D. R.; Hu, M.; Li, B.; Rhéaume, C.; Dai, X., 2006: The mouse Ovol2 gene is required for cranial neural tube development. *Developmental biology.*, **291**, 38–52.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations.

Malita, A.; Rewitz, K., 2021: Interorgan communication in the control of metamorphosis. *Current Opinion in Insect Science.*, **43**, 54–62.

Mannervik, M., 2013: Control of *Drosophila* embryo patterning by transcriptional co-regulators. *Experimental Cell Research.*, **321**, 47–57.

Markstein, M.; Markstein, P.; Markstein, V.; Levine, M. S., 2002: Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proceedings of the National Academy of Sciences of the United States of America.*, **99**, 763–768.

Marshall, O. J.; Brand, A. H., 2017: Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. *Nature Communications.*, **8**, 2271.

McGregor, A. P.; Orgogozo, V.; Delon, I.; Zanet, J.; Srinivasan, D. G.; Payre, F.; Stern, D. L., 2007: Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature.*, **448**, 587–590.

Menoret, D.; Santolini, M.; Fernandes, I.; Spokony, R.; Zanet, J.; Gonzalez, I.; Latapie, Y.; Ferrer, P.; Rouault, H.; White, K. P.; Besse, P.; Hakim, V.; Aerts, S.; Payre, F.; Plaza, S., 2013: Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome biology.*, **14**, R86.

Mével-Ninio, M.; Terracol, R.; Kafatos, F. C., 1991: The ovo gene of *Drosophila* encodes a zinc finger protein required for female germ line development. *The EMBO journal.*, **10**, 2259–2266.

Mével-Ninio, M.; Terracol, R.; Salles, C.; Vincent, A.; Payre, F., 1995: ovo, a *Drosophila* gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning. *Mechanisms of Development.*, **49**, 83–95.

Micchelli, C. A.; Perrimon, N., 2006a: Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature.*, **439**, 475–479.

Micchelli, C. A.; Perrimon, N., 2006b: Evidence that stem cells reside in the adult *Drosophila* midgut epithelium.

Micchelli, C. A.; Sudmeier, L.; Perrimon, N.; Tang, S.; Beehler-Evans, R., 2011: Identification of adult midgut precursors in *Drosophila*. *Gene Expression Patterns.*, **11**, 12–21.

Mirzoyan, Z.; Sollazzo, M.; Allocca, M.; Valenza, A. M.; Grifoni, D.; Bellosta, P., 2019: *Drosophila melanogaster*: A Model Organism to Study Cancer. *Frontiers in Genetics.*, **0**, 51.

Morcillo, P.; Rosen, C.; Baylies, M. K.; Dorsett, D., 1997: Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in *Drosophila*. *Genes and Development.*, **11**, 2729–2740.

Moreau, P.; Hen, R.; Wasylyk, B.; Everett, R.; Gaub, M. P.; Chambon, P., 1981: The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Research.*, **9**, 6047–6068.

Muerdter, F.; Boryń, Ł. M.; Arnold, C. D., 2015, September 1: STARR-seq - Principles and applications. *Genomics*.

Myers, E. W.; Sutton, G. G.; Delcher, A. L.; Dew, I. M.; Fasulo, D. P.; Flanigan, M. J.; Kravitz, S. A.; Mobarry, C. M.; Reinert, K. H. J.; Remington, K. A.; Anson, E. L.; Bolanos, R. A.; Chou, H. H.; Jordan, C. M.; Halpern, A. L.; Lonardi, S.; Beasley, E. M.; Brandon, R. C.; Chen, L. et al., 2000, March 24: A whole-genome assembly of *Drosophila*. *Science*.

Nair, M.; Bilanchone, V.; Ortt, K.; Sinha, S.; Dai, X., 2007: *Ovol1* represses its own transcription by competing with transcription activator c-Myb and by recruiting histone deacetylase activity. *Nucleic acids research.*, **35**, 1687–1697.

Nieto, M. A.; Huang, R. Y. Y. J.; Jackson, R. A. A.; Thiery, J. P. P., 2016: EMT: 2016. *Cell*.

Noyes, M. B.; Christensen, R. G.; Wakabayashi, A.; Stormo, G. D.; Brodsky, M. H.; Wolfe, S. A., 2008: Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Cell.*, **133**, 1277–1289.

Nüsslein-Volhard, C.; Wieschaus, E., 1980: Mutations affecting segment number and polarity in *Drosophila*. *Nature 1980 287:5785.*, **287**, 795–801.

Ogiyama, Y.; Schuettengruber, B.; Papadopoulos, G. L.; Chang, J. M.; Cavalli, G., 2018: Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during *Drosophila* Development. *Molecular Cell.*, **71**, 73-88.e5.

Oh, H.; Slattery, M.; Ma, L.; Crofts, A.; White, K. P.; Mann, R. S.; Irvine, K. D., 2013: Genome-wide Association of Yorkie with Chromatin and Chromatin-Remodeling Complexes. *Cell Reports.*, **3**, 309–318.

Ohlstein, B.; Spradling, A., 2006a: The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature.*, **439**, 470–474.

Ohlstein, B.; Spradling, A., 2006b: The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells.

Olins, A. L.; Olins, D. E., 1974: Spheroid Chromatin Units (ngr Bodies). *Science.*, **183**, 330–332.

Oliver, B.; Perrimon, N.; Mahowald, A. P., 1987: The *ovo* locus is required for sex-specific germ line maintenance in *Drosophila*. *Genes & development.*, **1**, 913–923.

Oliver, B.; Pauli, D.; Mahowald, A. P., 1990: Genetic evidence that the *ovo* locus is involved in *Drosophila* germ line sex determination. *Genetics.*, **125**, 535–550.

Oliver, B.; Singer, J.; Laget, V.; Pennetta, G.; Pauli, D., 1994: Function of *Drosophila ovo+* in germ-line sex determination depends on X-chromosome number. *Development.*, **120**, 3185–3195.

- Ong, C. T.; Corces, V. G., 2014: CTCF: an architectural protein bridging genome topology and function. *Nature reviews. Genetics.*, **15**, 234–246.
- Pan, X.; Connacher, R. P.; O'Connor, M. B., 2021: Control of the insect metamorphic transition by ecdysteroid production and secretion. *Current Opinion in Insect Science.*, **43**, 11–20.
- Papoulas, O.; Beek, S. J.; Moseley, S. L.; McCallum, C. M.; Sarte, M.; Shearn, A.; Tamkun, J. W., 1998: The *Drosophila* trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. *Development.*, **125**, 3955–3966.
- Pauli, D.; Mahowald, A. P., 1990: Germ-line sex determination in *Drosophila melanogaster*. *Trends in genetics : TIG.*, **6**, 259–264.
- Payre, F.; Vincent, A.; Carreno, S., 1999: ovo/svb integrates Wingless and DER pathways to control epidermis differentiation. *Nature.*, **400**, 271–275.
- Payre, F., 2004: Genetic control of epidermis differentiation in *Drosophila*. *The International journal of developmental biology.*, **48**, 207–215.
- Phillips-Cremins, J. E.; Corces, V. G., 2013, May 23: Chromatin Insulators: Linking Genome Organization to Cellular Function. *Molecular Cell*.
- Pichon, X.; Lagha, M.; Mueller, F.; Bertrand, E., 2018: A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges. *Molecular Cell.*, **71**, 468–480.
- Plaza, S.; Menschaert, G.; Payre, F., 2017: In Search of Lost Small Peptides. *Annual Review of Cell and Developmental Biology.*, **33**, 391–416.
- Preger-Ben Noon, E.; Sabarís, G.; Ortiz, D. M.; Sager, J.; Liebowitz, A.; Stern, D. L.; Frankel, N., 2018: Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function. *Cell Reports.*, **22**, 3021–3031.
- Puisieux, A.; Brabletz, T.; Caramel, J., 2014: Oncogenic roles of EMT-inducing transcription factors. *Nature Cell Biology*.
- Quinlan, A. R.; Hall, I. M., 2010: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.*, **26**, 841–842.
- Ramírez, F.; Ryan, D. P.; Grüning, B.; Bhardwaj, V.; Kilpert, F.; Richter, A. S.; Heyne, S.; Dündar, F.; Manke, T., 2016: deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research.*, **44**, W160–W165.
- Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T., 2000: Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature.*, **406**, 593–599.
- Regadas, I.; Dahlberg, O.; Vaid, R.; Ho, O.; Belikov, S.; Dixit, G.; Deindl, S.; Wen, J.; Mannervik, M., 2021: A unique histone 3 lysine 14 chromatin signature underlies tissue-specific gene regulation. *Molecular Cell.*, **81**, 1766-1780.e10.
- Rheault, M. R.; O'Donnell, M. J., 2004: Organic cation transport by Malpighian tubules of *Drosophila melanogaster* : application of two novel electrophysiological methods. *Journal of Experimental Biology.*, **207**, 2173–2184.
- LM, R., 1993: Hormone receptors and the regulation of insect metamorphosis. *Receptor.*, **3**, 203–209.
- Riddihough, G.; Ish-Horowicz, D., 1991: *Individual stripe regulatory elements in the Drosophila hairy promoter respond to maternal, gap, and pair-rule genes. Genes and Development.*, Vol. 5pp. 840–854.
- Robertson, A. G.; Bilenky, M.; Tam, A.; Zhao, Y.; Zeng, T.; Thiessen, N.; Cezard, T.; Fejes, A. P.; Wederell, E. D.; Cullum, R.; Euskirchen, G.; Krzywinski, M.; Birol, I.; Snyder, M.; Hoodless,

P. A.; Hirst, M.; Marra, M. A.; Jones, S. J. M., 2008: Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Research.*, **18**, 1906–1917.

Robinson, M. D.; McCarthy, D. J.; Smyth, G. K., 2010: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England).*, **26**, 139–140.

Roca, H.; Hernandez, J.; Weidner, S.; McEachin, R. C.; Fuller, D.; Sud, S.; Schumann, T.; Wilkinson, J. E.; Zaslavsky, A.; Li, H.; Maher, C. A.; Daignault-Newton, S.; Healy, P. N.; Pienta, K. J., 2013: Transcription Factors OVOL1 and OVOL2 Induce the Mesenchymal to Epithelial Transition in Human Cancer. *PLoS ONE.*, **8**.

Rodan, A. R., 2019: The Drosophila Malpighian tubule as a model for mammalian tubule function. *Current opinion in nephrology and hypertension.*, **28**, 455–464.

Rosa, S.; Shaw, P., 2013: Insights into chromatin structure and dynamics in plants. *Biology.*, **2**, 1378–1410.

Sanger, F.; Tuppy, H., 1951: The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochemical Journal.*, **49**, 463.

Sanger, F.; Nicklen, S.; Coulson, A. R., 1977a: DNA sequencing with chain-terminating. *Proc Natl Acad Sci USA.*, **74**, 5463–5467.

Sanger, F.; Air, G. M.; Barrell, B. G.; Brown, N. L.; Coulson, A. R.; Fiddes, J. C.; Hutchison, C. A.; Slocombe, P. M.; Smith, M., 1977b: Nucleotide sequence of bacteriophage ϕ x174 DNA. *Nature.*, **265**, 687–695.

Sanger, F., 1988: *Further ANNUAL REVIEWS SEQUENCES, SEQUENCES, AND SEQUENCES. Annual Review of Biochemistry.*, Vol. 57pp. 1–29.

Savard, J.; Marques-Souza, H.; Aranda, M.; Tautz, D., 2006: A Segmentation Gene in Tribolium Produces a Polycistronic mRNA that Codes for Multiple Conserved Peptides. *Cell.*, **126**, 559–569.

Saxena, K.; Srikrishnan, S.; Celia-Terrassa, T.; Jolly, M. K., 2020: OVOL1/2: Drivers of Epithelial Differentiation in Development, Disease, and Reprogramming. *Cells Tissues Organs.*

Schaaf, C. A.; Misulovin, Z.; Gause, M.; Koenig, A.; Gohara, D. W.; Watson, A.; Dorsett, D., 2013: Cohesin and Polycomb Proteins Functionally Interact to Control Transcription at Silenced and Active Genes. (Cavalli, G., Ed.) *PLoS Genetics.*, **9**, e1003560.

Schoenfelder, S.; Fraser, P., 2019: Long-range enhancer-promoter contacts in gene expression control. *Nature reviews. Genetics.*, **20**, 437–455.

Schubert, T.; Pusch, M. C.; Diermeier, S.; Benes, V.; Kremmer, E.; Imhof, A.; Längst, G., 2012: Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Molecular cell.*, **48**, 434–444.

Schuettengruber, B.; Chourrout, D.; Vervoort, M.; Leblanc, B.; Cavalli, G., 2007, February 23: Genome Regulation by Polycomb and Trithorax Proteins. *Cell.*

Schulz, K. N.; Bondra, E. R.; Moshe, A.; Villalta, J. E.; Lieb, J. D.; Kaplan, T.; McKay, D. J.; Harrison, M. M., 2015: Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Research.*, **25**, 1715–1726.

Schwartz, Y. B.; Cavalli, G., 2017: Three-dimensional genome organization and function in Drosophila. *Genetics.*, **205**, 5–24.

Sexton, T.; Yaffe, E.; Kenigsberg, E.; Bantignies, F.; Leblanc, B.; Hoichman, M.; Parrinello, H.; Tanay, A.; Cavalli, G., 2012: Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell.*, **148**, 458–472.

Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y., 2004: Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell.*, **119**, 941–953.

Shin, S. H.; Kim, D.; Hwang, J.; Kim, M. K.; Kim, J. C.; Sung, Y. K., 2014: OVO homolog-like 1, a target gene of the Wnt/ β -catenin pathway, controls hair follicle neogenesis. *The Journal of investigative dermatology.*, **134**, 838–840.

Singh, S. R.; Liu, W.; Hou, S. X., 2007a: The Adult Drosophila Malpighian Tubules Are Maintained by Multipotent Stem Cells. *Cell Stem Cell.*, **1**, 191–203.

Singh, S. R.; Liu, W.; Hou, S. X., 2007b: The Adult Drosophila Malpighian Tubules Are Maintained by Multipotent Stem Cells. *Cell Stem Cell.*, **1**, 191–203.

Singh, S. R.; Aggarwal, P.; Hou, S. X., 2019: Cancer Stem Cells and Stem Cell Tumors in Drosophila. In: Deng, W. M. (ed.), *The Drosophila Model in Cancer*. Springer International Publishing, Cham, pp. 175–190.

Slatko, B. E.; Gardner, A. F.; Ausubel, F. M., 2018: Overview of Next Generation Sequencing Technologies. *Current protocols in molecular biology.*, **122**, e59.

Small, S.; Arnosti, D. N., 2020: Transcriptional enhancers in Drosophila. *Genetics.*, **216**, 1–26.

Smith, B.; Bhowmick, N., 2016: Role of EMT in Metastasis and Therapy Resistance. *Journal of Clinical Medicine.*, **5**, 17.

Spearman, C., 1904: “General Intelligence,” Objectively Determined and Measured. *The American Journal of Psychology.*, **15**, 201.

Staden, R., 1996: The Staden Sequence Analysis Package. *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology*.

Stern, D. L.; Frankel, N., 2013: The structure and evolution of cis -regulatory regions: the shavenbaby story. *Philosophical Transactions of the Royal Society B: Biological Sciences.*, **368**, 20130028.

Sterner, D. E.; Berger, S. L., 2000: Acetylation of Histones and Transcription-Related Factors. *Microbiology and Molecular Biology Reviews.*, **64**, 435–459.

Stoiber, M.; Celniker, S.; Cherbas, L.; Brown, B.; Cherbas, P., 2016: Diverse Hormone Response Networks in 41 Independent Drosophila Cell Lines. *G3: Genes, Genomes, Genetics.*, **6**, 683–694.

Strasser, B. J., 2010: Collecting, Comparing, and Computing Sequences: The Making of Margaret O. Dayhoff’s Atlas of Protein Sequence and Structure, 1954-1965. *Journal of the History of Biology.*, **43**, 623–660.

Sucena, E.; Delon, I.; Jones, I.; Payre, F.; Stern, D. L., 2003: Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature.*, **424**, 935–938.

Szabo, Q.; Bantignies, F.; Cavalli, G., 2019: Principles of genome folding into topologically associating domains. *Science Advances.*, **5**, eaaw1668.

Tam, W. L.; Weinberg, R. A., 2013: The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nature Medicine.*, **19**, 1438–1449.

Tauc, H. M.; Rodriguez-Fernandez, I. A.; Hackney, J. A.; Pawlak, M.; Ronnen Oron, T.; Korzelius, J.; Moussa, H. F.; Chaudhuri, S.; Modrusan, Z.; Edgar, B. A.; Jasper, H., 2021: Age-

related changes in polycomb gene regulation disrupt lineage fidelity in intestinal stem cells. *eLife.*, **10**.

Ting, S. B.; Caddy, J.; Hislop, N.; Wilanowski, T.; Auden, A.; Zhao, L. L.; Ellis, S.; Kaur, P.; Uchida, Y.; Holleran, W. M.; Elias, P. M.; Cunningham, J. M.; Jane, S. M., 2005: A homolog of *Drosophila* grainy head is essential for epidermal integrity in mice. *Science (New York, N.Y.)*, **308**, 411–413.

Tsuji, G.; Ito, T.; Chiba, T.; Mitoma, C.; Nakahara, T.; Uchi, H.; Furue, M., 2018: The role of the OVOL1–OVOL2 axis in normal and diseased human skin. *Journal of Dermatological Science.*, **90**, 227–231.

Turner, B. M.; Birley, A. J.; Lavender, J., 1992: Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell.*, **69**, 375–384.

Ugur, B.; Chen, K.; Bellen, H. J., 2016: *Drosophila* tools and assays for the study of human diseases. *Disease Models & Mechanisms.*, **9**, 235–244.

Umlauf, D., 2015: Le génome intime... et en trois dimensions. *médecine/sciences.*, **31**, 304–311.

Unezaki, S.; Nishizawa, M.; Okuda-Ashitaka, E.; Masu, Y.; Mukai, M.; Kobayashi, S.; Sawamoto, K.; Okano, H.; Ito, S., 2004a: Characterization of the isoforms of MOVO zinc finger protein, a mouse homologue of *Drosophila* Ovo, as transcription factors. *Gene.*, **336**, 47–58.

Unezaki, S.; Nishizawa, M.; Okuda-Ashitaka, E.; Masu, Y.; Mukai, M.; Kobayashi, S.; Sawamoto, K.; Okano, H.; Ito, S., 2004b: Characterization of the isoforms of MOVO zinc finger protein, a mouse homologue of *Drosophila* Ovo, as transcription factors. *Gene.*, **336**, 47–58.

van Bortle, K.; Corces, V. G., 2013: The role of chromatin insulators in nuclear architecture and genome function. *Current Opinion in Genetics & Development.*, **23**, 212–218.

Verza, F. A.; Das, U.; Fachin, A. L.; Dimmock, J. R.; Marins, M., 2020: Roles of Histone Deacetylases and Inhibitors in Anticancer Therapy. *Cancers.*, **12**, 1664.

Wang, C.; Spradling, A. C., 2020a: An abundant quiescent stem cell population in *Drosophila* Malpighian tubules protects principal cells from kidney stones. *eLife.*, **9**.

Wang, C.; Spradling, A. C., 2020b: An abundant quiescent stem cell population in *Drosophila* Malpighian tubules protects principal cells from kidney stones. *eLife.*, **9**.

Wang, Z. H.; Li, Z.; Hu, M.; Yang, Q. J.; Yan, S.; Wu, R. S.; Li, B. A.; Guo, M., 2017: *Ovol2* gene inhibits the Epithelial-to-Mesenchymal Transition in lung adenocarcinoma by transcriptionally repressing *Twist1*. *Gene.*, **600**, 1–8.

Ward, J. H., 1963: Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association.*, **58**, 236–244.

WATSON, J. D.; CRICK, F. H. C., 1953: Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature 1953 171:4356.*, **171**, 737–738.

Wieschaus, E.; Nüsslein-Volhard, C.; Jürgens, G., 1984: Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* : III. Zygotic loci on the X-chromosome and fourth chromosome. *Wilhelm Roux's archives of developmental biology.*, **193**, 296–307.

Woodcock, C. L. F.; Safer, J. P.; Stanchfield, J. E., 1976: Structural repeating units in chromatin. *Experimental Cell Research.*, **97**, 101–110.

Wu, R. si; Hong, J. J.; Wu, J. fa; Yan, S.; Wu, D.; Liu, N., 2017: *OVOL2* antagonizes TGF- β signaling to regulate epithelial to mesenchymal transition during mammary tumor metastasis. *Oncotarget.*, **2**, 39401–39416.

Xu, T.; Jiang, X.; Denton, D.; Kumar, S., 2020: Ecdysone controlled cell and tissue deletion. *Cell Death & Differentiation.*, **27**, 1–14.

Yamamoto, S.; Jaiswal, M.; Charng, W. L.; Gambin, T.; Karaca, E.; Mirzaa, G.; Wiszniewski, W.; Sandoval, H.; Haelterman, N. A.; Xiong, B.; Zhang, K.; Bayat, V.; David, G.; Li, T.; Chen, K.; Gala, U.; Harel, T.; Pehlivan, D.; Penney, S. et al., 2014: A Drosophila Genetic Resource of Mutants to Study Mechanisms Underlying Human Genetic Diseases. *Cell.*, **159**, 200–214.

Yamanaka, N.; Rewitz, K. F.; O’connor, M. B., 2012a: Ecdysone Control of Developmental Transitions: Lessons from Drosophila Research. *Annual Review of Entomology.*, **58**, 497–516.

Hill, R. J.; Billas, I. M. L.; Bonneton, F.; Graham, L. D.; Lawrence, M. C., 2012b: Ecdysone Receptors: From the Ashburner Model to Structural Biology *. *Annual review of entomology.*, **58**, 497–516.

Yang, J.; Antin, P.; Berx, G.; Blanpain, C.; Brabletz, T.; Bronner, M.; Campbell, K.; Cano, A.; Casanova, J.; Christofori, G.; Dedhar, S.; Derynck, R.; Ford, H. L.; Fuxe, J.; García de Herreros, A.; Goodall, G. J.; Hadjantonakis, A. K.; Huang, R. J. Y.; Kalchauer, C. et al., 2020: Guidelines and definitions for research on epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology.*, **21**, 341–352.

Ye, G. D.; Sun, G. bin; Jiao, P.; Chen, C.; Liu, Q. F.; Huang, X. L.; Zhang, R.; Cai, W. Y.; Li, S. N.; Wu, J. F.; Liu, Y. J.; Wu, R. S.; Xie, Y. Y.; Chan, E. C.; Liou, Y. C.; Li, B. A., 2016: OVOL2, an Inhibitor of WNT Signaling, Reduces Invasive Activities of Human and Mouse Cancer Cells and Is Down-regulated in Human Colorectal Tumors. *Gastroenterology.*, **150**, 659-671e16.

Yin, H.; Sweeney, S.; Raha, D.; Snyder, M.; Lin, H., 2011: A High-Resolution Whole-Genome Map of Key Chromatin Modifications in the Adult Drosophila melanogaster. (Reik, W., Ed.) *PLoS Genetics.*, **7**, e1002380.

Zabidi, M. A.; Arnold, C. D.; Schernhuber, K.; Pagani, M.; Rath, M.; Frank, O.; Stark, A., 2015: Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature.*, **518**, 556–559.

Zanet, J.; Benrabah, E.; Li, T.; Péliissier-Monier, A.; Chanut-Delalande, H.; Ronsin, B.; Bellen, H. J.; Payre, F.; Plaza, S., 2015: Pri sORF peptides induce selective proteasome-mediated protein processing. *Science.*, **349**, 1356–1358.

Zanet, J.; Chanut-Delalande, H.; Plaza, S.; Payre, F., 2016: Small Peptides as Newcomers in the Control of Drosophila Development. *Current Topics in Developmental Biology.*, Vol. 117pp. 199–219.

Zang, C.; Schones, D. E.; Zeng, C.; Cui, K.; Zhao, K.; Peng, W., 2009: A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics (Oxford, England).*, **25**, 1952–1958.

Zeitlinger, J.; Stark, A.; Kellis, M.; Hong, J. W.; Nechaev, S.; Adelman, K.; Levine, M.; Young, R. A., 2007: RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature Genetics.*, **39**, 1512–1516.

Zeitlinger, J., 2020: Seven myths of how transcription factors read the cis-regulatory code. *Current Opinion in Systems Biology.*, **2020**, 22–31.

Zhang, Y.; Liu, T.; Meyer, C. A.; Eeckhoute, J.; Johnson, D. S.; Bernstein, B. E.; Nusbaum, C.; Myers, R. M.; Brown, M.; Li, W.; Liu, X. S., 2008: Model-based analysis of ChIP-Seq (MACS). *Genome biology.*, **9**, R137.

Zhu, L. J.; Gazin, C.; Lawson, N. D.; Pagès, H.; Lin, S. M.; Lapointe, D. S.; Green, M. R., 2010: ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics.*, **11**, 237.

Zipper, L.; Jassmann, D.; Burgmer, S.; Görlich, B.; Reiff, T., 2020: Ecdysone steroid hormone remote controls intestinal stem cell fate decisions via the PPAR γ -homolog Eip75B in *Drosophila*. *eLife.*, **9**.

Table des abbréviations

2

20E: 20-hydroxyecdysone, 90

4

4C: Chromosome Conformation Capture Sequencing, 51

A

Ac: groupe acétyle, 59
 ACP: Analyse en Composantes Principales, 62
 ADN: Acide Désoxyribonucléique, 10, 31, 37, 38, 39, 40, 41, 42, 43, 44, 45, 49, 50, 51, 52, 54, 56, 57, 58, 59, 62, 66, 67, 68, 76, 108, 128, 129, 130, 135, 140, 146
 AMP: précurseurs de cellules souches rénales, 74
aph-4: gène codant pour la phosphatase alcaline 4, 74
 ARN: Acide Ribonucléique, 27; Acide Rubonucléique, 14, 17, 27, 28, 29, 35, 36, 38, 50, 51, 52, 53, 59, 67, 70, 92
 Asx: Additional sex combs, 65
 ATP: adénosine triphosphate, 51, 136

B

BAC: Bacterial Artificial Chromosome, 42

bam: gène bag of marbles, 13
 BEAF-32: Boundary element-associated factor of 32kD, 71, 141
 BLAST: Basic Local Alignment Search Tool, 43
 BmOvo: forme d'Ovo chez le ver à soie, 12

C

CDS: CoDing Sequence, 10
 CE: cellules étoilés, 73, 74
 ChIP: Immunoprécipitation de Chromatine, 24, 56, 57, 58, 61, 71, 108, 131
 ChIP-seq: Chromatine Immunoprécipitation suivi d'un séquençage des régions fixées, 31, 37, 56, 57, 58, 61, 62, 69, 71, 72, 76, 91, 108, 131, 135
 CHRAC: Chromatin accessibility complex, 136
 Cl: anion chlorure, 74
 CP: cellules principales, 73, 74
 CPG: cellules primaires germinales, 13
 CSI: cellules souches intestinales, 34, 36, 74, 76, 128, 130, 143
 CSR: cellules souches rénales, 74, 76
ct: gène *cut*, 74
Ct: Cut, 73

D

Dam: adénine méthyltransférase, 62

DamID: DNA adenine méthyltransférase identification, 62, 64
 dATP: désoxyadénosine triphosphate, 40
 dCP: gènes développementaux, 71
 dCTP: désoxycytidine triphosphate, 40
 ddATP: didésoxyadénosine triphosphate, 40
 ddCTP: didésoxycytidine triphosphate, 40
 ddGTP: didésoxyguanosine, 40
 ddNTP: didésoxyribonucléotides, 40, 44
 ddTTP: didésoxythymidine triphosphate, 40
 dGTP: désoxyguanosine triphosphate, 40
 DI: facteur Delta, 34
 DNAase: enzyme dégradant l'ADN, 56, 129
 dRAF: Complexe de facteurs associés à dRing, 65, 66
 dRing: SexComb Extra, 64, 65
 dTTP: désoxythymidine triphosphate, 40

E

E(z): Enhancer de zeste, 65, 66
 EB: entéroblastes, 34, 35, 130
 EC: entérocyte, 34, 35, 36, 74, 146
 EcR: Récepteur à l'Ecdysone, 68, 90, 91, 92
 EE: entéro-endocrine, 34, 35

EGF: Epidermal Growth Factor, 23, 26, 36, 74, 90, 91
 EMBL: European Molecular Biology Laboratory, 40
 EMT: Transition Epithélio-Mésenchymateuse, 11, 12, 17, 35, 36, 130
 eRNAs: ARN issu de la transcription de région enhancer, 50
 Esc: ExtraseXComb, 65
esg: gène *escargot*, 35
 Esg: facteur Escargot, 34

F

FISH: d'hybridation in situ par fluorescence, 47
 FS: Frame Shift décalage du cadre de lecture, 29

G

GAF: facteur de type GAGA, 68, 135, 136, 137, 138, 148
 GFP: Green Fluorescent Protein, 29, 53, 54
 Grh: Grainyhead, 130, 131, 132
 GRHL: Grainyhead-like, 130
 GRHL2: Grainyhead-like 2 homologue humain de Grainyhead, 15

H

H⁺: cation hydrogène, 73
 H3K27ac: histone H3 acétylé sur la lysine 27, 59, 68, 69, 70, 71, 108, 142
 H3K27me3: Histone H3 triméthylé sur la

lysine 27, 62, 63, 65, 66, 71, 141

H3K36me3: Histone H3 triméthylé sur la lysine 36, 65, 67, 68, 70, 141

H3K4me1: histone H3 monométhylé sur la lysine 4, 68, 69, 70, 71, 108, 142

H3K4me2: Histone 3 diméthylé sur la lysine 4, 63, 67; Histone H3 diméthylé sur la lysine 4, 62

H3K4me3: histone H3 triméthylé sur la lysine 4, 66, 68, 70, 71, 141

H3K79me3: Histone H3 triméthylé sur la lysine 79, 62, 63, 67

H3K9me2: Histone H3 diméthylé sur la lysine 9, 62, 63, 64, 141

H4K16ac: l'histone H4 acétylé sur la lysine 16, 61

HAT: histone acétyltransférases, 59, 60, 72

HDAC: histones désacétylases, 59, 60, 67, 72

Hh: Hedgehog, 23, 26

Hi-C: high throughput 3C, 128, 139, 140, 141, 142

HisC: Histone Complexe locus codant pour les histones chez la Drosophile, 61

hkCP: gènes de ménage, 71

HKMT: lysine méthyltransférases, 59; lysines méthyltransférases, 60, 64, 72

HMM: Modèle Caché de Markov, 62

K

K: lysine, 59, 60

K⁺: anion potassium, 73

KC167: lignée cellulaire chez la drosophile, 62

M

M2: Master 2ème année, 4, 5, 36

MACS: Model-based Analysis of ChIP-Seq, 40, 76

MCP: Ms2 Coat Protein, 53

me1: monométhylation, 59

me2: diméthylation, 59

me3: triméthylation, 59, 64

MED31: sous-unité 32 du complexe Médiateur, 68

MIGS: Minimum Information about Genome Sequence, 40

ModENCODE: model organism ENCYclopedia Of DNA Elements, 61, 69

mOvo: forme murine d'Ovo, 12

N

Na⁺: cation sodium, 73

NGS: Séquençage Nouvelle Génération, 44, 45

NURF: Nucleosome Remodelling Factor, 90, 136

Nω: composé nitrogène guanidinium terminal de l'arginine, 60

O

ORF: Open Reding Frame (cadre ouvert de lecture), 27, 92

Ovo: Protéine codée par le locus ovo/svb, 10, 12, 13, 31, 108, 128

OvoA: forme répressive germinale de Ovo/Shavenbaby, 10, 12, 13, 23, 27, 28

OvoB: forme germinale activatrice de Ovo/Shavenbaby, 10, 12, 13, 23, 27, 28

ovoD1: Mutation antimorphe dominante de ovoL1, 13

ovoD2: Mutation antimorphe dominante d'ovoL2, 13

ovoD3: Mutation antimorphe dominante d'ovoL3, 13

OvoL: Ovo-like, 10, 11, 13, 14, 15, 16, 17, 108, 130, 146

OvoL1: homologue humain d'ovo, 11, 12, 13, 15, 16

OvoL1-3: formes humaines d'Ovo, 12

OvoL2: homologue humain d'ovo, 13, 14, 15, 16

OvoL3: homologue humain d'ovo, 13

OvoQ6: motif de fixation de Svb, 31, 108, 128

P

Pan: Pangolin, 26

Pc: Polycomb, 64, 65, 66, 146

PcG: complexe protéique Polycomb, 64, 65, 66, 141

Pcl: Polycomb-like, 65, 146

PCR: Réaction en chaîne par polymérase, 40, 42, 44, 54, 140

Ph: Polyhoméotie, 64, 66

Pho: Pleiohomeotie, 65

PhoRC: Complexe Répressif Pho, 65, 66

PMW: matrices poids position, 56, 129

Pol II: ARN Polymérase, 52

polyA: polyAdénosine, 54

PPCD: Dystrophie Cornéenne

Postérieure Polymorphe, 14, 15, 130

PRC1: Complexe Répressif Polycomb 1, 64, 65, 66

PRC2: Complexe Répressif Polycomb 2, 65, 66

PR-DUB: Deubiquitinase Repressive Polycomb, 65, 66

PRE: Polycomb Responsive Element, 51, 65, 66

pri: gène *Polished-rice*, 27, 28, 29, 36, 72, 91, 92

Pri: Peptides Polished-rice, 23, 27, 28, 29, 36, 76, 90, 92, 108, 142

PRMT: arginine méthyltransférases, 59, 60; arginines méthyltransférases, 60

Psc: Posterior Sex Comb, 64, 65

PsqL: forme longue de Pipsqueak, 138

PsqS: forme courte de Pipsqueak, 138

R

R: arginine, 41, 59, 60

Rme1: Composé Arginine terminal monométhylé, 60; Groupement Arginine monométhylé, 60; Groupement Arginine terminal monométhylé, 60

Rme1δ: Composé interne de l'Arginine monométhylé, 60

Rme2as: Composé Arigine diméthyle asymétriquement, 60

Rme2s: Composé Arginine terminal diméthylé symétriquement, 60

RNAi: ARN interférant, 15, 25, 28, 50, 76, 143, 145; expérience

d'interférence à l'ARN, 14

RNA-seq: Séquençage de l'ARN, 5, 14, 27, 34, 36, 37, 57, 64, 72, 92, 108, 131, 143, 144, 146

S

S2: lignée cellulaire de drosophiles, 28, 29, 72, 92, 108, 109, 128, 130, 131, 133, 135

SAM: S- adénosylméthionine, 59, 60

Sc: SexComb Extra, 64, 65

scRNA-seq: Séquençage de l'ARN en cellules uniques, 34, 35

Sd: Scalloped, 26

shh: gène sonic hedgehog, 24

Shh: Sonic hedgehog, 24

smad-ovo: homologue d'ovo chez le ver plat, 14

SmedOvo: forme d'Ovo chez le ver plat, 12

smORF: small Open Reading Frame, 23, 27, 28, 29, 90

sn: gène *singed*, 30

SoxN: SoxNeuro, 26

STARR-seq: self-transcribing active regulatory region sequencing, 54, 55, 56, 71

Su(Hw): facteur Suppressor of Hair wing, 50, 63, 71

Su(z)12: Suppresseur de zeste 12, 65

SUUR: Suppressor of Underreplication, 63, 64

svb: gène *shavenbaby*, 10, 15, 23, 24, 25, 26, 27, 28, 30, 31, 35, 36, 49, 75, 76, 90, 143

Svb: Shavenbaby, 10, 12, 15, 16, 23, 27, 28, 29, 30, 31, 36, 57, 72, 73, 76, 90, 91, 108, 109, 128, 129, 130, 131, 133, 134, 135, 137, 138, 139, 141, 142, 143, 144, 145, 146, 148

SvbAct: forme activatrice somatique de Svb,

28, 29, 36, 76, 108, 128, 139, 141, 142, 143, 145, 148

SvbF7: motif de fixation étendue de Svb, 31, 108, 128

SvbRep: forme répressive somatique de Svb, 28, 29, 30, 36, 76, 108, 128, 133, 139, 141, 142, 143, 145, 146, 147, 148

T

TAD: Domaine associé à la topologie, 139, 141

tal: gène *tarsal-less* (autre nom de *pri*), 27

TF: facteur de transcription, 51, 52, 56, 58, 129, 130

TRR: Trithorax Related, 90

Tsh: Teashirt, 73

TSS: Transcription Start Site, 55, 71, 142

U

Ubr3: ubiquitine ligase E3, 29, 76

Usp: Ecdysone Receptor Ultraspine, 90

UTR: Untranslated Transcribed Region, 50

W

Wg: Wingless, 23, 26, 36, 74, 90

WGS: whole genome shotgun sequencing, 41, 47

Y

Yki: facteur Yorkie, 57, 76

Z

ZAP: Zone de polarisation active, 24

ZGA: activation du génome zygotique, 141

RESUME DE THESE FRANÇAIS

Les facteurs de transcription à doigt de zinc de la famille OvoL, au nombre de trois chez les mammifères (OvoL1, 2 et 3), sont des acteurs majeurs de la régulation épithéliale durant le développement des vertébrés. Chez l'adulte, ils participent à la fois à la différenciation terminale des cellules épithéliales et au maintien de ce caractère pour l'homéostasie tissulaire. Pour ce faire, les facteurs OvoL inhibent l'action des facteurs pro-EMT (Transition Epithélio-Mésenchymateuse) ZEB et TWIST qui eux poussent vers l'état mésenchymateux. Ainsi, lors de cancer humain, OvoL1 empêche la transformation des cellules tumorales épithéliales en cellules invasives mésenchymateuses métastatiques. De plus, des maladies héréditaires, dont des pathologies sévères des yeux et de la peau, ont été associées à la dérégulation des gènes OvoL.

À ce jour et malgré des intérêts biomédicaux potentiels, les mécanismes moléculaires qui soutiennent l'action des facteurs OvoL et notamment leurs gènes cibles spécifiques restent mal compris. Il est possible que la redondance fonctionnelle des facteurs OvoL chez les vertébrés complique l'accès à leur fonction.

Chez la Drosophile (*Drosophila melanogaster*) les facteurs OvoL, fortement conservés au sein du règne animal, sont seulement représentés par le gène *ovo/shavenbaby* (*ovo/svb*), membre fondateur de cette famille. Les deux formes germinales, OvoA et OvoB soutiennent l'ovogenèse alors que la forme somatique, Svb, contrôle le remodelage tridimensionnel des cellules épithéliales. Svb est également nécessaire pour le maintien, la prolifération, mais également la différenciation des cellules souches digestives adultes. Svb est synthétisée sous une forme longue, appelée SvbREP qui, sous l'action conjuguée des petits peptides Polished-rice (Pri) et de l'Ubiquitine-E3-ligase Ubr3, est clivée par le protéasome pour obtenir une forme plus courte, SvbACT, transcriptionnellement active.

Pour comprendre le mode d'action individuel de chacune des deux formes de Svb, nous avons développé un modèle cellulaire (S2) permettant l'expression de l'une ou l'autre des deux formes. Nous avons effectué des expériences de génomique pour déterminer de façon globale le comportement de SvbAct et SvbRep. Pour cela, nous avons analysé leur profil de fixation à l'ADN (ChIP-seq) d'une part, et identifié les gènes cibles (RNA-seq) pour chaque forme d'autre part. L'intégration de ces données a permis de mettre en évidence que SvbREP et ACT se fixent sur les mêmes régions génomiques cis-régulatrices (enhancers) pour contrôler de manière antagoniste l'expression d'environ 250 gènes cibles directs. Nous avons également montré que Svb régule des groupes de gènes différents en fonction du type cellulaire au sein duquel il évolue et mis en évidence que l'environnement chromatinien contraint son activité.

Mes travaux ont non seulement permis de mieux appréhender le mode de fixation de facteurs OvoL sur le génome, mais apportent une nouvelle perspective sur la nature de leurs gènes cibles impliqués dans le contrôle du cycle cellulaire et/ou l'adhésion cellulaire.

ABSTRACT

The OvoL family of zinc finger transcription factors, three in number in mammals (OvoL1, 2 and 3) are major players in epithelial regulation during vertebrate development. In the adult, they participate in both the terminal differentiation of epithelial cells and the maintenance of this character for tissue homeostasis. To do this, OvoL factors inhibit the action of the pro-EMT (Epithelial-Mesenchymal Transition) factors ZEB and TWIST, which push towards the mesenchymal state. Thus, in human cancer, OvoL1 prevents the transformation of epithelial tumour cells into invasive metastatic mesenchymal cells. In addition, hereditary diseases, including severe eye and skin diseases, have been associated with the deregulation of OvoL genes. Currently, despite potential biomedical interests, the molecular mechanisms that underpin the action of OvoL factors, including their specific target genes, remain poorly understood. It is possible that the functional redundancy of OvoL factors in vertebrates complicates access to their function.

In *Drosophila* (*Drosophila melanogaster*), OvoL factors are highly conserved in the animal kingdom and are represented only by the *ovo/shavenbaby* (*ovo/svb*) gene, a founding member of this family. The two germline forms, OvoA and OvoB, support oogenesis while the somatic form, Svb, controls three-dimensional epithelial cell remodelling. Svb is also required for the maintenance, proliferation and differentiation of adult digestive stem cells.

Svb is synthesised in a long form, called SvbREP, which, under the combined action of the small Polished-rice peptides (Pri) and the Ubiquitin-E3 ligase Ubr3, is cleaved by the proteasome to obtain a shorter, transcriptionally active form, SvbACT. To understand the individual mode of action of each of the two forms of Svb, we developed a cell model (S2) that allows the expression of either form. We performed genomic experiments to determine the overall behaviour of SvbAct and SvbRep. To do this, we analysed their DNA binding profile (ChIP-seq) on the one hand, and identified the target genes (RNA-seq) for each form on the other. Integration of these data revealed that SvbREP and ACT bind to the same cis-regulatory genomic regions (enhancers) to antagonistically control the expression of approximately 250 direct target genes. We have also shown that Svb regulates different groups of genes depending on the cell type in which it evolves and that the chromatin environment constrains its activity.

My work has not only provided a better understanding of how OvoL factors bind to the genome, but also provides a new perspective on the nature of their target genes involved in cell cycle control and/or cell adhesion.