

Université  
de Toulouse

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

---

**Présentée et soutenue par :**

**Delphine MENORET**

**le** mardi 3 décembre 2013

**Titre :**

Décodage des réseaux géniques impliqués dans le remodelage cellulaire chez  
l'embryon de *Drosophila melanogaster*

---

**École doctorale et discipline ou spécialité :**

ED BSB : Biologie du développement

**Unité de recherche :**

Centre de Biologie du Développement

**Directeur(s) de Thèse :**

Serge PLAZA

François PAYRE

**Jury :**

Pr Philippe Besse, Président

Dr Pascale Gilardi, Rapporteur

Dr Krzysztof Jagla, Rapporteur

Dr Denis Thieffry, Examineur

Pr Christel Lutz, Invitée



*Au bout de 4 années de travail, voici enfin venu le temps des remerciements pour tous ceux sans qui cette thèse n'aurait pas pu être ce qu'elle est.*

*Dans un premier temps je tiens à remercier bien évidemment François et Serge, mes encadrants de thèse qui ont pris le risque de me choisir pour mener ce projet ambitieux. Je tiens également à remercier Marc Haenlin et Gwen, qui m'ont apporté leur soutien lors des moments difficiles, et qui m'ont aidée à prendre les bonnes décisions.*

*De nombreux collaborateurs ont participé à ce travail : S. Aerts, R. Spokony, V. Hakim, I. Gonzalez. Un grand merci à M. Santolini, qui a mené de nombreuses analyses en « CRMologie », et qui nous a fait part de son expérience bio-informatique, permettant d'arriver à une publication de grande qualité.*

*Bien qu'elle ne figure pas en tant qu'encadrante officielle sur le manuscrit, je tiens à adresser un message particulier à Hélène. Tu m'as prise sous ton aile quand je suis arrivée dans l'équipe, et tu m'as initiée à toutes les techniques qui m'étaient alors inconnues. On se l'est souvent dit entre nous mais je le répète, un de mes grands regrets est de n'avoir pas pu travailler avec toi en plus étroite collaboration, nos projets respectifs ayant pris le dessus à chaque fois. Tu m'as également soutenue lors des étapes plus critiques de ma thèse : papier CRM, révisions, écriture du manuscrit, en n'hésitant pas à me consacrer du temps pour m'aider. Tu fais partie des personnes grâce auxquelles j'ai pu achever cette thèse alors que cela me paraissait insurmontable ;op.*

*Phiphi, je t'ai fais souffrir pendant 3 ans en réunions d'équipe en te submergeant de « vulcania plot » et autres histogrammes en tout genre, mais tu m'as toujours apporté ton aide au quotidien, par exemple lors de mes galères de clonage ;op !!!! Grâce à toi je suis enfin convaincue que « pour que ça réussisse il suffit d'une colonie !!!!! ». Je te remercie donc pour tous ces conseils et ton amitié !!*

*Jenny, Cédric et Pierre, nos discussions rugbystiques étaient pour moi un vrai moment de détente. Je savais qu'un match du Stade ou de l'équipe de France pendant le weekend amènerait une bonne discussion le lundi !*

*Yvan, l'homme aux 12 millions d'hybridations in situ, je voudrais te remercier pour toute l'aide que tu m'as apportée lorsqu'il s'agissait de tester tous mes candidats. Tu m'as ouvert les yeux sur le monde merveilleux des in situ ou immuno-in situ, avec tes protocoles simples ou améliorés, 2, 3 ou 4 jours, et adaptables à toutes les situations !!!!!*

*Je tiens à adresser un message particulier aux thésards de l'équipe, Ahmad et Emilie les « vieux », et les « nouveaux » Azza et Jérôme. Vous m'avez vu galérer jusqu'au bout pour parvenir à finaliser cette thèse, et je souhaiterais que vous vous accrochiez pour finir vous aussi. Dites-vous que vous faites tout cela pour vous et non pour les autres, que cela peut paraître juste infaisable et insurmontable à certains moments, mais comme quelqu'un me l'a très justement conseillé, il faut à ce moment-là savoir se reposer sur nos proches et éliminer toute interaction que vous jugerez néfaste.*

*Audrey, ma morue préférée, ton arrivée dans l'équipe nous a permis de créer une belle amitié entre nous, tu m'as apporté un soutien immense durant ces 2 dernières années, en n'hésitant pas à mêler leçon de morale et réconfort version « maman Audrey » quand tu voyais que je carburais aux paninis nutella, kinder bueno et petits écoliers trempés dans un chocolat chaud, signe d'une grande détresse de*

*ma part ;op !!!! Ta ténacité pour te construire vie privée et professionnelle qui te conviennent m'a toujours impressionnée, et je fais tout pour copier ton optimisme !!!*

*Juju, ma tutrice, je tiens à te remercier pour ton amitié, tu as su voir quand j'étais à bout de nerfs, et tu m'as soutenue sans chercher à m'influencer sur les décisions à prendre. Ton enthousiasme pour mon « super projet » comme tu l'appelais m'a aidée à passer les moments difficiles.*

*Je voudrais aussi remercier tous les thésards du CBD, devenus de vrais amis au fil du temps : Amir, Ismaël, Bibi, Sandra, Clément, Aurélie, Mathilde, Luis.*

*Emily, je ne pouvais pas ne pas finir ces remerciements CBD par toi, tu t'en doutes bien. Mon pétage de plombs « je vais courir nue dans les bois » quand j'ai reçu les résultats de transcriptome a été le début d'une grande amitié entre nous. Je pense que nos nombreux points communs (trop nombreux parfois, c'est limite flippant) ont contribué à nous rapprocher, quoi de plus beau que de se raconter ses looses improbables de la vie quotidienne, enchainements ininterrompus de galères qui nous font passer nos pires journées de merde, mais qui représentent tellement bien nos vies !!!! C'est moche à dire, mais de savoir que tu serais là sur le palier à commencer à sourire dès que je dirais « tu veux savoir un truc drôle ? Enfin drôle... » puis rire à mes galères et me faire partager les tiennes était d'un grand réconfort !!! Tu as également été présente lors des périodes les plus difficiles, et tu as été la première à ne pas me dire « mais si ! Faut absolument que tu continues !!!! », la bonne réaction que toi seule pouvait avoir... Et pour m'encourager lors de la dernière ligne droite, comme tu ne pouvais pas être présente à mes côtés, tu m'as envoyé ton meilleur agent, l'amie Cahouète qui a contribué à me changer les idées un mois avant la soutenance !!! Qui d'autres d'ailleurs que toi pouvait se douter qu'une poule pourrait me faire le plus grand plaisir comme cadeau de thèse... Maintenant cette amitié continue en dehors du labo, toujours plus solide je l'espère, et qui s'enrichit tous les jours !!!!!*

*Je tiens à remercier aussi les membres du club de rugby de Labastide-Beauvoir, anciens et actuels, pour leur amitié sincère : Mathias, Geoffrey, Poupougne, Jérôme R et Jérôme G, Joël, Pip, Antoine, Gaëlle, Elodie, Corinne, Caro, Stéphanie ; quand moi-même je n'avais plus goût à ma thèse, les deux Jérôme ont su, par l'intérêt qu'ils y portaient et les questions qu'ils me posaient, me refaire entrevoir que mon projet pouvait être intéressant.*

*Bien évidemment, un remerciement énorme à mes parents et mon frère, qui m'ont soutenue au quotidien, prêts à tout pour m'apporter leur aide, et qui ont également su me changer les idées quand cela s'est avéré nécessaire. Rien de tel qu'un séjour au grand air à Noirmoutier pour recharger mes batteries et repartir au combat !!!*

*Enfin, il n'y a pas de mots pour décrire ce que je dois à Guigui. Ton amour, ton écoute et ton soutien m'ont permis de finaliser cette thèse, un challenge que je ne me serais jamais sentie capable de relever sans toi. Tu as su m'apporter du réconfort, mais aussi être compréhensif face à mes innombrables pondoirs, minipreps ou transfo du dimanche après-midi. Si ce travail a abouti, c'est à toi que je le dois.*

*Gilles, j'ai toujours senti ta présence à mes côtés pendant ces 4 ans, c'est en pensant à toi que j'ai réussi à finir ce travail.*

*Cette thèse est pour toi.*

## Table des matières

Préambule .....	1
-----------------	---

### INTRODUCTION

<b>I. Régulation génique et développement .....</b>	<b>2</b>
<b>Transcription et régulation transcriptionnelle .....</b>	<b>2</b>
La transcription eucaryote, quelques généralités .....	2
Les Facteurs de Transcription (FTs) .....	3
<b>Des régions régulatrices essentielles : les CRMs .....</b>	<b>4</b>
Des « receptor genes » vers les CRMs .....	4
Recherche des éléments cis-régulateurs <i>in vivo</i> .....	5
<b>Les différents types de CRMs .....</b>	<b>6</b>
Le modèle enhanceosome .....	6
Le modèle billboard .....	7
Le CRM <i>eve2</i> , paradigme d'élément cis-régulateur .....	8
<b>L'apport des approches globales et de la bio-informatique .....</b>	<b>9</b>
Le décodage du génome .....	9
Le CHIP-seq .....	12
L'analyse du transcriptome via les puces à ADN .....	13
<b>II. La morphogénèse épidermique, modèle d'étude de la spécificité tissulaire d'un FT .....</b>	<b>15</b>
<b>Mise en place du tissu épidermique au cours du développement .....</b>	<b>15</b>
La cuticule .....	15
Différentes couches superposées forment la cuticule .....	16
De nombreuses protéines composent la cuticule .....	17
Développement des cellules épidermiques .....	18
Différenciation des cellules épidermiques .....	18
Morphologie des cellules épidermiques à trichomes .....	18
<b>Le couple <i>Svb</i> / <i>Pri</i>, déterminants du patron des trichomes .....</b>	<b>19</b>
<i>Svb</i> , régulateur transcriptionnel de la morphogénèse .....	19
Le locus <i>Ovo</i> / <i>svb</i> .....	20
<i>Svb</i> , nécessaire et suffisant pour la formation des trichomes .....	20
<i>Svb</i> intégrateur de voies de signalisation antagonistes .....	21
<i>Pri</i> , un régulateur inattendu de la morphogénèse .....	23
<i>pri</i> est exprimé à différentes étapes du développement .....	23
<i>Pri</i> , régulateur de la protéine <i>Svb</i> .....	24
<b><i>Svb</i> active directement une batterie d'effecteurs cellulaires .....</b>	<b>24</b>
<i>Svb</i> , dernier FT de la différenciation des cellules épidermiques .....	24
Trois familles de gènes cibles identifiées .....	25
Protéines de liaison à l'actine .....	25
Protéines de la cuticule .....	26
Protéines de la matrice extracellulaire .....	26
Identification des CRMs associés aux gènes cibles .....	27

## RESULTATS

<b>Partie I. Identification des cibles <i>in vivo</i> du facteur Svb et décryptage de leur code de régulation .....</b>	<b>33</b>
<b>INTRODUCTION.....</b>	<b>34</b>
<b>RESULTATS ADDITIONNELS ET DISCUSSION .....</b>	<b>36</b>
Réalisation et validation du transcriptome.....	36
Vers l'identification des cofacteurs de Svb.....	40
Organisation fonctionnelle des CRMs dépendants de Svb.....	42
Svb, facteur pionnier de la transcription.....	46
<b>CONCLUSION .....</b>	<b>48</b>
<b>Partie II. Identification de nouvelles fonctions cellulaires en aval de Svb pour la morphogénèse des cellules à trichomes.....</b>	<b>49</b>
<b>INTRODUCTION.....</b>	<b>50</b>
<b>RESULTATS .....</b>	<b>51</b>
Validation par hybridation <i>in situ</i> de 30 nouveaux effecteurs de Svb.....	51
De nouvelles fonctions dépendantes de Svb pour la morphogénèse .....	53
Un nouveau rôle de Svb dans la définition du site d'implantation des trichomes.....	55
Participation des modules GLOBAL et CORE dans la morphogénèse épidermique...58	
<b>DISCUSSION.....</b>	<b>64</b>
<b>Partie III. Nouvelles fonctions embryonnaires des peptides Pri et identification des effecteurs cibles .....</b>	<b>69</b>
<b>INTRODUCTION.....</b>	<b>70</b>
<b>RESULTATS .....</b>	<b>71</b>
Rôle de Pri dans la différenciation des trachées .....	71
A la recherche des gènes dont l'expression est augmentée .....	71
Les peptides Pri régulent l'expression de composants de la cuticule.....	74
<b>DISCUSSION.....</b>	<b>77</b>

## CONCLUSION et PERSPECTIVES

<b>A. Vers un changement d'échelle de l'analyse des systèmes biologiques. ....</b>	<b>83</b>
1. Opportunités et enjeu.....	83
2. Apports et limites des approches de transcriptome .....	85
3. Apports et limites des approches de CHIP-seq.....	91
4. Apports et limites de la bio-informatique.....	93
5. Coupler des approches globales pour répondre à une question biologique .....	94
<b>B. Les enhanceurs de la morphogénèse épidermiques, progrès et futurs challenges. ....</b>	<b>95</b>
1. Présence et rôles potentiels des « sites fantômes » de fixation des FTs.....	95
2. Des modèles plus souples pour l'architecture des CRMs.....	96
3. Conservation évolutive des CRMs : contrainte ou mirage ? .....	96
<b>C. La découverte des effecteurs de la morphogénèse. ....</b>	<b>97</b>
1. Les limites de la prédiction fonctionnelle .....	97
2. Qu'avons-nous appris de l'analyse genome-wide ? .....	98
3. PCP : vive la différence ! .....	98

<b>D. Une nouvelle classe de régulateurs du développement : les petits peptides.....</b>	<b>101</b>
1. L'importance quantitative et fonctionnelle des lncRNAs .....	101
2. You said long NON-CODING RNAs.....	101
3. De nouvelles fonctions régulatrices des peptides Pri .....	102
<b>E. Conclusion .....</b>	<b>103</b>

<b>REFERENCES .....</b>	<b>105</b>
-------------------------	------------

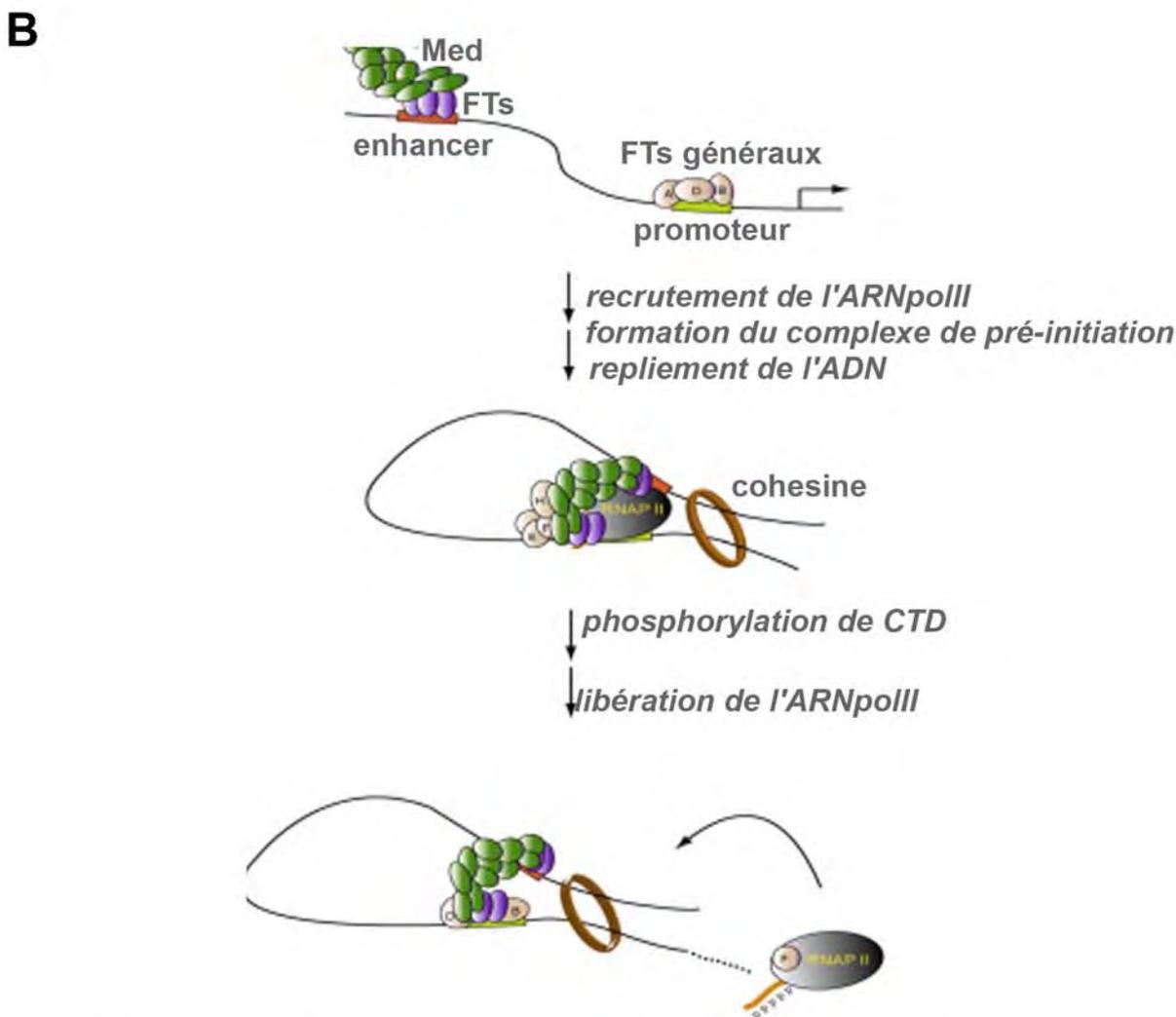
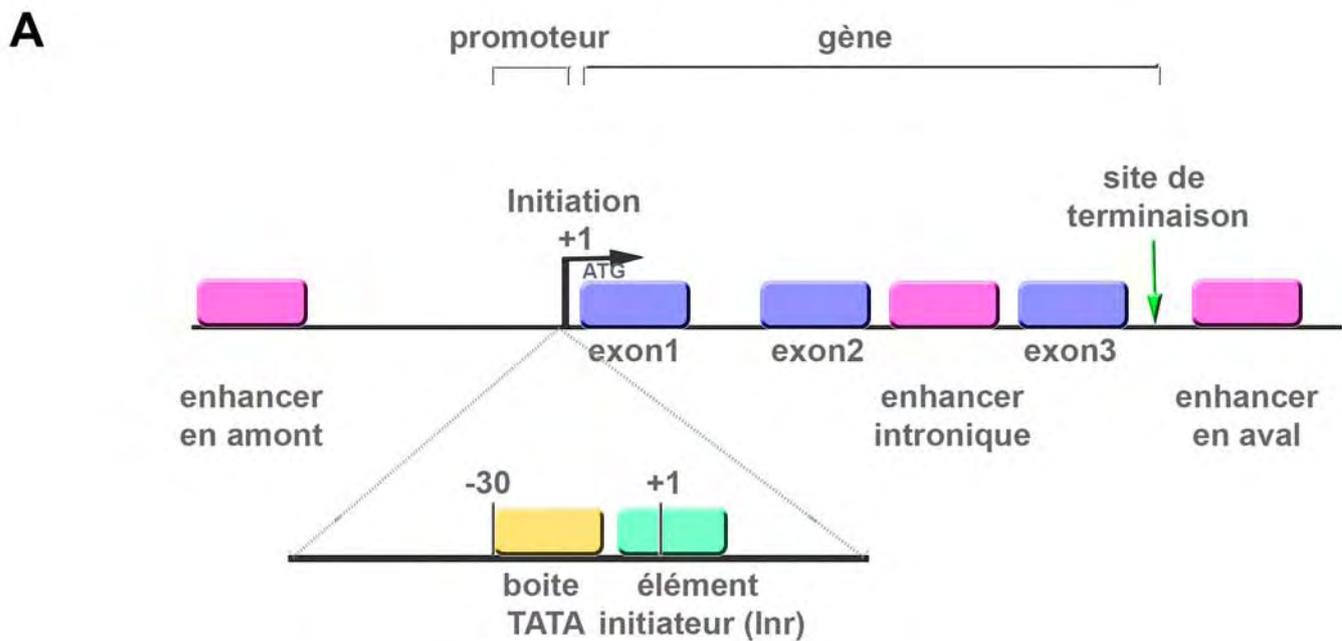


## Préambule

Une caractéristique frappante du monde animal est l'extraordinaire variété des types cellulaires qui composent un organisme adulte, illustrée par exemple par la diversité morphologique des cellules. Il est bien établi que les différents types cellulaires se spécifient puis se différencient au cours de l'embryogenèse, grâce à la définition progressive des sous-ensembles de gènes qui s'expriment dans chaque cellule. On sait aujourd'hui que virtuellement toutes les étapes de l'expression génique sont soumises à des régulations, depuis la structure de la chromatine, la transcription, l'épissage, l'export des ARN, leur localisation et traduction, et bien sûr le large spectre des modifications post-traductionnelles.

Dans ce manuscrit, nous nous concentrerons sur les mécanismes régulant la transcription des gènes au cours du développement. Si les concepts généraux sont formulés depuis longtemps, il semble bon de rappeler que nous connaissons toujours très mal : i) les éléments physiques implantant la différenciation terminale des cellules, ii) la logique du code de régulation de l'expression du génome. En effet, nous verrons que notre compréhension des Réseaux de Régulation Génique reste dominée par quelques exemples phares régissant la mise en place des territoires embryonnaires au cours développement précoce. D'autre part, contrairement aux unités de transcription codant des protéines dont le code génétique a été décrypté au cours des années 60, la nature du (ou des) code(s) de régulation de leur expression reste globalement mal comprise.

**L'objectif de mes travaux de thèse a été d'aborder ces deux problèmes, par une analyse systématique des mécanismes de régulation géniques responsables de la différenciation de l'épiderme, au cours de l'embryogenèse de drosophile.**



**Figure 1 : Organisation d'un locus et transcription chez les Eucaryotes.**

**A. Organisation classique d'un locus Eucaryote.**

Le gène est organisé en exons successifs (violet) en aval du promoteur qui contient le site d'initiation de la transcription (+1) et les boîtes TATA (orange) et Inr (vert) reconnues par la machinerie de transcription. Des régions enhancer (roses) peuvent être localisées en amont ou en aval du gène, ainsi que dans un intron.

**B. Initiation de la transcription Eucaryote.**

La fixation de FTs (ovales roses) sur un enhancer (rectangle orange) permet le recrutement du complexe Mediateur (Med, vert foncé). L'interaction de celui-ci avec les FTs généraux fixés sur le promoteur du gène induit le recrutement de l'ARNpolIII (complexe de pré-initiation) et permet le repliement de l'ADN maintenu par la cohésine (anneau). L'ARNpolIII est libérée après phosphorylation de CTD.

D'après Borggreffe and Yue 2011.

## I. Régulation génique et développement

Alors que toutes les cellules d'un organisme contiennent le même génome, celles-ci vont acquérir au cours de l'embryogénèse des identités différentes suite à l'expression d'un sous-ensemble spécifique de gènes. Comprendre en détail le contrôle de la transcription représente donc un enjeu fondamental pour la biologie du développement.

### I.1. Transcription et régulation transcriptionnelle

La transcription dirige l'expression des gènes assurant les fonctions cellulaires générales. Elle est également finement régulée spatialement et temporellement pour définir, dans chaque type cellulaire, l'expression des gènes de différenciation.

#### I.1.1. La transcription eucaryote, quelques généralités

Chez les eucaryotes, la transcription s'effectue par l'ARN polymérase II recrutée sur la région promotrice du gène, par l'intermédiaire de complexes protéiques adaptateurs se fixant à l'ADN (FTIID). La transcription est contrôlée par des régions génomiques cis régulatrices indépendantes, dites enhancers et silencers, ou collectivement modules cis régulateurs (CRM), servant de plateforme pour la fixation des Facteurs de Transcription (FT). Ces régions peuvent être localisées au voisinage du site d'initiation, en amont ou en aval du gène, ainsi que dans les introns, rendant difficile leur identification (Fig 1A). La communication entre FTs et le complexe d'initiation implique le complexe Médiateur (Borggreffe and Yue 2011) agissant comme interface transcriptionnelle, pour transmettre les informations du CRM au promoteur (Fig 1B) (Malik and Roeder 2011).

De nombreux travaux ont bien établi l'importance du contrôle de l'expression des gènes au cours du développement. Le décryptage des mécanismes de reconnaissance et de sélection des régions régulatrices par les FTs est donc primordial pour comprendre la spécificité transcriptionnelle responsable de la formation et différenciation des tissus.



### I.1.2 Les Facteurs de Transcription (FTs)

Notamment en réponse à des signaux cellulaires, l'activation ou la répression de la transcription des gènes est assurée par les FTs dont le rôle est de spécifier l'expression des gènes dans un groupe de cellules déterminé, à un stade développemental donné. L'expression des gènes codants les FTs est bien sûr contrôlée par les mêmes mécanismes et son altération peut provoquer des défauts spectaculaires. Par exemple le FT *eyeless* contrôle la formation des yeux de drosophile. L'expression artificielle d'*eyeless* entraîne la formation d'yeux supplémentaires, par exemple suite à son expression dans les pattes (Halder, Callaerts et al. 1995). Les FTs sont composés de domaines fonctionnels indépendants. Chaque FT contient au moins un domaine de liaison spécifique à l'ADN, ainsi que des domaines additionnels responsables des propriétés d'activation et/ou répression de la transcription.

La nature des séquences reconnues par un FT constitue donc le premier paramètre de sa spécificité d'action. Diverses approches ont été développées pour identifier le site de fixation pour un FT donné. Les FT fixent en général de courtes séquences, souvent avec code de reconnaissance assez relâché (ou dégénéré). La diversité des séquences reconnues par un FT est représentée par une matrice poids position (PWM), pondérant chaque position par la fréquence des variations nucléotidiques observées (Fig 2). Malheureusement on ne dispose de PWMs que pour une fraction restreinte des différents FTs d'une espèce. Par exemple chez la drosophile, la base de données Redfly (<http://redfly.ccr.buffalo.edu/>) répertorie 2051 sites de fixation, correspondant à 158 FTs (d'après 662 publications). Le génome de drosophile code environ 1000 FTs, montrant que l'on dispose d'informations de PWMs pour moins d'un cinquième des différents FTs. En outre, ces PWM sont souvent bâties sur l'analyse d'une faible fraction des sites génomiques, en moyenne une douzaine de sites par facteur ! On mesure donc combien les connaissances actuelles sur les interactions FT/PWM restent lacunaires.



## I.2. Des régions régulatrices essentielles : les CRMs

### I.2.1. Des « receptor genes » vers les CRMs

L'idée selon laquelle la différenciation cellulaire résulte du contrôle de l'expression des gènes est apparue dès les années 50, suite à l'observation de différences d'activité des histones entre les tissus (Stedman 1950). Britten et Stedman émettent l'hypothèse de l'existence de « receptor » gène, qu'ils définissent comme une séquence d'ADN liée à un gène « producer ». Cette séquence régulatrice est responsable de la transcription du gène « producer », quand un complexe spécifique est formé entre une molécule activatrice et le « receptor » (Britten and Davidson 1969). Le concept de CRM est donc déjà sous-jacent, de même que la possibilité de redondance entre sets de « receptor » et/ou de « producer ». La notion d'enhancer est explicitement formulée au cours des années 80, avec l'identification d'une région localisée 150 pb en amont du gène de l'antigène T de SV40, nécessaire à sa transcription (Benoist and Chambon 1981).

Une étape importante est ensuite la caractérisation d'enhancers associés à des gènes importants pour le développement, notamment chez la drosophile. En 1989, Doyle analyse la régulation de l'expression embryonnaire de *zerknüllt* (*zen*), un gène impliqué dans la spécification dorso-ventrale. Il identifie plusieurs éléments de régulation de *zen*, chacun participant à la mise en place du patron dorso-ventral (Doyle, Kraut et al. 1989). La même année Logan dissèque le locus *yp1* et *yp2* et identifie deux régions distinctes, l'une dirigeant l'expression de ces gènes dans le corps gras et l'autre dans les ovaires (Logan, Garabedian et al. 1989). Il apparaît rapidement que les régions importantes pour diriger l'expression des gènes semblent particulièrement conservées au cours de l'évolution. Par exemple, la conservation de régions génomiques entre *Drosophila affinis* et *mojavensis* suggère leur importance pour la régulation de *Adh* (Ayer and Benyajati 1990). Des cas particuliers apparaissent également, comme la découverte d'enhancers autorégulateurs pour *Deformed*, un gène homéotique impliqué dans la régionalisation antéro-postérieure (Bergson and

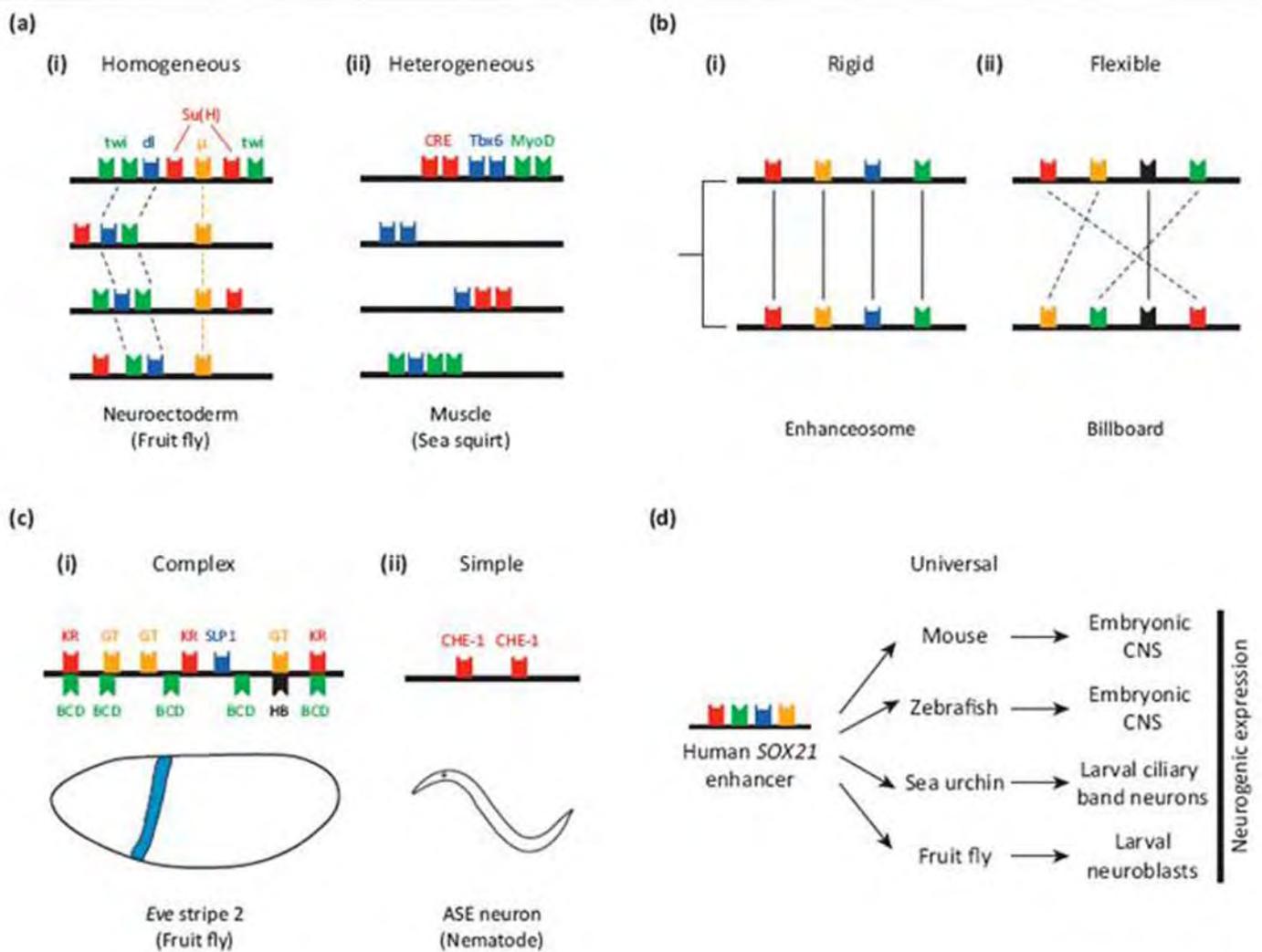


McGinnis 1990), et *fushi tarazu* (Schier and Gehring 1993). Ces études pionnières ont révélé que les CRMs jouent un rôle important pour l'expression de gènes du développement et peuvent être associés à des îlots de conservation évolutive au sein des régions non codantes du génome.

Pour accélérer l'identification des séquences cibles des FTs, on a développé des approches haut débit : "bacterial one-hybrid" (Noyes, Meng et al. 2008), "protein binding microarrays" (Berger and Bulyk 2006), "MITOMI" (Maerkl and Quake 2007), permettant l'analyse d'un plus grand nombre de FTs. Ces PWMs sont répertoriées dans des bases de données publiques : JASPAR (Portales-Casamar, Thongjuea et al. 2010), UNIProbe (Robasky and Bulyk 2010), FlyFactorSurvey (Zhu, Christensen et al. 2010) ou commerciales, *e.g.* TRANSFAC. En l'état, cet effort de rassemblement des données disponibles souligne finalement la relative pauvreté des informations actuelles.

### 1.2.2. Recherche des éléments cis régulateurs *in vivo*.

La transgénèse par élément P chez la drosophile, permettant de tester l'effet d'une région génomique sur la transcription d'un gène rapporteur, a constitué un saut technologique pour l'analyse des séquences régulatrices *in vivo* (Hiromi, Kuroiwa et al. 1985; Glaser, Wolfner et al. 1986). Malgré sa lourdeur, on a largement utilisé cette approche pour localiser, fonctionnellement, les CRMs de régulateurs du développement (Gallo 2013). Le développement de la transgénèse site spécifique  $\text{ØC31}$  a amélioré la technique, en augmentant le taux de transgénèse (de 5% à 50%) et prévenant l'influence des effets de position, dû à l'insertion aléatoire de l'élément P, sur l'expression des transgènes (Bischof, Maeda et al. 2007). Le test par transgénèse de l'activité *in vivo* d'un CRM reste cependant un goulot d'étranglement, incitant au développement d'approches alternatives, expérimentales



**Figure 3. Modèles d'organisation des CRMs.**

**(a). Homogène et hétérogène.**

Des CRMs homogènes présentent une organisation globalement identique concernant l'identité des sites de fixation (i). Le cas des CRMs dirigeant une expression dans le neuroectoderme chez la drosophile est montré : les sites Twist (vert), Dorsal (bleu), Su(H) (rouge) et mystery (jaune) sont présents. Une contrainte de localisation est spécifique au site mystery. Des CRMs hétérogènes ne présentent pas les mêmes sites de fixation ni la même organisation (ii). L'exemple des CRMs dirigeant une expression musculaire chez *Ciona* est montré.

**(b). Enhanceosome et billboard.**

Des CRMs peuvent présenter une contrainte forte sur le nombre et l'organisation des sites les uns par rapport aux autres (i). Certains CRMs présentent une organisation flexible sans contrainte organisationnelle (ii).

**(c). Simplicité et complexité.**

L'organisation d'un CRM peut être complexe (nombre de sites différents, orientation) (i) alors que certains CRMs présentent les mêmes sites en petit nombre (ii). Le cas de *eve2* est donné en tant que CRM complexe car il présente des sites pour 5 FTs différents, alors qu'un CRM dirigeant une expression neuronal chez le nématode est composé de 2 sites identiques.

**(d). Conservation.**

L'organisation d'un CRM peut être conservée entre différentes espèces (cas du CRM de *sox21*, dirigeant une expression neurogénique).

D'après Yanez Cuna, Stark 2012.

ou prédictives, à l'échelle du génome entier (Rister and Desplan 2010; Stark 2013) (Bulyk 2013).

### 1.3 Les différents types de CRM.

Les CRMs régulant les étapes précoces du développement doivent intégrer de multiples informations (*i.e.*, lier de nombreux FTs) pour définir un registre spatial d'expression. Il est souvent proposé que ces éléments doivent présenter une architecture fonctionnelle plus complexe que des CRMs dirigeant les étapes de différenciation terminale (Fig 3c), dont l'activité pourrait répondre à un nombre plus restreint (un seul ?) de FT. On a donc cherché à classifier les CRMs selon l'organisation de leurs différents sites, et deux modèles alternatifs ont émergés : le modèle "enhanceosome" et le modèle "billboard" (Meireles-Filho and Stark 2009).

#### 1.3.1 Le modèle enhanceosome

L'enhanceosome représente un CRM avec une architecture fonctionnelle très contrainte des différents motifs qui le composent (Fig 3b). A la fois le nombre, l'arrangement et la distance respectifs des motifs les uns par rapport aux autres sont critiques pour l'activité du CRM, probablement car elle nécessite des interactions synergiques entre différents FTs. Ce modèle dérive de l'analyse du gène de l'interféron- $\beta$ , dont un CRM principal requiert une organisation précise des sites de fixation (*e.g.*, du nombre de tours d'hélice d'ADN qui les séparent) pour être fonctionnel (Thanos and Maniatis 1995). L'absence d'un FT ou d'un site de fixation abolit généralement l'activité du CRM.

Suite à ces travaux, on a pu identifier des CRMs co-régulés en recherchant des régions présentant une organisation similaire de motifs. Ainsi 4 CRMs neurogéniques contenant des sites de fixation pour les FT Dl, Tw, Su(H), et le "mystery site", présentent des caractéristiques semblables : une taille d'environ 300bp, 20pb entre les sites Dl et Tw,

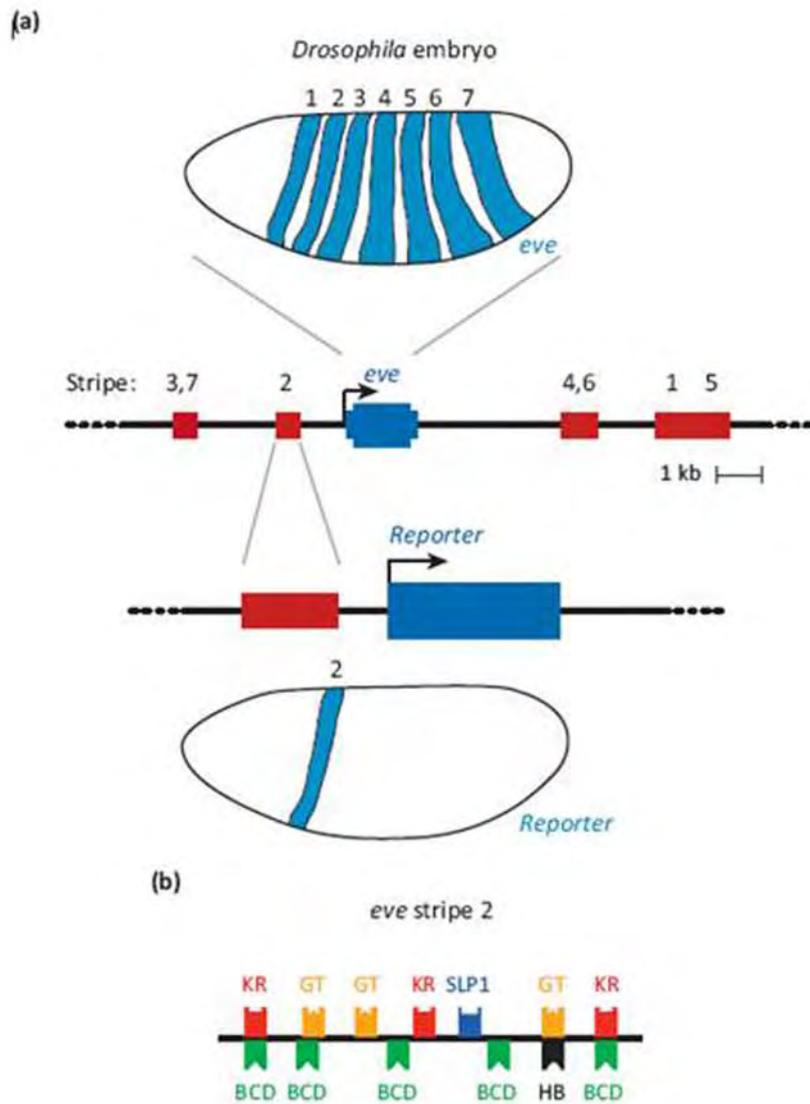


100 à 150 bp entre les sites mystery et Tw, une orientation caractéristique du site Su(H) (Fig 3a) (Erives and Levine 2004). L'analyse de CRMs associés au développement précoce montre que certaines combinaisons de sites de fixation semblent sur-représentées, soutenant que leur organisation est elle aussi contrainte et structurée (Papatsenko, Goltsev et al. 2009). Cependant, il semble que les CRM aussi contraints que l'enhanceosome de l'interféron- $\beta$  soient assez rares. Ils pourraient correspondre à des éléments intégrateurs de différents FTs pour une résultante transcriptionnelle discrète, de type ON/OFF.

### I.3.2 Le modèle billboard

Le modèle billboard présente au contraire une architecture très flexible, apparemment aléatoire, des différents motifs régulateurs nécessaires à l'activité du CRM (Fig 3b). Par exemple, 19 CRMs dirigent une expression musculaire similaire chez *Ciona*, sans présenter une architecture apparemment stéréotypée de différents sites (Fig 3a) (Brown, Johnson et al. 2007). Ces multiples sites peuvent agir de manière indépendante, et ensemble permettent des réponses transcriptionnelles variées. Par exemple, Kulkarni et Arnosti (Kulkarni and Arnosti 2003) ont montré qu'un CRM est capable de transmettre un signal activateur ou répresseur dans le même noyau. Il intègre donc différents signaux de façon indépendante.

La conservation évolutive des motifs composants un CRM a été souvent vue comme une preuve, au moins indirecte, d'une architecture de type enhanceosome (Fig 3d). En revanche, la flexibilité "apparente" des sites au sein d'un même CRM à travers l'évolution de différentes espèces suggère une organisation de type billboard.



**Figure 4. Caractéristiques du CRM *eve2* associé au gène *even-skipped*.**

**a. Locus *eve* et CRM *eve2*.**

*eve* (rectangle bleu) est exprimé dans 7 bandes au stade blastoderme syncitial. Ce patron est récapitulé par l'activité de 5 CRMs (rectangles rouges). Le CRM *eve2* dirige une activité spécifiquement dans la bande 2.

**b. Organisation de *eve2*.**

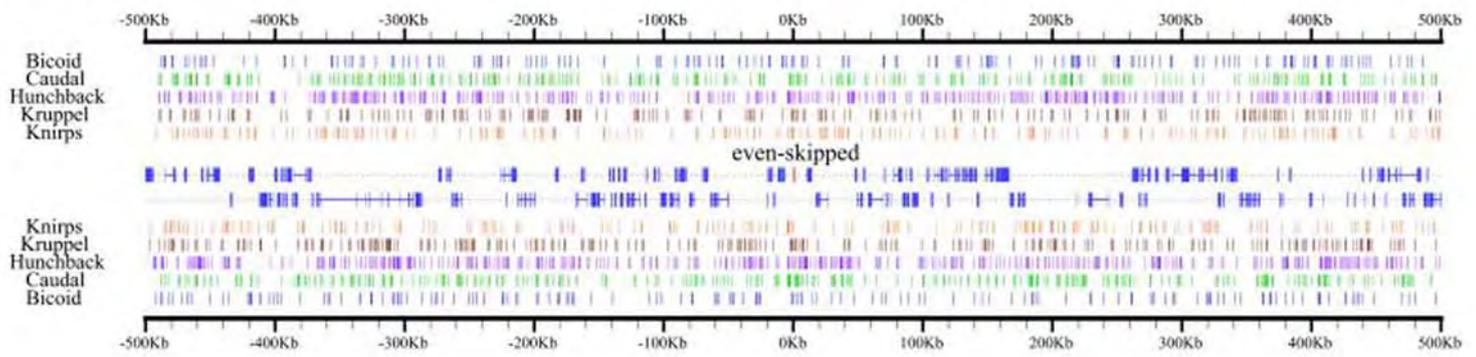
*eve2* contient différents sites de fixation liés par les FTs Kruppel (rouge), Giant (jaune) Sloopy-paired (bleu) Bicoid (vert) Hunchback (noir). Leur organisation est nécessaire pour diriger une expression dans la bande 2. D'après Yanez Cuna, Stark 2012.

### I.3.3 Le CRM *eve2*, un paradigme d'élément cis-régulateur.

L'un des CRMs les plus étudiés est probablement celui dirigeant l'expression de la bande 2 d'*even-skipped* (*eve*). *eve* est exprimé en 7 bandes au stade blastoderme syncytial via l'action de 5 CRMs : *eve1*, *eve2*, *eve5*, *eve 3+7* et *eve4+6*. Le CRM *eve2* a été identifié par analyse systématique (Goto, Macdonald et al. 1989), puis réduit à une taille de 480bp (Small, Blair et al. 1992) (Fig 4a). Certains régulateurs transcriptionnels ont été identifiés par approche génétique dégageant le concept d'une action concertée de FTs de la segmentation (Frasch and Levine 1987). Des approches moléculaires ont établi que les FTs Bicoid (Bcd), Hunchback (Hb), Krüppel (Kr), Giant (Gt), régulent directement le CRM *eve2* (Small, Kraut et al. 1991; Stanojevic, Small et al. 1991; Small, Blair et al. 1992), qui contient 6 sites Kr, 5 Bcd, 3 Gt, et 3 sites de liaison de Hb (Small, Blair et al. 1992) (Fig 4b). L'analyse d'*eve2* a mis en évidence la coopération synergique de ces différents FTs pour contrôler précisément son activité (Small, Blair et al. 1992; Arnosti, Barolo et al. 1996). Des clusters (hétéro- et/ou homotypiques) de sites de fixation ont été retrouvés dans d'autres CRM contrôlant la segmentation, et il est souvent inféré que les CRMs correspondent à l'accumulation locale de sites de fixation de FTs dans les régions non codantes (Read, Nishigaki et al. 1990). Ludwig et Kreitman (Ludwig and Kreitman 1995; Ludwig, Patel et al. 1998) proposent que la présence de sites multiples probablement (au moins en partie) redondants confère au CRM une certaine robustesse (ou adaptation) face aux mutations. La présence de sites multiples pourrait aussi favoriser une réponse graduelle en fonction de la concentration locale en FT.

*Eve-2* a été utilisé comme modèle de l'évolution d'un CRM entre espèces. Hare et al. (Hare, Peterson et al. 2008) ont effectué une analyse comparative des séquences d'*eve-2* dans différentes espèces et ont mesuré l'activité de ces régions orthologues chez *D. melanogaster*. Puisque des éléments orthologues, avec une composition et arrangement en sites de TFs très variables, dirigent un profil d'expression similaire, les auteurs concluent que l'architecture fonctionnelle d'*eve-2* est très flexible. Ces arguments ont été cependant remis en cause par

## (A) High stringency matches



## (B) High stringency matches and clustering filter

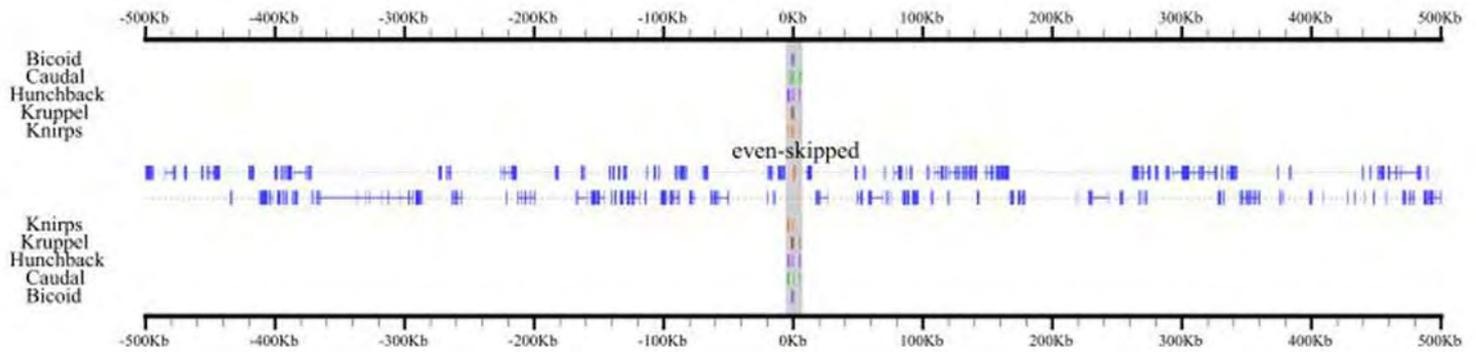


Figure 5. Un exemple des limites des prédictions bio-informatiques

### A. Distribution des prédictions de sites de fixation autour du locus *even-skipped* pour les FTs Bcd, Cd, Hb, Kr et Kni.

Les sites de fixation prédits sont représentés par des traits bleus pour Bicoid, verts pour Caudal, violets pour Hunchback, marrons pour Kruppel et oranges pour Knirps. Les prédictions pour le brin sens sont présentées au-dessus du locus (en bleu); celles du brin antisens en-dessous.

Une multitude de sites potentiels de fixation pour tous ces facteurs sont retrouvés autour du locus.

### B. Distribution des mêmes prédictions après filtrage par cluster de sites de fixation.

Une seule région correspond à un cluster d'au moins 12 sites sur une région de 700pb.

D'après Berman Nibu 2002

Crocker et al (Crocker and Erives 2008), montrant la présence de plusieurs séquences bien conservées tout au long du CRM, soutenant l'existence de contraintes pour l'architecture fonctionnelle de ce CRM. Il semble donc qu'*eve-2* contienne des éléments non encore caractérisés et contribuant sans doute à son activité. Ces différentes données, à tout le moins leur interprétation, ajoutent à la confusion et attisent des débats assez vifs dans la communauté sur la façon, billboard ou enhanceosome, dont seraient généralement construits les CRMs.

## I.4. L'apport des approches globales

### I.4.1 Le décodage du génome

Le séquençage du génome de *Drosophila melanogaster* (Adams, Celniker et al. 2000) a donné accès à l'ensemble des séquences non-codantes, une véritable mine pour étudier les mécanismes du contrôle de l'expression génique.

Théoriquement, la prédiction des sites de liaison pour un FT dont on connaît la séquence de reconnaissance apparaît assez simple, en recherchant les occurrences de motifs présentant un seuil d'homologie significatif avec la PWM. Plusieurs méthodes ont été développées : *e.g.*, Patser (Hertz and Stormo 1999), Match (Kel, Gossling et al. 2003), MotifLocator (Aerts, Van Loo et al. 2003), MotifScanner (Aerts, Van Loo et al. 2003), Matrix-Scan (Thomas-Chollier, Sand et al. 2008). Même si ces méthodes de PWM-scanning ont été sensiblement améliorées, elles restent peu efficaces pour identifier des CRMs. Elles génèrent un grand nombre de prédictions, principalement dues à la courte taille et les variations des sites de liaison des FTs (Fig 5).

Le séquençage du génome d'autres espèces de drosophile a permis des études comparatives systématiques (Bergman, Pfeiffer et al. 2002; Stark, Lin et al. 2007). Les prédictions de sites peuvent être filtrées par empreinte phylogénétique, *i.e.* de ne considérer que les motifs évolutivement conservés (Tagle, Koop et al. 1988; Wasserman, Palumbo et al.



2000). Des exemples de tels outils sont Consite (Sandelin, Wasserman et al. 2004) et rVista (Loots and Ovcharenko 2004), bien adaptés pour l'analyse du promoteur d'un gène particulier. En effet, pour limiter les faux positifs, la recherche est généralement restreinte aux régions proximales (Kheradpour, Stark et al. 2007; Stark, Lin et al. 2007).

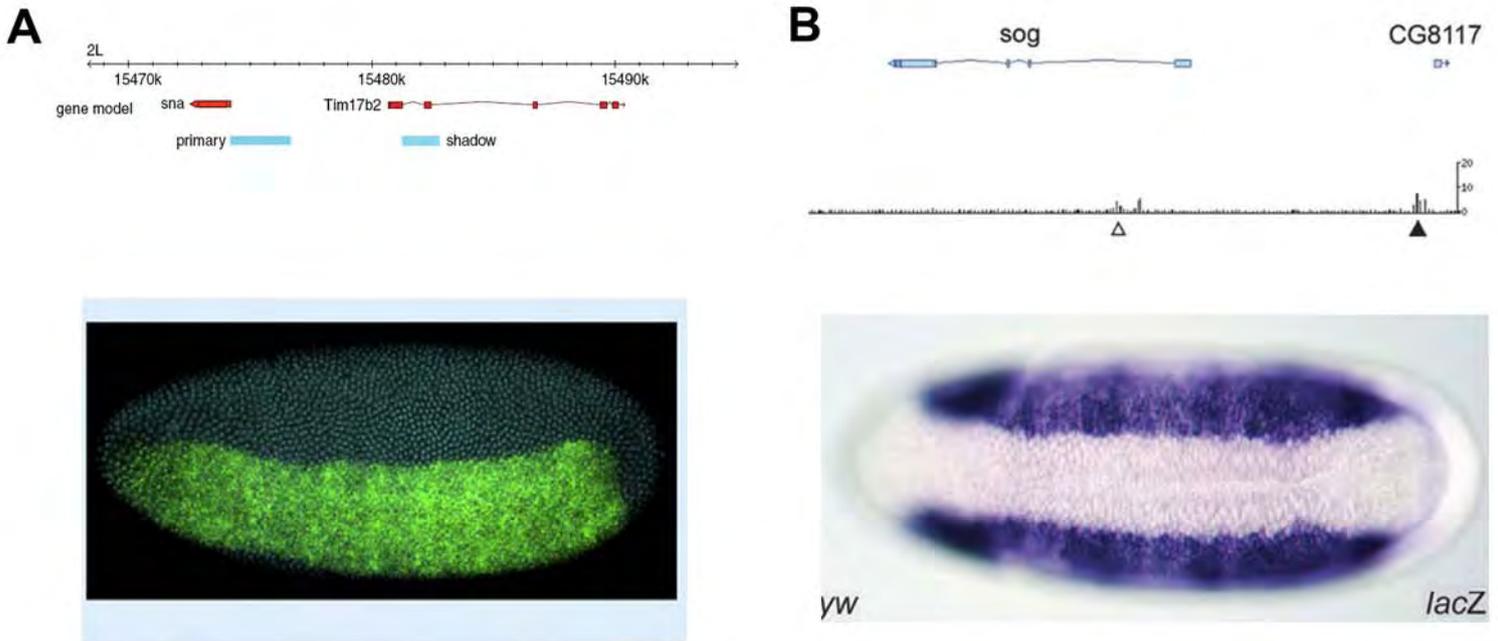
Une stratégie différente commence par la détection d'une région génomique évolutivement conservée, suivie d'une recherche de PWMs au sein de cette région, par exemple EvoPrinter (Odenwald, Rasband et al. 2005). De manière un peu similaire, des études ont recherché des mots d'ADN présentant une conservation évolutive globale (sans contrainte préalable d'alignement) supérieure à celle du génome, notamment à l'ensemble des mots de même taille (Pritsker, Liu et al. 2004; Elemento and Tavazoie 2005). Cette approche a permis de découvrir des motifs sur-conservés au cours de l'évolution, suggérant une pression résultant de leur fonction présumée, pouvant par exemple correspondre à des sites de liaison de FTs. Une limitation bien sûr de cette approche est qu'elle repose sur l'identification de mots exacts, et non des PWMs qui décrivent mieux les interactions ADN/protéine.

Suite notamment à la dissection fonctionnelle de CRMs du développement précoce, une notion répandue est qu'ils correspondent à l'accumulation locale (clusters) de sites de fixation pour des FTs distincts dans une courte région génomique (Davidson, Dogra et al. 2001). Cette idée a été appliquée avec succès pour identifier des CRM impliqués dans le développement de la drosophile. Par exemple, Berman et al. (Berman, Nibu et al. 2002) ont identifié de nouveaux CRMs cibles de Cad, Bcd, Kr, Kni et Hb en recherchant dans une fenêtre de 700bp au moins 15 sites de fixation (Fig 5). La recherche de cluster de motifs améliore donc la spécificité des prédictions de CRM, mais reste peu efficace à l'échelle du génome entier (Aerts, van Helden et al. 2007). La combinaison de recherche de clusters et de la conservation évolutive permet d'améliorer les prédictions. La conservation de clusters composés de différents motifs (hétérotypiques) entre *D. melanogaster* et *D. pseudoobscura*



permet de mieux prédire les CRMs, en les distinguant des régions négatives (Berman, Pfeiffer et al. 2004). Une étude réalisée sur 34 FTs de drosophile montre que la recherche de clusters de motifs homotypiques conservés permet d'améliorer sensiblement la prédiction de CRMs (Aerts, van Helden et al. 2007). Globalement l'efficacité de prédiction des CRM reste limitée, mettant en lumière le faible niveau de compréhension de l'architecture fonctionnelle des CRMs.

En effet la prédiction des CRMs reste complètement dépendante de la connaissance *a priori* des motifs fonctionnels qui les composent. Nous avons vu la pauvreté des données actuellement disponibles sur le répertoire des motifs de liaison pour l'ensemble des FTs d'une espèce. D'autre part, il n'est pas exclu que d'autres séquences que celles reconnues par des FTs jouent un rôle pour bâtir fonctionnellement un CRM (Khoueir, Rothbacher et al. 2010). Enfin, une fois un CRM cible d'un ou de plusieurs FTs identifié, comment découvrir l'ensemble des éléments cis-régulateurs qui le composent reste un challenge particulièrement difficile. La découverte du facteur Zelda illustre bien ce problème. Le motif « TAG » (CAGGTAG) a été initialement identifié en comparant les CRM reconnus par 21 FTs impliqués dans le développement précoce de la drosophile (Satija and Bradley 2012). Une observation étonnante est que différentes régions du génome semblent propices à la fixation *in vivo* de ces FTs, y compris quand ces régions ne contiennent pas de sites de fixation pour le FT considéré ! On a découvert ensuite le facteur Vielfaltig ou Zelda qui lie le motif « TAG » (ten Bosch, Benavides et al. 2006) et favoriserait le recrutement de différents FTs, probablement par interaction protéine/protéine. Ce facteur Zelda, qui jouerait un rôle déterminant et très général pour l'établissement de la segmentation, a néanmoins échappé à de très nombreux cribles génétiques de grande ampleur. De même, des études comparatives entre CRMs cibles du FT Dorsal montrent l'enrichissement et le rôle fonctionnel de deux motifs « orphelins », GCTGGAA et CACATG, qui participent à l'activité des CRMs (Stathopoulos, Van Drenth et al. 2002; Liberman and Stathopoulos 2009). On voit donc la



**Figure 6. Exemple de CRMs présentant une localisation atypique.**

**A. Cas d'un CRM de *snail*.**

Le locus *snail* est représenté en orange. Un CRM classique est représenté (primary), un second CRM localisé 7kb en amont est représenté (shadow). Une immunocoloration de ce CRM en amont d'un rapporteur *lacZ* montre le patron d'expression spécifique dirigé par cette région.

D'après Hong, Levine 2008.

**B. Cas d'un CRM de *sog*.**

Le locus *sog* est représenté en bleu. Un CRM classique est représenté par le triangle vide, le CRM localisé 20kb en aval est représenté par un triangle noir. Une immunocoloration du CRM shadow en amont d'un rapporteur *lacZ* montre le patron stéréotypé dirigé par ce CRM.

D'après Perry, Levine, 2010.

nécessité du va et vient entre démarches prédictives et caractérisation fonctionnelle des CRMs.

#### I.4.2 Le ChIP-seq

Le développement de la technique de Chromatine Immuno Precipitation (ChIP) a constitué une révolution en permettant d'identifier les interactions FT/ADN *in vivo* (Johnson, Mortazavi et al. 2007). Le séquençage à haut débit (Next Generation Sequencing ou NGS) a favorisé son application à l'échelle du génome entier et permis une meilleure résolution des sites de fixation (Park 2009).

Des études réalisées chez la drosophile par les laboratoires d'E. Furlong et M. Levine pour la formation du mésoderme embryonnaire (Sandmann, Jensen et al. 2006; Jakobsen, Braun et al. 2007; Sandmann, Girardot et al. 2007) et de la segmentation (Zeitlinger, Zinzen et al. 2007), ont conduit à un premier décryptage « genome-wide » des réseaux géniques impliqués, identifiant ainsi de nouvelles fonctions pour un FT donné (Zeitlinger, Zinzen et al. 2007). Le consortium modENCODE a analysé les profils de ChIP pour 30 FTs, à plusieurs stades de développement durant l'embryogenèse, ainsi que dans plusieurs lignées cellulaires (Roy, Ernst et al. 2010; Negre, Brown et al. 2011). Le projet BDTNP a analysé 21 FTs lors de la segmentation de blastoderme (MacArthur, Li et al. 2009), entamant ainsi une cartographie systématique des sites de liaisons des FTs sur le génome.

Ces études révèlent l'existence de CRMs apparemment redondants, dits « shadow enhancers », (Hong, Hendrix et al. 2008). Ils reproduisent tout ou partie du patron d'expression du CRM principal, présentent ou non les mêmes sites de fixation de FTs, et peuvent être très éloignés du démarrage de transcription (Fig 6). La conservation de clusters composés de différents motifs (hétérotypiques) entre *D. melanogaster* et *D. pseudoobscura* permet de mieux prédire les CRMs, en les distinguant des régions négatives (Cande, Goltsev et al. 2009).



La plupart des FT analysés par ChIP présentent un (très) grand nombre de sites de fixation, à une étape précise du développement: plus de 2000 pour Twist (Sandmann, Girardot et al. 2007), 670 pour Mef2 (Cunha, Sandmann et al. 2010), etc... Combien de ces sites de fixation correspondent à des CRMs fonctionnels reste difficile à évaluer. Plusieurs études suggèrent qu'une partie importante de ces régions ne contribuent pas au contrôle de la transcription, au moins aux étapes de développement analysées (Li, MacArthur et al. 2008; Fisher, Li et al. 2012). Il est donc important, pour analyser les données de ChIP d'avoir en même temps accès à l'ensemble des gènes dont l'expression dépend du facteur de transcription considéré.

#### I.4.3 L'analyse du transcriptome via les puces à ADN

Les puces à ADN ont été le premier moyen permettant de quantifier les variations du transcriptome, *i.e.* de déterminer l'ensemble des gènes dont l'expression est affectée par la perte de fonction, ou la surexpression, d'un FT (Schena, Shalon et al. 1995). Par exemple, chez la drosophile la comparaison du transcriptome de disques imaginaux de patte, entre condition sauvage et suite à l'expression forcée du gène *eyeless* a permis d'identifier des acteurs du développement de l'œil (Michaut, Flister et al. 2003). Le NGS permet aujourd'hui des approches alternatives, comme le RNAseq, qui peut aussi identifier de nouvelles unités de transcription (Wang, Gerstein et al. 2009). Cependant une limitation importante à l'analyse du transcriptome est qu'elle cumule des effets directs (*e.g.*, les cibles d'un FT dont on étudie les conséquences de son inactivation) et indirects, les variations d'expression observées pouvant découler d'une cascade d'évènements moléculaires successifs.



En résumé, nous venons de voir l'enjeu que représente l'élucidation des mécanismes intimes de régulation de la transcription au cours du développement. Il s'agit en particulier d'identifier le répertoire complet des gènes effecteurs de la différenciation terminale. Dans une deuxième étape, il s'agit de cartographier les éléments cis-régulateurs associés. Ainsi on pourra comparer leur architecture fonctionnelle aux modèles actuels, issus de l'analyse de l'embryogenèse précoce, pour mieux comprendre le contrôle de l'expression du génome pour les différentes étapes du développement.

Pour aborder ces questions, j'ai donc voulu identifier à l'échelle du génome entier les effecteurs cellulaires, et décrypter leur code de régulation, sous-tendant la morphogénèse épidermique chez *Drosophila melanogaster* en fin de développement embryonnaire.

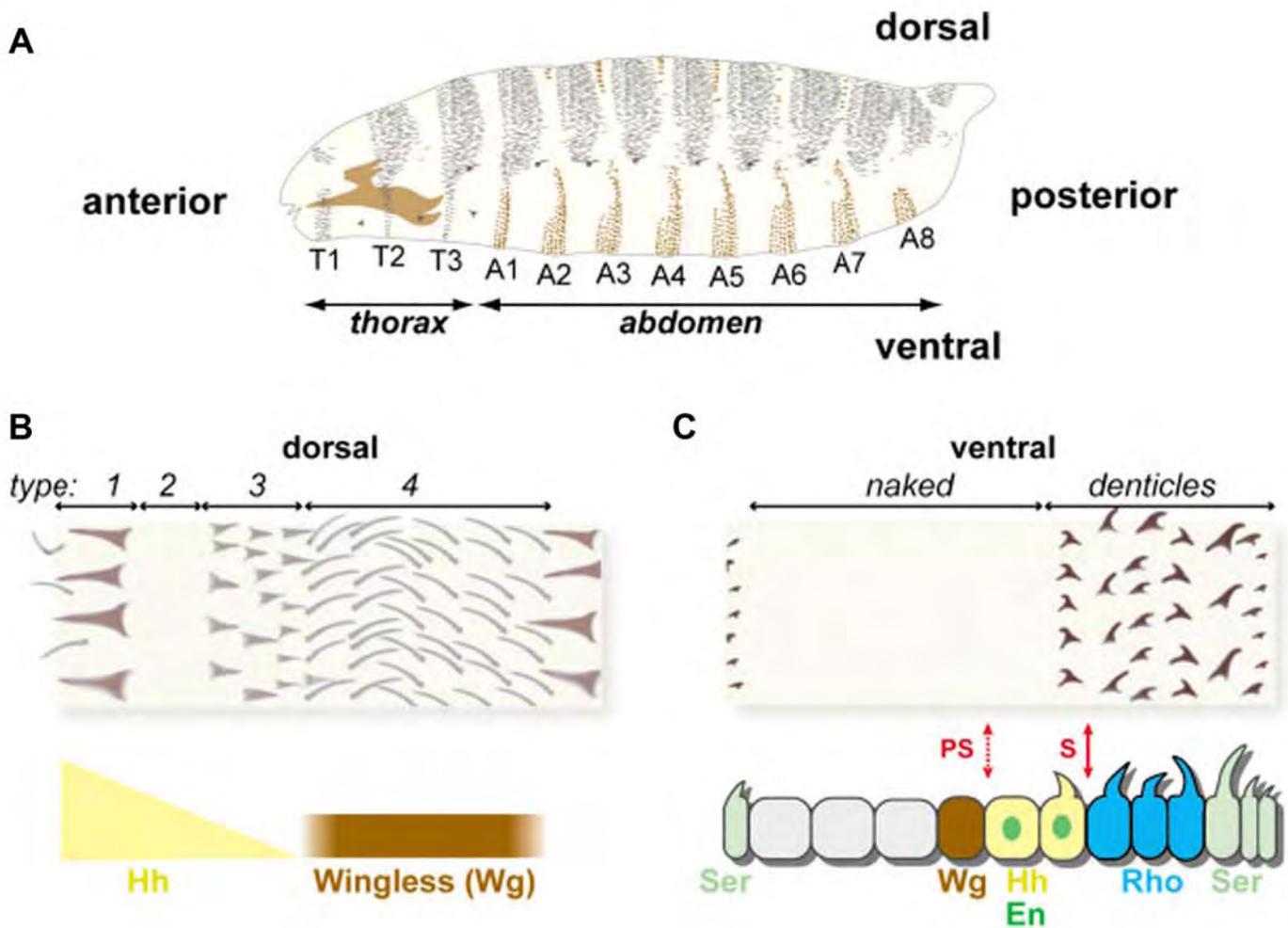


Figure 7. L'organisation stéréotypée des cellules épidermiques.

**A. Représentation schématique d'une larve.**

L'épiderme présente une organisation stéréotypée avec alternance de cellules lisses et de cellules présentant des extensions cytoplasmiques.

**B. Mise en place des trichomes dorsaux.**

Les trichomes de types 1 à 3 sont définis par un gradient de Hedgehog (Hh), les trichomes de type 4 sont définis par un gradient de Wingless (Wg).

**C. Mise en place des denticules ventrales.**

Wingless (Wg) définit les cellules lisses alors que des combinaisons des facteurs Hedgehog (Hh), Rhomboid (Rho), Engrailed (En) et Serrate (Ser) définissent les différents types de denticules.

D'après Payre 2004.

## II. La morphogénèse épidermique, modèle d'étude de la spécificité

### tissulaire d'un FT

Les mécanismes qui dirigent le remodelage de la forme tridimensionnelle des cellules, la morphogénèse, restent mal compris. On ignore le répertoire complet des différents effecteurs impliqués. Si des études se focalisent sur des protéines réputées clés, par exemple petites GTPases, Cadherin, Myosin etc..., les mécanismes de leur régulation sont principalement post-traductionnels. L'importance des FTs dans la différenciation terminale suggère cependant la nécessité de contrôler l'expression du génome, y compris pour la morphogénèse. Les mécanismes de régulation génique responsables de la morphogénèse restent donc à découvrir. Pour aborder ce problème, notre laboratoire étudie les cellules de l'épiderme embryonnaire de drosophile comme modèle *in vivo* de la morphogénèse.

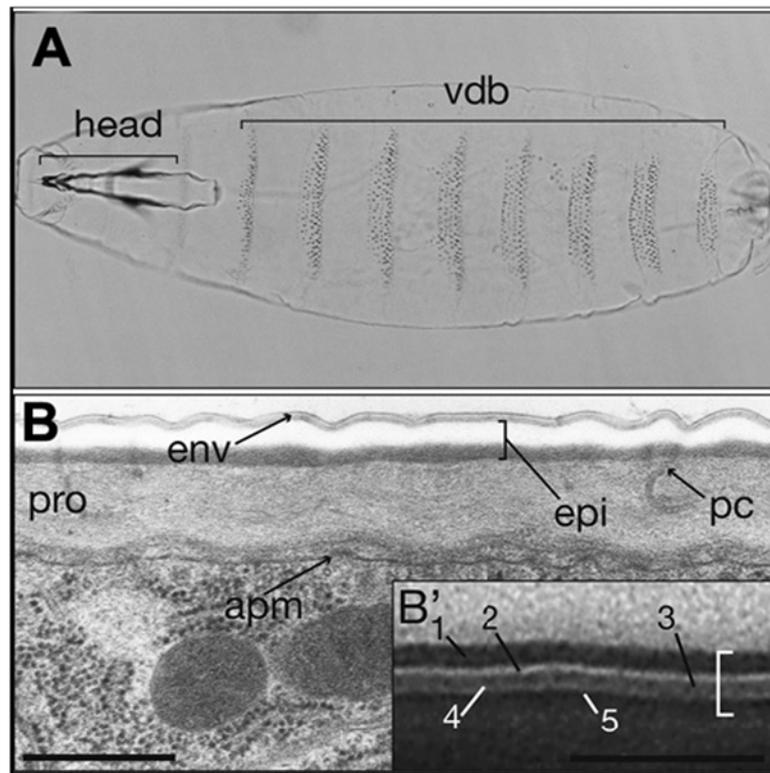
### II.1. Mise en place du tissu épidermique au cours du développement

Le tissu épidermique dérive du neurectoderme bilatéral et va recouvrir tout l'embryon suite à la fermeture dorsale. Il s'agit d'un mouvement coordonné de migration des cellules latéro-dorsales. Ces cellules s'aplatissent et s'allongent selon l'axe dorso-ventral, pour recouvrir progressivement l'amnioséreuse, sans division ni intercalation cellulaire. Les parties gauches et droites de l'épiderme fusionnent ensuite dans la région la plus dorsale.

L'épiderme est composé d'une monocouche de cellules post-mitotiques. Dans chaque segment, l'épiderme présente l'alternance stéréotypée de cellules lisses et de cellules avec des extensions cytoplasmiques apicales, les trichomes (Fig 7A). Nous verrons dans la suite une description plus détaillée de leur organisation.

#### II.1.1. La cuticule

Les trichomes sont en réalité des structures spécialisées de l'exosquelette des insectes, la cuticule. La cuticule recouvre tous les tissus en contact avec le milieu extérieur: l'épiderme,



**Figure 8. Structure de la cuticule larvaire chez *Drosophila melanogaster*.**

**A. Cuticule larvaire en vue ventrale.**

L'ensemble de la larve est couverte d'une cuticule qui est plus épaisse au niveau de la tête et des denticules (vdb).

**B. Coupe transversale d'une cuticule et analyse en microscopie électronique.**

La procuticule (pro) est liée à la membrane plasmique des cellules épithéliales (apm).

L'épicuticule (épi) la recouvre, puis l'enveloppe (env) est déposée.

D'après Moussian 2010.

mais aussi les régions antérieure et postérieure de l'intestin et le système respiratoire. La cuticule protège contre les stress mécaniques, la déshydratation, l'agression par des microorganismes pathogènes et les prédateurs, et participe à la locomotion. La cuticule est une structure extrêmement complexe, et on admet généralement qu'elle contribue sans doute à l'extraordinaire succès évolutif des insectes.

*a) Différentes couches superposées forment la cuticule*

Dès le milieu du 19<sup>ème</sup> siècle, on s'est attaché à analyser la cuticule, notamment sa composition, organisation, sécrétion, etc... (voir (Locke 2001) pour revue). La cuticule contiendrait des centaines de constituants différents. La majeure partie de la cuticule est composée de chitine, un polymère de N-acetyl Glucosamine, le deuxième polysaccharide (en masse, derrière la cellulose) à la surface du globe. La cuticule contient aussi une très grande variété de protéines, ainsi que différents lipides et cires (Locke, Kiss et al. 1994). La composition exacte, dont on sait qu'il existe des variations importantes (Fig 8A) (par exemple entre les régions flexibles ou rigides, opaques ou transparentes, etc...), de la cuticule reste mal connue. L'étude structurale en microscopie électronique permet de distinguer plusieurs couches successives (Fig 8B). Schématiquement, les différentes couches de cuticule se mettent en place séquentiellement, mais il est bien sur possible que leur déposition soit plus intimement coordonnée.

La première couche, synthétisée dès le stade 15 chez l'embryon de drosophile, est l'enveloppe, riche en lipides, qui va constituer une barrière vis à vis du milieu extérieur (Fig 8B). Son dépôt n'est pas continu, on observe d'abord sa déposition préférentielle au niveau des trichomes, pour finalement recouvrir l'ensemble de l'épiderme au stade 16.

L'épicuticule majoritairement composée de protéines est déposée sous l'enveloppe (Fig 8B). Elle contient divers composants structuraux mais aussi enzymatiques, notamment



responsables de la formation de quinones impliquées dans la rigidification et la pigmentation de la cuticule (Locke and Krishnan 1971).

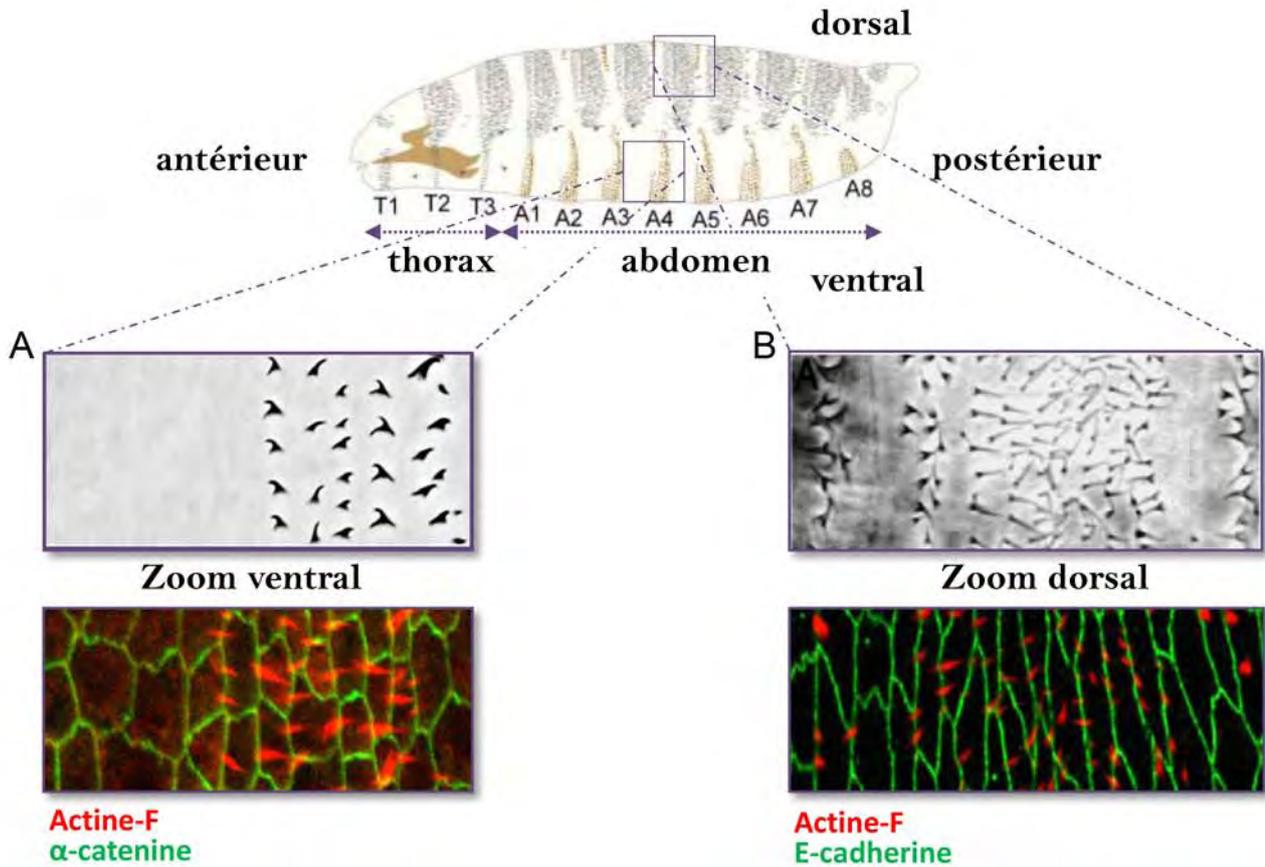
La dernière couche sécrétée par les cellules épidermiques est la procuticule (Fig 8B). En contact direct avec la membrane plasmique, elle est composée de feuillets successifs de chitine, dont le dépôt avec un angle constant entre deux couches conduit à un arrangement hélicoïdal. La procuticule s'étoffe au stade 17, puis les trois couches continuent de s'épaissir jusqu'à éclosion de la larve, 6h plus tard. La dépense métabolique des cellules de l'épiderme pour la formation de la cuticule est très importante, on présente ainsi souvent l'épiderme embryonnaire comme un véritable organe de sécrétion.

#### *b) De nombreuses protéines composent la cuticule*

La base de données Flybase répertorie plus de 300 gènes codant des protéines en lien avec la cuticule, dont celles contenant des motifs d'interaction avec la chitine. Il existe toute une variété de protéines mal connues, qui seraient impliquées dans la formation, maintien, modification, etc..., de la cuticule. Ces gènes sont souvent regroupés en familles multigéniques, qui peuvent évoluer rapidement entre espèces d'insectes.

La famille des Larval Cuticle Protein (LCP) contient 18 gènes regroupés en clusters (Snyder, Hirsh et al. 1981), et ces protéines LCP présentent un domaine R&R, réputé fixer la chitine (Rebers and Willis 2001).

La famille Tweedle (Twdl) regroupe 27 gènes regroupés en trois clusters génomiques. Leur expression commence chez l'embryon au stade 13, dans les dérivés de l'ectoderme (trachées, épiderme, intestin), et se maintient jusqu'au premier stade larvaire. Les mutants pour *TwdlD* présentent une taille réduite, avec une forme générale anormalement arrondie (Guan, Middlebrooks et al. 2006). On a pu montrer que ces protéines Twdl sont incorporées dans la cuticule larvaire, dans des territoires distincts correspondant au registre d'expression des différents gènes analysés (Guan, Middlebrooks et al. 2006).



### Figure 9. Organisation des trichomes.

Des zooms de cuticules de larves L1 sont montrés au niveau ventral (A) et dorsal (B), ainsi que des immunocolorations marquant les contours cellulaires (alpha-caténine ou DE-cadherine en vert) et les faisceaux d'actine (phalloïdine en rouge).

#### A. Les denticules.

Au stade larvaire les denticules de chaque rangée présentent une orientation particulière (antérieure pour les rangées 1 et 4, postérieure pour les rangées 2, 3 et 5). Les denticules lors de leur formation sont d'abord toutes orientés vers la région postérieure. Toutes les extensions sont localisées en région postérieure de la cellule.

#### B. Les trichomes.

Les trichomes dorsaux présentent une organisation stéréotypée. Ceux de la rangée 1 sont orientés vers l'antérieur, alors que ceux de la rangée 3 ont orientés vers le postérieur. Les 3 premières rangées de type 3 pointent vers le postérieur, puis les 3 dernières vers l'antérieur.

La famille la plus abondante des gènes de cuticule est cependant la famille des Cpr (Cuticule protein related), contenant 102 gènes chez la drosophile. Ils codent des protéines capables de lier la chitine *in vitro* via le domaine R&R (Rebers and Willis 2001; Togawa, Nakato et al. 2004), mais là encore leur fonction reste mal connue.

## II.1.2. Développement des cellules épidermiques

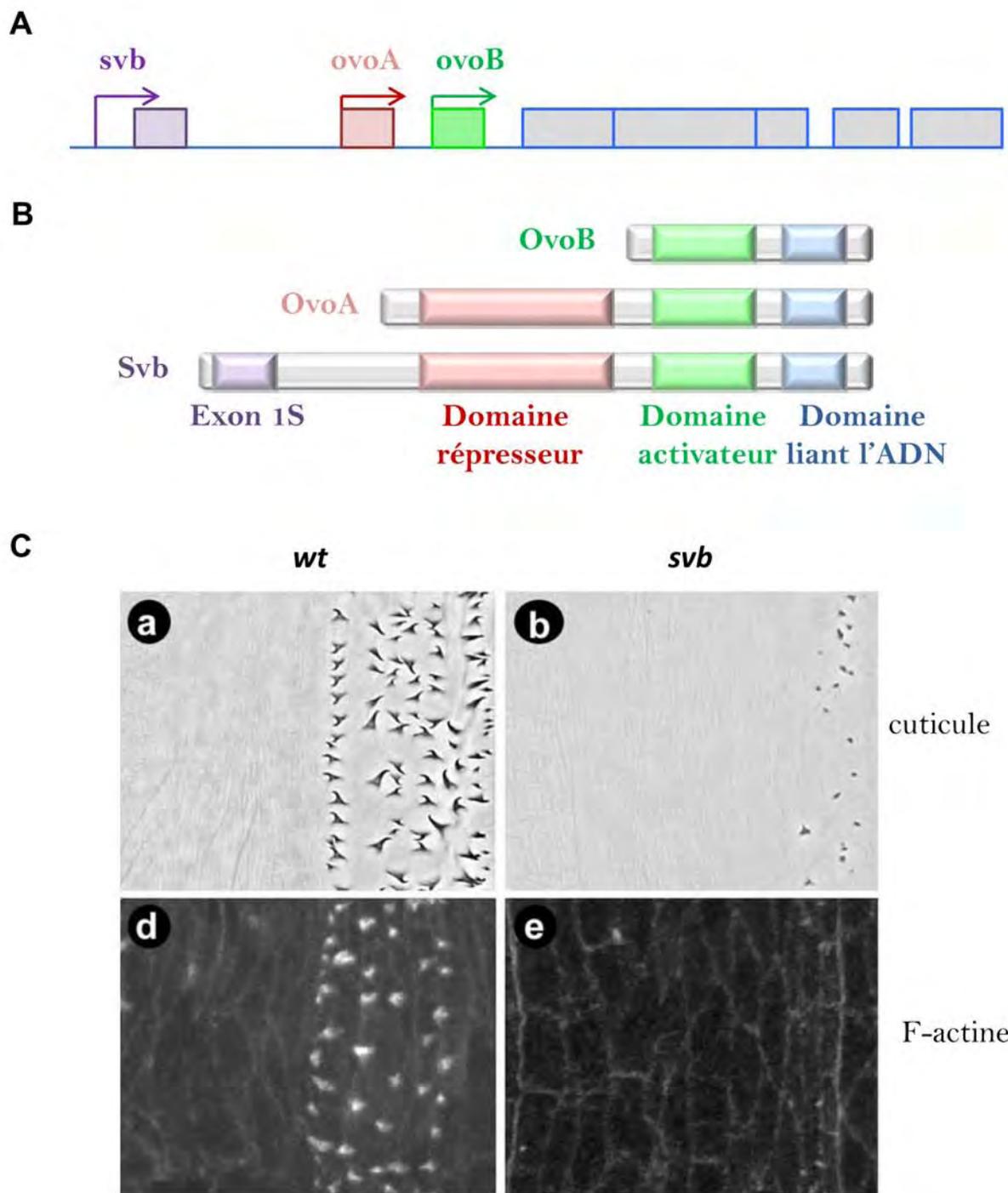
### a) *Différentiation des cellules épidermiques*

La spécification des cellules épidermiques entre le destin lisse ou à trichome s'effectue au sein de chaque segment de l'embryon, de façon différente entre ventre et le dos.

En milieu d'embryogénèse (stade 14), les cellules épidermiques présentent toutes une morphologie similaire, avec une surface apicale lisse et un allongement selon l'axe apico-basal. A ce stade, les microfilaments sont concentrés aux jonctions cellulaires. Les cellules qui vont former les trichomes commencent alors à changer de forme : elles s'allongent selon l'axe dorso-ventral et s'alignent en rangées régulières, donnant l'apparence de barreaux d'échelle (Price, Roberts et al. 2006). Les filaments d'actine s'accumulent au cortex apical, d'abord uniformément, puis à la marge postérieure de la cellule (Price, Roberts et al. 2006). Plusieurs foci d'actine apparaissent, se regroupant en général en 1-3 faisceaux, qui croissent perpendiculairement à la face apicale (Dickinson and Thatcher 1997; Price, Roberts et al. 2006). Dès le stade 15, les extensions d'actine préfigurent le patron complet des trichomes (Fig 9A,B). A partir du stade 16, on observe un désassemblage des faisceaux d'actine ; la cuticule conservant l'empreinte des extensions cellulaire pour donner les denticules et trichomes dorsaux matures.

### b) *Morphologie des cellules épidermiques à trichomes*

Les trichomes ont une morphologie différente entre le ventre et le dos. Dans le ventre, malgré des différences de taille ou d'orientation, les denticules présentent une morphologie en forme de crochet et sont fortement pigmentés. Chaque segment abdominal



**Figure 10. Locus *Ovo/svb*, FTs et phénotypes associés.**

**A. Organisation moléculaire du locus *Ovo/svb*.**

Les transcrits sont issus de promoteurs indépendants. *Svb* est issu d'un promoteur somatique alors que *OvoA* et *OvoB* sont issus de promoteurs germinaux. Les exons communs sont représentés en gris et les exons spécifiques en vert pour *OvoB*, rouge pour *OvoA* et violet pour *svb*.

**B. Organisation des protéines *Ovo* et *Svb*.**

Le domaine de liaison à l'ADN est représenté en bleu, le domaine activateur en vert et le domaine répresseur en rouge. L'exon 1S spécifique de *Svb* est représenté en violet.

**C. Phénotype dû à l'absence de *svb*.**

La mutation *svb* entraîne une absence de denticules corrélée avec une absence de faisceaux d'actine au niveau apical.

D'après Delon et al.; 2003.

contient 6 à 7 rangées de trichomes, et 6 rangées de cellules lisses (Fig 7C). Les denticules des rangées 1 et 4 pointent vers l'avant, celles des rangées 2, 3 et 5 vers l'arrière (Fig 7C, 9A).

Les trichomes dorsaux sont plus fins, sans crochet et peu ou pas pigmentés. L'épiderme dorsal de chaque segment abdominal peut être sous divisé en 4 régions. Le type 1 correspond à une seule rangée de cellules présentant de larges extensions pigmentées, orientées vers l'avant. Le type 2 correspond à une rangée de cellules lisses, alors que le type 3 est composé de 3 rangées de cellules avec des extensions de taille intermédiaire, orientées vers l'arrière. Le type 4 est composé de 6-7 rangées cellulaires présentant des extensions apicales très fines et souples. Les extensions des 3 rangées les plus avant sont orientées vers l'arrière, alors que les extensions des 3 ou 4 rangées les plus arrière sont orientées vers l'avant (Fig 7B, 9B).

## II.2 Le couple *Svb/Pri*, déterminants du patron des trichomes

L'analyse de la spécification des destins épidermiques démontre qu'ils sont sous le contrôle des voies de signalisation Wg et EGF-R, en interaction avec Notch, Hedgehog (Fig 7, Fig 11). Un des résultats fondateur de l'équipe a été l'identification de l'intégrateur de ces différentes voies de signalisation : le facteur de transcription Shavenbaby (*Svb*).

### II.2.1 *Svb*, régulateur transcriptionnel de la morphogénèse

Les mutations *svb* ont été initialement isolées dans l'un des premiers cribles génétiques réalisés chez la drosophile (Nusslein-Volhard and Wieschaus 1980). L'inactivation de *svb* entraîne une réduction massive des trichomes, tant dans la région ventrale que dorsale (Fig 10C).



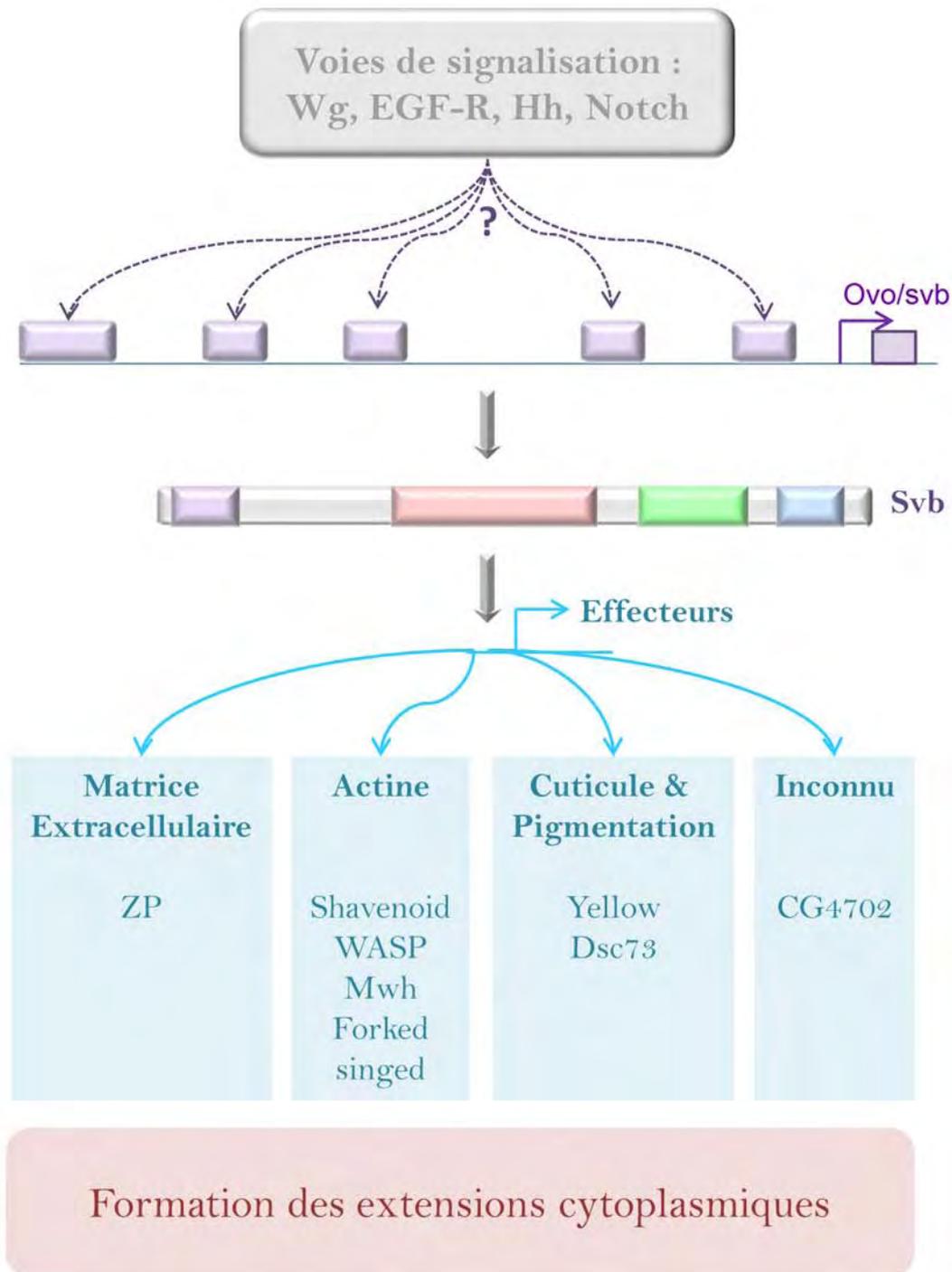
a) *Le locus Ovo/svb*

*svb* code un FT à doigt de zinc dont l'expression épidermique est spécifique des cellules à trichomes (Garfinkel, Wang et al. 1994; Mevel-Ninio, Terracol et al. 1995; Delon, Chanut-Delalande et al. 2003). Il fait partie du locus *ovo/svb*, qui dirige en outre la formation de la lignée germinale femelle, via les isoformes Ovo A & B, des versions tronquées dans la région N-terminale de la protéine Svb (Fig 10 A,B). Les 3 isoformes partagent la même région C-terminale, avec le domaine de liaison à l'ADN ainsi qu'un domaine activateur de transcription (Fig 10B). Ces trois isoformes sont produites chacune à partir d'un promoteur différent, deux promoteurs germinaux pour OvoA et OvoB, et le promoteur somatique, responsable de la production de Svb, notamment dans les cellules à trichomes (Fig 10A). OvoB agit comme activateur de transcription. Par contre OvoA est un répresseur de transcription, grâce à la présence d'un domaine N-terminal additionnel (Andrews, Garcia-Estefania et al. 2000).

L'étude des protéines Ovo a identifié leur séquence de fixation *in vitro*. On a d'abord découvert 8 sites de fixation, partageant le motif « cœur » CnGTT (Lu, Andrews et al. 1998). Des approches additionnelles de type SELEX ont ensuite précisé ce motif sur une vingtaine de séquences (Lee and Garfinkel 2000; Bielinska, Lu et al. 2005) aboutissant à la PWM définie actuellement pour les FT Ovo/Svb : aCnGTTa.

b) *Svb, nécessaire et suffisant pour la morphogénèse épidermique*

La comparaison de l'activité des différentes formes OvoA, OvoB et Svb a permis de définir la fonction transcriptionnelle de Svb. L'expression artificielle d'OvoB (activateur) dans des cellules lisses suffit à forcer la formation de trichomes, comme l'expression ectopique de Svb (Payre, Vincent et al. 1999). Au contraire, l'expression d'OvoA (répresseur) dans les cellules à denticules les transforme en cellules lisses (Payre, Vincent et al. 1999). Svb agit donc comme activateur de la transcription pour la formation des trichomes. D'autre



**Figure 11. Réseau génique impliqué dans la morphogénèse gouvernée par Svb.**

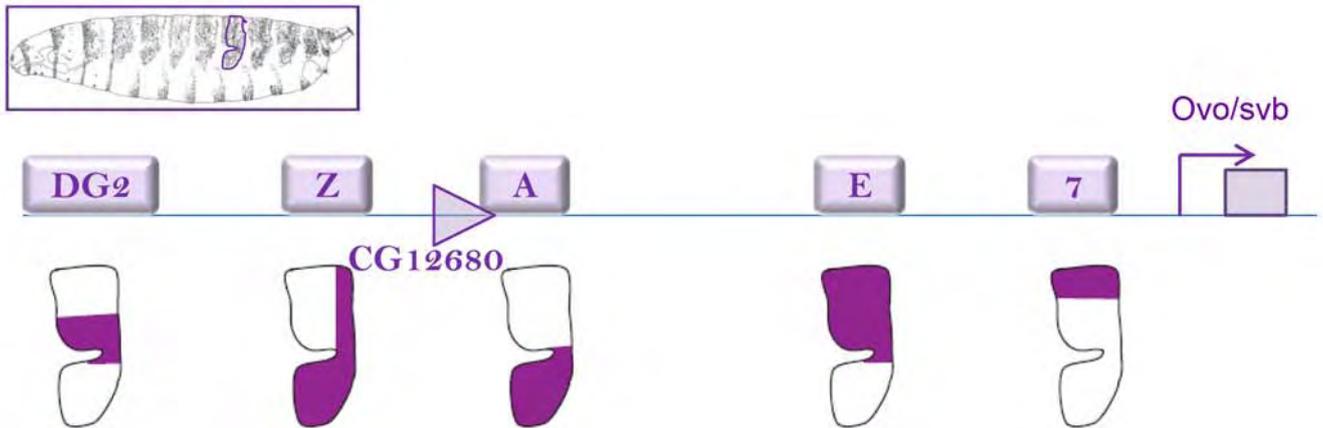
Des signaux issus des voies de signalisation Wingless, EGF-R, Notch et Hedgehog (Hh) sont intégrés par les régions régulatrices de *svb*, entraînant son expression spécifiquement dans les cellules à trichomes. Svb agit comme activateur de la transcription de différentes classes d'effecteurs.

part, quand les isoformes *Ovo* sont exprimés de manière artificielle dans l'épiderme (elles sont normalement restreintes à la lignée germinale), elles sont capables de reconnaître les mêmes effecteurs que *Svb*, pour réprimer (*OvoA*) ou activer (*OvoB*) leur expression (Chanut-Delalande, Fernandes et al. 2006). L'ensemble de ces données montre que *Svb* est nécessaire et suffisant pour induire la formation des trichomes, et suggère que les 3 isoformes partagent la même PWM.

*Svb* déclenche la réorganisation de l'actine lors de la mise en place des trichomes (Delon, Chanut-Delalande et al. 2003). *Svb* est également exprimé dans l'épiderme adulte, et son absence empêche là encore la formation des trichomes du thorax, aile, abdomen et antennes (Delon, Chanut-Delalande et al. 2003). Dans ces tissus adultes, la surexpression de *Svb* (ou *OvoB*) suffit à induire des extensions ectopiques, alors que l'expression d'*OvoA* empêche la formation des trichomes. L'expression de *Svb* dans un tissu où il n'est pas naturellement exprimé comme l'œil suffit également à induire la formation d'extensions anormales (Delon, Chanut-Delalande et al. 2003). Collectivement ces données soutiennent que dans différents tissus épidermiques, *svb* est nécessaire et suffisant pour induire la formation des trichomes, via la réorganisation de l'actine corticale.

### *c) Svb, intégrateur de voies de signalisation antagonistes*

De nombreux travaux montrent que l'activation de la voie Wnt par le ligand Wingless (*Wg*) détermine la formation des cellules épidermiques lisses. L'inactivation de *wg* entraîne la formation de denticules ectopiques, alors que la sur-activation de la voie augmente l'étendue de cuticule nue. En revanche, la voie EGF-R promeut la formation de denticules, et une forme dominante négative du récepteur à l'EGF induit la transformation des denticules en cellules lisses. Les travaux de l'équipe ont montré que l'action antagoniste de ces deux voies implique leur action contraire sur l'expression de *svb* (Payre, Vincent et al.



**Figure 12. Les régions régulatrices de *svb*.**

Le locus *Ovo/svb* est représenté en violet. Les 5 CRMs associés à *svb* sont représentés par des rectangles violets. Pour chaque région, un schéma représente le patron produit par le CRM en amont d'un rapporteur *lacZ*, après immuno-coloration.

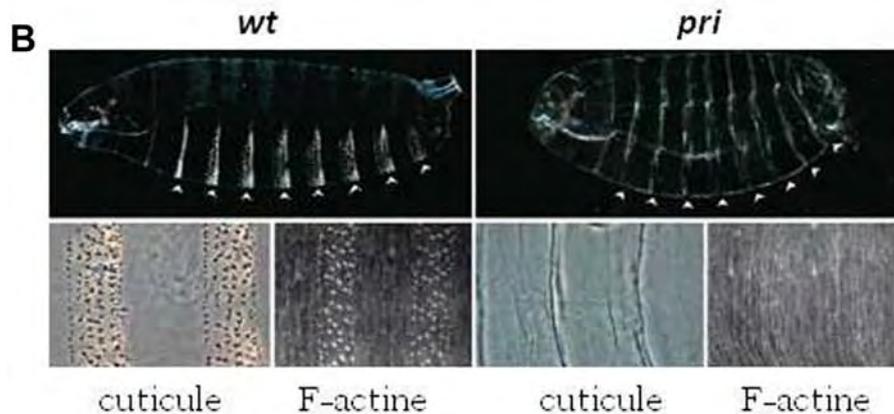
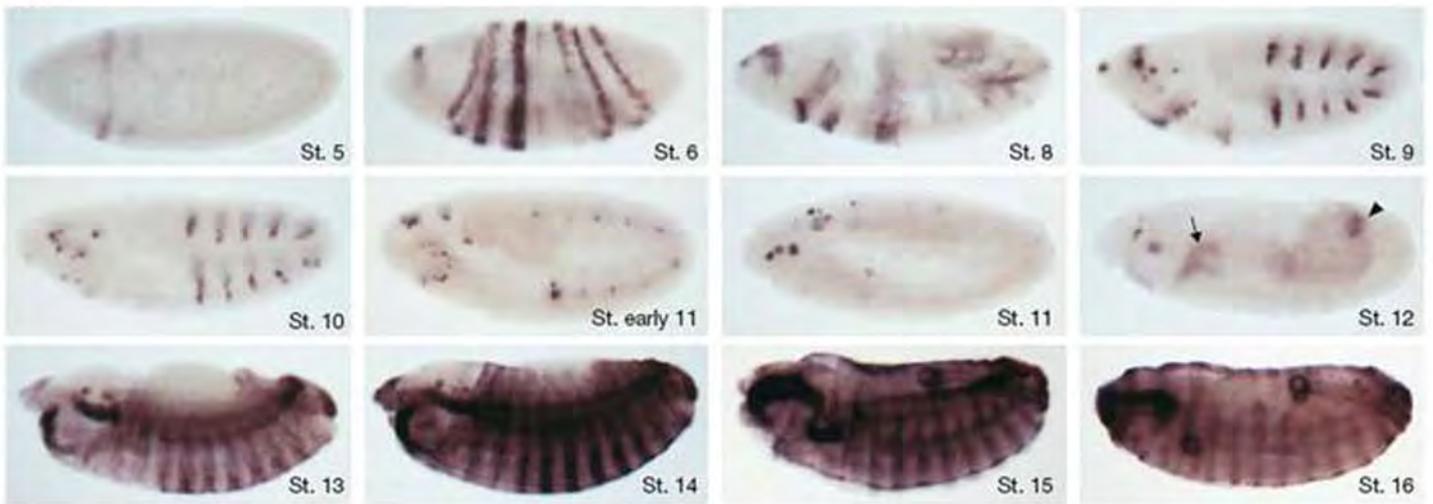
Les 5 CRMs ont des patrons partiellement redondants. Le CRM 7 dirige une expression dorsale, la région E dirige une expression dorso-latérale et la région 1 dirige une expression ventrale. Les shadow enhancers Z et DG2 reproduisent partiellement le patron des CRMs classiques.  
D'après Frankel, 2010.

1999). La voie Wg va ainsi inhiber l'expression de *svb*, pour la formation des cellules lisses. L'activation de la voie EGF-R induit l'expression de *svb* pour la formation des denticules.

Svb apparait donc comme l'intégrateur des voies de signalisation Wg et EGF-R antagonistes de la morphogénèse épidermique. Ces voies déterminent donc le destin des différentes cellules épidermiques via l'activation ou la répression de l'expression de *svb* (Fig 11).

L'analyse du locus *ovo/svb* par transgénèse systématique a identifié les régions régulant sa transcription dans l'épiderme embryonnaire. Trois régions principales dirigent une expression dans des sous-populations partiellement chevauchantes des cellules à trichomes (McGregor, Orgogozo et al. 2007) (Fig 12). La région 7 ou proximale dirige l'expression de *svb* dans 4 rangées dorsales de l'embryon (type 1 et 3) et dans les cellules ventrales à denticules. La région médiale E dirige une expression dorso-latérale (type 4) ainsi que dans un sous ensemble des cellules ventrales. La région distale A dirige essentiellement une expression latérale (type 4). Collectivement ces trois CRMs récapitulent le patron d'expression de *svb* dans l'épiderme (McGregor, Orgogozo et al. 2007) (Fig 12). L'évolution du patron d'expression de *svb* observé dans d'autres espèces de drosophile est due à l'accumulation de mutations dans chacun de ces CRM dirigeant l'expression épidermique de *svb*. Ensemble, ces différentes modifications sont responsables de l'évolution morphologique du patron des trichomes (Frankel, Davis et al. 2010; Frankel, Erezylmaz et al. 2011). Ces résultats démontrent le rôle de la modification des CRMs dans l'évolution d'un caractère morphologique.

Nous avons vu que certains gènes importants pour le développement présentent des CRMs secondaires ou « shadow » éloignés du gène et reproduisant le patron des CRMs primaires. Il existe au moins 2 CRMs shadow pour *svb*. Z reproduit partiellement le patron des CRMs proximal médial et distal, notamment pour les trichomes de type 4. DG2 dirige

**A**

**Figure 13. MRE29, ou *pri* (*polished rice*), un nouveau régulateur de la morphogénèse épidermique.**

**A. Expression de *pri* au cours du développement chez *Drosophila melanogaster*.**

*pri* est exprimé en bandes entre les stades 5 et 8. Une expression dans les placodes trachéales et la tête apparaît au stade 9. L'expression disparaît au stade 11-12. Aux stades tardifs une expression est visible dans l'épiderme, les trachées, et l'intestin.

**B. Phénotype des larves *pri*-.**

Cuticule et immuno-coloration des faisceaux d'actine d'embryons *wt* et *pri*-. Des larves *pri*- présentent une absence de denticules liée à un défaut de formation des faisceaux d'actine.

D'après Kondo et al.; 2007.

une expression réduite dans un sous-domaine du CRM E (Frankel, Davis et al. 2010) (Fig 12). Ces CRMs assurent la robustesse phénotypique du patron des trichomes face à des variations environnementales ou génétiques (Frankel, Davis et al. 2010).

## II.2.2 Pri, un régulateur inattendu de la morphogénèse

### a) Pri est exprimé à différentes étapes du développement

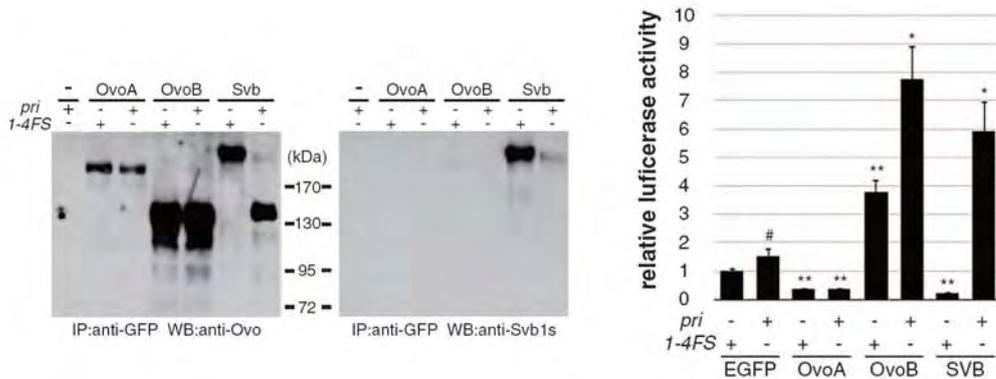
Svb est donc le FT contrôlant la morphogénèse des trichomes épidermiques chez la drosophile. Comment ce FT qui contient le domaine de répression présent dans OvoA peut-il néanmoins agir comme un activateur de transcription ?

L'identification d'un régulateur inattendu de la morphogénèse épidermique a permis de lever le voile. La recherche de nouveaux longs ARNs non-codants (lncARN) (Inagaki, Numata et al. 2005) a montré l'expression embryonnaire, avec un profil très dynamique, du candidat MRE29, rebaptisé plus tard *polished-rice* (*pri*) suite à l'analyse de sa fonction. L'ARN *polished-rice* est d'abord exprimé dans 7 bandes antéropostérieures au stade blastoderme, puis dans des clusters bilatéraux de cellules aux stades 9 et 10 (Kondo, Hashimoto et al. 2007) (Fig 13 A). A partir du stade 12-13, *pri* s'exprime principalement dans les spiracles, l'intestin, les trachées et les cellules épidermiques (Kondo, Hashimoto et al. 2007) (Fig13 A). Des embryons mutants *pri* présentent des défauts épidermiques semblables à ceux des mutants *svb*, c'est-à-dire une absence des trichomes ventraux et dorsaux (Fig13 B), un phénotype à l'origine du nom *polished-rice* (Kondo, Hashimoto et al. 2007). On observe aussi des défauts des trachées, avec notamment l'absence des *taenidia*, des anneaux riches en filaments actine qui soutiennent la lumière des trachées (Kondo, Hashimoto et al. 2007).

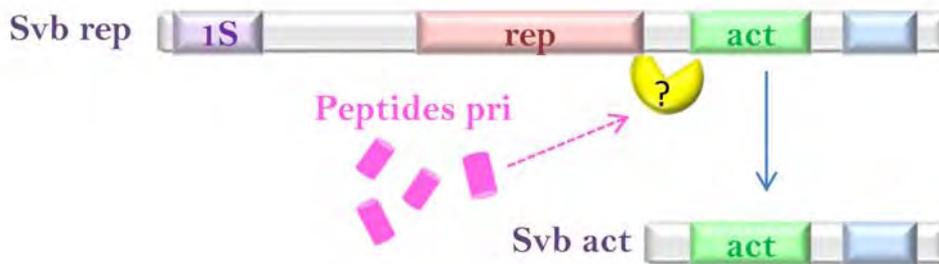
**A**

Pri\_ORF1 **MAAYLDPTGQY**  
 Pri\_ORF2 **MAAYLDPTGQY**  
 Pri\_ORF3 **MSHDLDPTGTY**  
 Pri\_ORF4 **M-----LDPTGTYRRPRDTQDSRQKRRQDCLDPTGQY**

**B**



**C**



**Figure 14. Structure des peptides Pri et mode d'action sur Svb.**

**A. Alignement des 4 ORFs redondantes de l'ARNm pri.**

Chaque ORF contient un motif LDPTG(Q/T)Y conservé et représenté en rouge.

**B. Les peptides Pri induisent un clivage de la protéine Svb qui devient activatrice.**

Western blots montrant l'effet de Pri sur les protéines OvoA, OvoB et Svb. Alors que la présence de Pri induit un clivage de Svb, les protéines OvoA et OvoB ne sont pas sensibles aux peptides, montrant que l'effet des peptides passerait par l'exon1S spécifique de Svb.

Essai luciférase de l'activité des protéines OvoA, OvoB et Svb en présence ou absence des peptides Pri. Seul Svb devient activatrice de la transcription en présence de Pri. Les activités répressive d'OvoA et activatrice d'OvoB ne sont pas affectées par Pri.

D'après Kondo et al.; 2010.

**C. Schéma du mode d'action des peptides Pri (roses) sur la protéine Svb.**

Svb est d'abord présente sous forme longue répressive de la transcription. La présence des peptides Pri induit un clivage de la région N-terminale de Svb qui perd son domaine répressif. Svb devient activateur de la transcription.

### b) *Pri*, régulateur de la protéine *Svb*

L'ARNm de *pri* ne contient que 5 courtes ORFs, présentant une organisation polycistronique. Différentes études ont finalement démontré que les 4 premières ORF, codant des peptides de 11 à 32 acides aminés (Fig 14 A), sont nécessaires et suffisantes pour les fonctions embryonnaires de *pri*. Ces quatre peptides, ou microprotéines, possèdent un motif C-terminal commun et agissent de manière redondante. L'expression d'un seul peptide *pri* à partir d'un ARN entièrement synthétique est ainsi suffisante pour assurer la fonction *pri*, au moins dans un système cellulaire reconstitué (J. Zanet). Comment ces peptides Pri peuvent-ils contrôler la formation des trichomes, et quelle est leur interaction avec le FT *Svb* ?

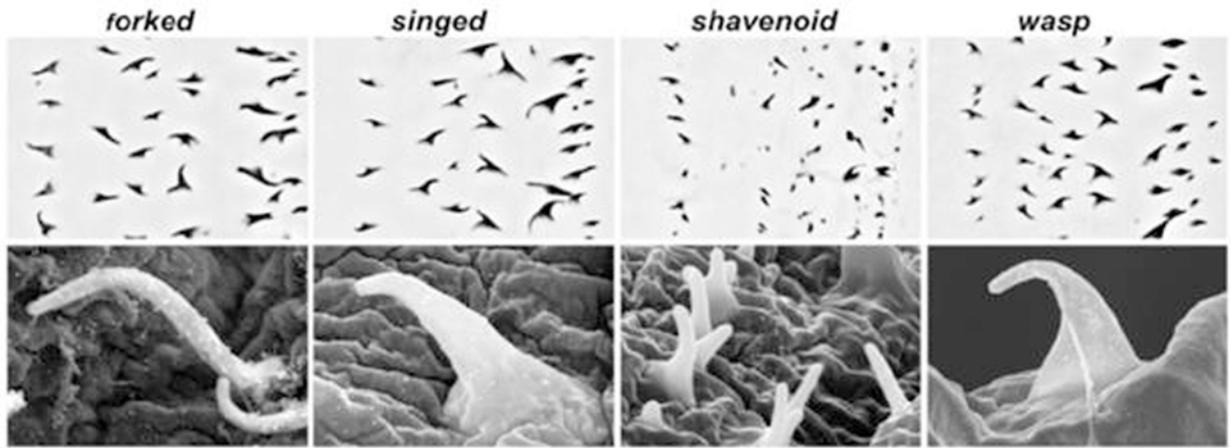
En collaboration avec l'équipe de Yuji Kageyama, les travaux de l'équipe montrent que, pour la morphogénèse épidermique, les peptides Pri agissent via la régulation post-traductionnelle du FT *Svb* (Fig 14 B). *pri* induit en effet un clivage protéolytique de la région N-terminale de *Svb*, ce qui conduit à la perte du domaine répresseur de la transcription (Kondo, Plaza et al. 2010). La forme activée de manière *pri*-dépendante de *Svb* devient donc un activateur de la transcription, dont la structure primaire est très proche d'OvoB (Kondo, Plaza et al. 2010) (Fig 14 C).

Cette étude démontre donc le rôle clé d'un ARN apparemment non-codant dans le contrôle du développement de la drosophile par la production de 4 peptides régulant l'activité transcriptionnelle de *Svb*.

## II.3 *Svb* active directement une batterie d'effecteurs cellulaires

### II.3.1 *Svb*, dernier FT de la différenciation des cellules épidermiques

Toutes les mutations qui modifient le patron des trichomes altèrent soit des gènes qui contrôlent l'expression de *svb* (*Wg*, *nkd*, *arm*, *TCF*, *Hh*, *EGF-R*, *N*, etc...) ou, dans le cas de *pri*, empêche la maturation activant le FT *Svb*. De plus, les variations de patron des



**Figure 15. Le rôle des protéines de liaison à l'actine dans la morphogénèse des trichomes.** Cuticules et MEB (microscopie électronique à balayage) de larves mutantes pour les effecteurs *forked*, *singed*, *shavenoid* et *wasp* participant au maintien du faisceau d'actine. Les denticules sont plus petites et la pointe est cassée. D'après Chanut et al .; 2006.

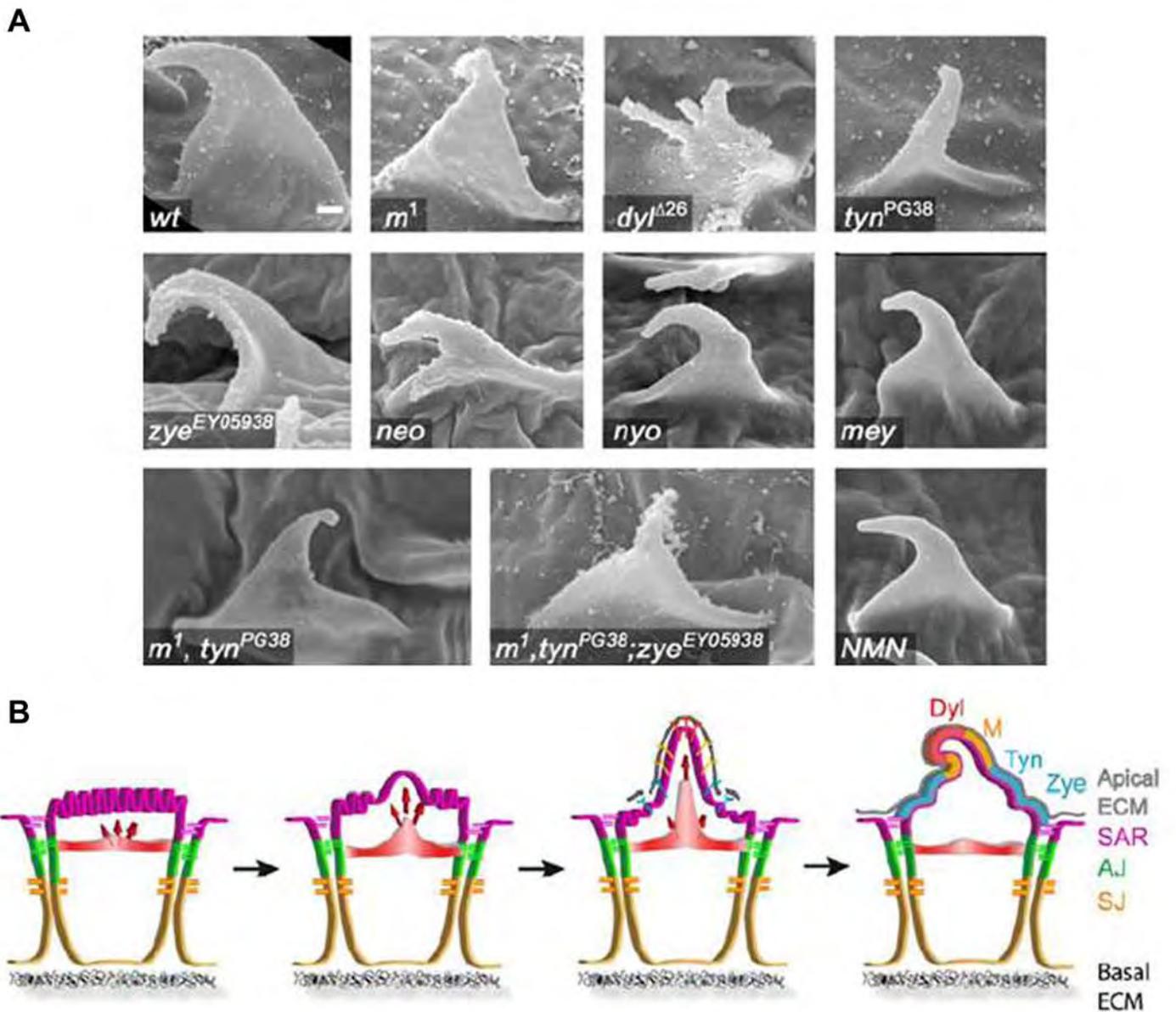
trichomes entre différentes espèces de drosophile sont dues à la modification du patron d'expression de *svb* au cours de l'évolution (Delon payre 2004, sucena et al 2003). Il semble donc que Svb soit le FT le plus aval des cascades régulatrices de la formation des trichomes. Pour tester cette hypothèse et identifier les mécanismes cellulaires responsables de la morphogénèse des trichomes épidermiques, l'équipe a entrepris la recherche des cibles directes de Svb.

### II.3.2 Trois familles de gènes cibles identifiées

Des approches candidates, ciblant des gènes sélectionnés pour leur profil d'expression épidermique, ont permis d'identifier une douzaine de gènes cibles dont l'expression dépend de Svb. Ces différents gènes sont ainsi exprimés dans tout ou partie des cellules à trichomes. L'absence de Svb conduit à une forte réduction de leur expression épidermique, sans toutefois altérer d'autres registres d'expression (par exemple dans l'intestin ou les cellules sanguines, dont la différenciation ne dépend pas de Svb). Au contraire, l'expression artificielle de Svb dans des cellules épidermiques lisses conduit à l'expression ectopique de ces cibles, de manière cellulaire autonome. Enfin, dans des embryons *D. sechellia*, dont la restriction de l'expression dorsale de Svb a conduit à la perte des trichomes de type 4, on retrouve une expression des différentes cibles, elle aussi restreinte aux mêmes cellules. Ces différents gènes cibles sont donc sous la régulation du facteur de transcription Svb, et nous allons voir le type de fonction connues ou présumées qu'ils remplissent.

#### a) Protéines de liaison à l'actine

La première classe de gènes dont l'expression dans les cellules à trichomes dépend de Svb code des protéines connues pour promouvoir la polymérisation et/ou la réorganisation des filaments d'actine. Ces données sont bien sûr importantes, car nous avons vu le rôle du cytosquelette d'actine pour la formation des trichomes. Des embryons mutants pour les protéines de liaison à l'actine WASP, Fascin (*singed*) et Espin (*forked*) présentent des



**Figure 16. Les protéines Zona Pellucida (ZP) participent au maintien des denticules.**

**A. Les phénotypes sont spécifiques pour chaque ZP.**

Images de microscopie électronique à balayage de larves mutantes pour une ou plusieurs ZPs. L'absence d'une ZP entraîne un défaut de forme des denticules spécifique selon la localisation de la protéine. Les phénotypes sont additionnels dans le cas de mutations doubles ou triples.

**B. Schéma de la mise en place des denticules.**

Les faisceaux d'actine participent à la formation initiale de l'extension. La localisation spécifique de chaque ZP dans une sous-région de la denticule maintient sa morphologie via des interactions avec la matrice extracellulaire.

D'après Fernandes et al.; 2010.

denticules plus fines, dont la pointe apparaît souvent cassée (Chanut-Delalande, Fernandes et al. 2006) (Fig 15). Chaque mutant produit cependant des défauts sensiblement différents (Fig 15). La combinaison de mutations individuelles conduit à des phénotypes additifs, renforçant l'idée d'une fonction spécifique de chacun de ces régulateurs d'actine pour la mise en place des extensions.

#### *b) Protéines de la cuticule*

Comme les denticules normales, les trichomes surnuméraires formés suite à l'expression ectopique de Svb apparaissent fortement pigmentés. L'équipe a donc recherché de nouvelles cibles de Svb parmi les enzymes de la voie de biosynthèse des catécholamines, indispensable à la pigmentation. Svb régule en effet l'expression du gène *yellow*, codant une protéine participant à la pigmentation/sclérotisation de la cuticule, nécessaire au durcissement des denticules (Chanut-Delalande, Fernandes et al. 2006). On retrouve aussi ChLD3, un gène codant un enzyme de la famille des chitines deacetylases, dont l'expression dans les trichomes dépend de *svb* mais dont la fonction reste à caractériser. De même, l'expression de *dsc73*, qui code une protéine sécrétée associée à la cuticule, est régulée par Svb (Andrew and Baker 2008). Ces données soutiennent donc un rôle de la modification de la cuticule dans la formation des trichomes épidermiques.

#### *c) Protéines de la matrice extracellulaire*

Une fonction importante pour la morphogenèse des trichomes est celle de différents composants de la matrice extracellulaire. En effet, l'équipe a montré que 8 gènes codant des protéines de la famille Zona Pellucida (ZP) (Plaza, Chanut-Delalande et al. 2010) sont régulés par Svb pour être exprimés dans les cellules à trichomes (Fernandes, Chanut-Delalande et al. 2010). L'inactivation individuelle de chacun des gènes ZP conduit à un défaut spécifique, altérant d'une manière particulière la forme de l'extension (Fig 16 A).



L'inactivation combinée de 2 ou 3 gènes ZP conduit à des phénotypes additifs, suggérant que les protéines ZPs présentent elles aussi des fonctions non redondantes pour la réorganisation apicale des cellules épidermiques (Fernandes, Chanut-Delalande et al. 2010) (Fig 16 A). En effet, chaque protéine ZP s'accumule dans un sous-domaine particulier de l'extension, contribuant localement à l'interaction entre la cuticule et la membrane plasmique (Fig 16 B). Cette localisation subcellulaire des protéines ZPs révèle l'existence de sous-domaines apicaux, suggérant une régulation inattendue du trafic vésiculaire en route vers le compartiment apical.

**Collectivement ces données montrent le FT Svb régule l'expression de différentes classes de gènes cibles, codant des régulateurs de l'actine, de la matrice extracellulaire, de la cuticule, ainsi que des protéines de fonction inconnue.**

Ces travaux ont ainsi permis d'identifier les premiers effecteurs cellulaires directement responsables de la morphogenèse des trichomes. Différentes projections suggéraient que ces acteurs ne représentaient en réalité qu'une sous fraction des effecteurs impliqués. Combien de protéines et familles fonctionnelles participent à la formation des trichomes restait donc à découvrir. De même comment l'expression de ces gènes est-elle régulée par le FT Svb : directement *vs* indirectement, nécessite-t-elle d'autres (co-)facteurs, quelle est l'architecture fonctionnelle de leur éléments régulateurs etc...

### II.3.3 Identification des CRMs spécifique des cellules à trichomes

Une première recherche des CRMs associés aux cibles de Svb a été menée par Serge Plaza via l'outil CIS-ANALYST (Berman, Nibu et al. 2002; Berman, Pfeiffer et al. 2004). Ce logiciel de prédiction de CRM se base sur la détection de clusters d'au moins 3 sites de fixation dans une fenêtre de 500bp, en tenant compte de la conservation entre *Drosophila*



*melanogaster* et *virilis*. Le seul CRM ainsi identifié, Emin (400pb), régule l'expression du gène *ZP miniature*. Emin dirige une expression spécifique des cellules à trichomes, perdue en absence de Svb, et l'expression de Svb dans les cellules lisses induit l'activation ectopique d'Emin, de manière cellulaire autonome (Chanut-Delalande, Fernandes et al. 2006). La protéine Svb se fixe *in vitro* sur Emin et la mutation du motif CnGTT (PWM Ovo) empêche cette fixation *in vitro*, et l'activité *in vivo* du CRM (Chanut-Delalande, Fernandes et al. 2006). Emin participe donc à l'expression de *miniature* dans les cellules à trichomes, sous le contrôle de Svb. Il s'agit de la première identification d'une cible, très probablement directe du FT Svb.

Afin d'identifier d'autres CRM cibles de Svb, Séverine Viala et Jennifer Zanet ont disséqué les régions génomiques des gènes *shavenoid (sha)* et *singed (sn)*. Par le test d'une batterie systématique de lignées reporter, ces travaux ont permis d'identifier un fragment de 600pb reproduisant le patron d'expression de *sn* dans les cellules à trichomes. Cette région appelée snE1 contient deux motifs CnGTT. Cette même approche a défini deux CRMs (env. 500bp) du gène *sha* dirigeant une expression spécifique des cellules à trichomes, sous la dépendance de *svb*. Ces régions présentent également des PWMs Ovo évolutivement conservés. Le fait que 2 autres régions *sha0* et *sha2* présentent les mêmes PWMs conservées mais ne dirigent pas d'expression dans l'épiderme souligne le fait que ce critère n'est pas suffisant pour identifier les CRMs associés aux cibles de *svb*.

A mon arrivée au laboratoire l'enjeu était alors d'identifier l'ensemble des gènes cibles de Svb afin de comprendre : i) les mécanismes de différenciation terminale des cellules épidermiques, ii) le mode de régulation utilisé par Svb pour reconnaître ses gènes cibles. J'ai donc entrepris des approches génome entier pour identifier les gènes et CRMs cibles de Svb. Mon projet s'est articulé autour de deux axes principaux :



- Existe-il un code régulateur associé aux gènes cibles de Svb ? Ce code suit-il les modèles généralement proposés pour l'architecture fonctionnelle des CRMs ? Le FT Svb agit-il avec des cofacteurs pour induire l'expression de ses effecteurs ?

- Quelle est la nature et la fonction des nouveaux effecteurs identifiés ? Existe-t-il d'autres processus nécessaires à la morphogénèse épidermique ? Si oui s'agit-il de mécanismes spécifiques des cellules à trichomes ou généraux de la différenciation épidermique ?



# Régulation transcriptionnelle de la différenciation épidermique



La formation des trichomes résulte de l'action concertée de nombreux facteurs cellulaires, dont certains sont exprimés spécifiquement dans les cellules à trichomes, sous la régulation directe du FT Svb. Une vingtaine de gènes cibles de Svb ont ainsi été identifiés par des approches candidates (Chanut-Delalande, Fernandes et al. 2006; Andrew and Baker 2008; Fernandes, Chanut-Delalande et al. 2010). Cette stratégie a révélé différentes familles de gènes effecteurs régulés par Svb (codant des protéines de liaison à l'actine, cuticule et matrice extra-cellulaire), mais le répertoire complet des fonctions cellulaires nécessaires à la formation des trichomes restait à identifier. Par exemple, la localisation spécifique des ZPs dans les trichomes implique l'existence d'un mécanisme particulier d'adressage, dont au moins une partie des acteurs pourraient être régulés par Svb pour être exprimés dans les cellules à trichomes. L'implantation stéréotypée des trichomes au sein de la cellule épidermique (position, orientation, etc...) suggère aussi l'existence de facteurs encore inconnus pour contrôler et/ou effectuer ce processus. Enfin, la modification de l'organisation et de la composition de la cuticule recouvrant les trichomes suggère l'existence de facteurs additionnels participant à cette interaction fonctionnelle entre cytosquelette et matrice extra-cellulaire. Par-delà l'identité des gènes cibles directs de Svb, la nature et la logique de leurs éléments de régulation transcriptionnelle restaient complètement à découvrir.

Nous avons donc entrepris des approches globales visant à la découverte de l'ensemble des cibles directes de Svb dans les cellules à trichomes, et l'identification de leur(s) CRMS associés. Nous avons complété ces études par la dissection fonctionnelle et la modélisation de l'architecture régulatrice de différents CRMs dépendants de Svb. Ces résultats font l'objet de la première partie de la section résultats.

Grace à ces approches globales, nous avons identifié de nouvelles fonctions régulées par Svb pour la morphogénèse épidermique. Nous avons poursuivi leur validation *in vivo* et débuté la caractérisation fonctionnelle de certains acteurs ainsi identifiés. Cet effort a été



accompli en interaction avec Hélène Chanut, qui a réalisé en parallèle un crible génétique systématique à la recherche de gènes nécessaires à la mise en place et/ou différenciation des trichomes. Ces deux approches globales indépendantes nous ont permis de converger sur plusieurs fonctions cellulaires et leurs acteurs respectifs. Ces travaux sont présentés dans la deuxième partie de la section résultats.

Nos approches non biaisées de profilage des variations du transcriptome ont aussi généré des résultats inattendus. Nous avons ainsi découvert des fonctions additionnelles des peptides Pri pour la différenciation épidermique, indépendamment de leur activité sur le facteur de transcription Svb. L'activité des peptides Pri est indispensable pour la différenciation épidermique en général, *i.e.* des cellules à trichomes et des cellules lisses. Nous avons défini un large ensemble de gènes dont les peptides Pri préviennent l'expression prématurée dans l'épiderme embryonnaire, et dont la dérégulation conduit à des défauts de l'organisation de la cuticule. Ces données constituent la troisième partie des résultats.

Ensemble, nos travaux ouvrent donc de nouveaux champs d'investigation afin de mieux comprendre les différents mécanismes de régulation transcriptionnelle responsables de la différenciation de l'épiderme embryonnaire.



## Partie I :

**Identification des cibles *in vivo* du facteur Svb  
et décryptage de leur code de régulation**



## INTRODUCTION

Le rôle des facteurs de transcription pour le contrôle de l'expression du génome est un élément clé du développement embryonnaire. C'est la reconnaissance par un FT des CRMs de ses gènes cibles qui va induire leur expression dans une population particulière de cellules, et à un stade donné du développement. Le déchiffrement du code utilisé par un FT pour reconnaître ses effecteurs constitue donc un challenge incontournable pour identifier les réseaux géniques contrôlant les différentes étapes du développement. Grâce au développement ces 10 dernières années de nouvelles technologies à haut débit, plusieurs études ont permis d'approcher le code utilisé par certains FTs pour activer leurs gènes cibles. La majorité des analyses se sont focalisées sur des FTs impliqués dans des étapes précoces de l'embryogenèse, comme par exemple la spécification du mésoderme. L'ensemble de ces découvertes a conduit à proposer des modèles d'organisation des différents sites de fixation pour les FT régulant l'activité des CRMs. Cependant, peu d'études ont concerné l'analyse des mécanismes transcriptionnels gouvernant un processus de différenciation terminale. Ainsi, très peu de choses sont connues sur la nature des gènes cibles, et la grammaire architecturale de leurs CRMs, pour diriger les étapes ultimes de régulation de l'expression du génome aboutissant à la différenciation cellulaire, notamment le contrôle de la forme tridimensionnelle des cellules.

Pour aborder cette question, nous utilisons le modèle de la morphogénèse épidermique au cours de l'embryogenèse tardive chez la drosophile. Ce processus est gouverné par le FT Svb, qui déclenche la formation d'extensions apicales dites trichomes. L'objectif de nos travaux est de définir comment ce facteur Svb reconnaît et sélectionne ses gènes cibles, indispensables au remodelage de la forme des cellules épidermiques. En couplant des approches globales de transcriptome et ChIP-seq, des approches fonctionnelles d'analyse systématique d'enhancers *in vivo*, ainsi que des approches de prédictions et



modélisations bio-informatiques, nous avons pu définir un large ensemble des gènes cibles directs de Svb et commencé à décrypter la logique de leurs éléments régulateurs. Nous montrons ainsi l'importance du contexte intime des sites de fixation de Svb pour leur activité *in vivo*, établissant une nouvelle PWM pour Svb plus performante pour prédire les sites fonctionnels que la matrice définie *in vitro*. Nous identifions d'autre part deux motifs « orphelins » additionnels, dont l'intégrité est généralement critique pour l'activité des CRMs. Les CRMs des cibles de Svb ne présentent cependant pas d'organisation stéréotypée, ni en terme de nombre ni en terme d'arrangement respectif des différents sites. En utilisant l'ensemble de ces critères et en croisant avec des données de transcriptome et de ChIP-seq, nous avons pu cartographier les régions génomiques fixant *in vivo* le FT Svb au voisinage de nouveaux gènes effecteurs. La combinaison de ces approches a permis d'identifier de nouveaux CRMs dépendants de Svb et dirigeant une expression dans l'épiderme, spécifiquement dans les cellules à trichomes. Notre étude a donc permis de décrypter différents aspects de la grammaire utilisée par Svb pour reconnaître ses effecteurs. Ces premières données sur les étapes de différenciation terminale montrent que les CRMs dépendants de Svb présentent une architecture générale à l'interface entre le modèle « enhanceosome » (organisation figée) et « billboard » (organisation flexible).

Ces résultats sont présentés sous forme d'un article récemment publié. Nous présentons aussi des résultats additionnels suivis d'une discussion générale.



RESEARCH

Open Access

# Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization

Delphine Menoret<sup>1,2†</sup>, Marc Santolini<sup>3†</sup>, Isabelle Fernandes<sup>1,2,4</sup>, Rebecca Spokony<sup>5</sup>, Jennifer Zanet<sup>1,2</sup>, Ignacio Gonzalez<sup>6,7</sup>, Yvan Latapie<sup>1,2</sup>, Pierre Ferrer<sup>1,2</sup>, Hervé Rouault<sup>3,8</sup>, Kevin P White<sup>5</sup>, Philippe Besse<sup>6,7</sup>, Vincent Hakim<sup>3</sup>, Stein Aerts<sup>9</sup>, Francois Payre<sup>1,2\*</sup> and Serge Plaza<sup>1,2\*</sup>

## Abstract

**Background:** Developmental programs are implemented by regulatory interactions between Transcription Factors (TFs) and their target genes, which remain poorly understood. While recent studies have focused on regulatory cascades of TFs that govern early development, little is known about how the ultimate effectors of cell differentiation are selected and controlled. We addressed this question during late *Drosophila* embryogenesis, when the finely tuned expression of the TF Ovo/Shavenbaby (Svb) triggers the morphological differentiation of epidermal trichomes.

**Results:** We defined a sizeable set of genes downstream of Svb and used *in vivo* assays to delineate 14 enhancers driving their specific expression in trichome cells. Coupling computational modeling to functional dissection, we investigated the regulatory logic of these enhancers. Extending the repertoire of epidermal effectors using genome-wide approaches showed that the regulatory models learned from this first sample are representative of the whole set of trichome enhancers. These enhancers harbor remarkable features with respect to their functional architectures, including a weak or non-existent clustering of Svb binding sites. The *in vivo* function of each site relies on its intimate context, notably the flanking nucleotides. Two additional *cis*-regulatory motifs, present in a broad diversity of composition and positioning among trichome enhancers, critically contribute to enhancer activity.

**Conclusions:** Our results show that Svb directly regulates a large set of terminal effectors of the remodeling of epidermal cells. Further, these data reveal that trichome formation is underpinned by unexpectedly diverse modes of regulation, providing fresh insights into the functional architecture of enhancers governing a terminal differentiation program.

## Background

Many studies have established that transcriptional networks control development, through determining specific programs of genome expression [1]. These gene regulatory networks (GRNs) are implemented by transcription factors (TFs) that bind to regulatory DNA sequences, known as enhancers or *cis*-regulatory modules (CRMs), to control the transcription of nearby genes. Although recruited to

target genes via their DNA binding properties [2], TFs recognize only short and often degenerate motifs (reviewed in [3,4]). Consequently, thousands of putative binding sites (BSs) are scattered throughout the genome, hampering efficient prediction of CRMs [3,5,6]. The fine structure of enhancers as well as putative general rule(s) underlying their organization remain, however, poorly understood.

Although animals encode hundreds of TFs, only a few of them have been studied in detail to elucidate the regulatory logic of their target enhancers [7,8]. In *Drosophila*, current knowledge of enhancer structure mainly comes from works on early development - for example,

\* Correspondence: francois.payre@univ-tlse3.fr; serge.plaza@univ-tlse3.fr

† Contributed equally

<sup>1</sup>Centre de Biologie du Développement, Université de Toulouse, 118 route de Narbonne, Toulouse, F-31062, France

Full list of author information is available at the end of the article

TFs controlling segmentation and mesoderm specification [9-12]. Within these early acting networks, several studies have shown that the local enrichment for BSs (homotypic or heterotypic clustering) in evolutionarily conserved regions is a general signature of active enhancers [13-15]. Functionally related enhancers (driving similar expression patterns) often share a combination or code of *cis*-regulatory motifs, together defining a specific program of expression [11,16-18]. Whether enhancers rely on a constrained organization of *cis*-regulatory motifs or can accommodate flexibility in their number, composition and positioning is still debated (reviewed in [4,19,20]). While several studies have shown that regulatory codes are efficient to predict expression pattern [9,11,16], recent large-scale work suggests that developmental enhancers may have a more flexible architecture [10,20]. However, in-depth analyses of individual enhancers [21-24] have revealed an unexpected level of functional constraint in their intimate architecture. It has been proposed that constrained enhancers could be critical when TFs display limiting concentrations [25] - for example, to accurately integrate gradients [26]. On the other hand, enhancers that do not hold integrative properties might be of simpler architecture [27,28]. Distinguishing between these possibilities thus requires detailed analyses of the structure and regulatory logic of CRM-TF interactions that occur at late developmental stages.

Here, we focus on a GRN that controls cell morphogenesis during terminal differentiation of the *Drosophila* embryonic epidermis. The subset of epidermal cells that express the TF Ovo/Shavenbaby (Svb) [29] undergo localized changes in cell shape leading to the formation of dorsal hairs and ventral denticles, collectively referred to as trichomes [30]. Svb triggers the expression of various classes of cellular effectors in trichome cells. Developmental and genetic analyses have established that trichome formation relies on their collective action, acting together as a developmental module to promote cell shape reorganization [31-33]. The mechanisms underlying the co-expression of Svb-regulated genes in trichome cells remained yet poorly understood. A first level of regulation resides in the activity of Svb itself, which is controlled in a post-translational manner in response to small peptides encoded by the gene *polished-rice* (*pri*) [34]. Pri peptides trigger amino-terminal truncation of the Svb protein, switching its activity from a full-length repressor to a cleaved activator [34], therefore providing temporal control to the program of trichome formation [32]. However, little is known concerning how this TF recognizes and selects its target genes. Besides definition of DNA-binding specificity *in vitro* [35] and the identification of a few targets regulated by Ovo germline-specific isoforms [35,36], only a single epidermal enhancer dependent on Svb has been identified so far [31].

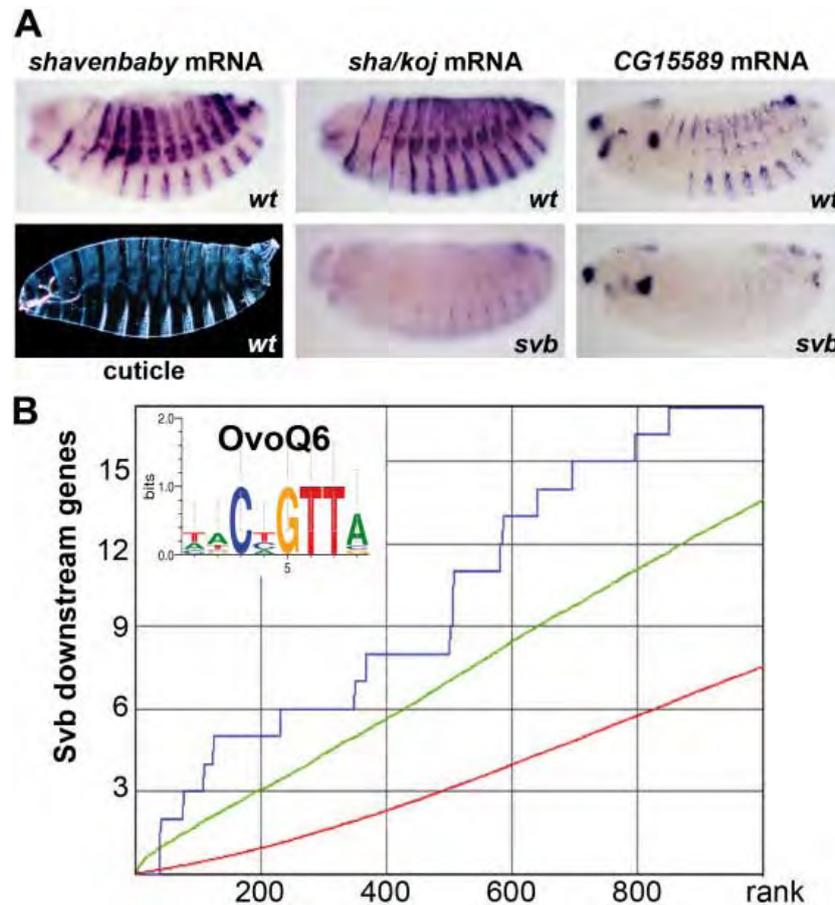
Thus, whether or not Svb targets genes that are co-expressed in trichome cells and have similar *cis*-regulatory elements remained an open question.

To address this question, we designed a set of computational modeling coupled to experimental approaches to identify and investigate the *cis*-regulatory logic of Svb-dependent enhancers. By systematic *in vivo* assays, we first identified a robust set of Svb target effectors, specifically expressed in trichome cells at the time of their morphological differentiation. We then searched for and identified 14 Svb-dependent epidermal enhancers driving their expression in trichome cells and investigated their functional organization. Computational analyses and experimental dissection led to a refinement of the Svb BSs bound *in vivo* and the identification of two additional motifs required for enhancer activity. Our studies further reveal that the distribution of these *cis*-regulatory motifs does not follow a stereotypical organization. Coupled to chromatin immunoprecipitation (ChIP)-seq and microarray profiling, the models built from these fine scale experiments allow efficient genome-wide identification of new enhancers that drive the specific expression of trichome effectors. In summary, our results show that enhancers driving co-expression in cells of a late GRN have variable composition and respective organization of *cis*-regulatory motifs, extending the idea that co-expressed developmental enhancers can have diverse *cis*-regulatory architectures [11,37], including for those mediating terminal stages of cell differentiation.

## Results

### Enrichment of conserved binding sites in Svb downstream genes

Previous work has identified a dozen genes activated by Svb, each contributing to epidermal cell remodeling [31,33,38,39]. To investigate the *cis*-regulatory logic of Svb-dependent targets, we first sought to define a larger set of Svb downstream genes appropriate for *in silico* analyses. We therefore analyzed additional candidates selected because of their expression in subsets of epidermal cells (from the Berkeley *Drosophila* Genome Project) using *in situ* hybridization. Of 57 candidates, we identified 21 Svb-dependent genes, that is, those downregulated in *svb* mutants and upregulated following *svb* ectopic expression (Figure 1a; Figure S1A in Additional file 1 (legend in Additional file 2)), while the other 36 epidermal genes were found to be independent of Svb (Additional file 1, Figure S1B). Together with genes identified previously [31,33,38,39], this constitutes a robust set of 39 genes activated by Svb to be expressed in trichome cells. We used these 39 Svb targets to examine whether they display an evolutionarily conserved signature in their non-coding regions when compared with all *Drosophila* genes, or the 36 epidermal genes independent of *svb* as a negative



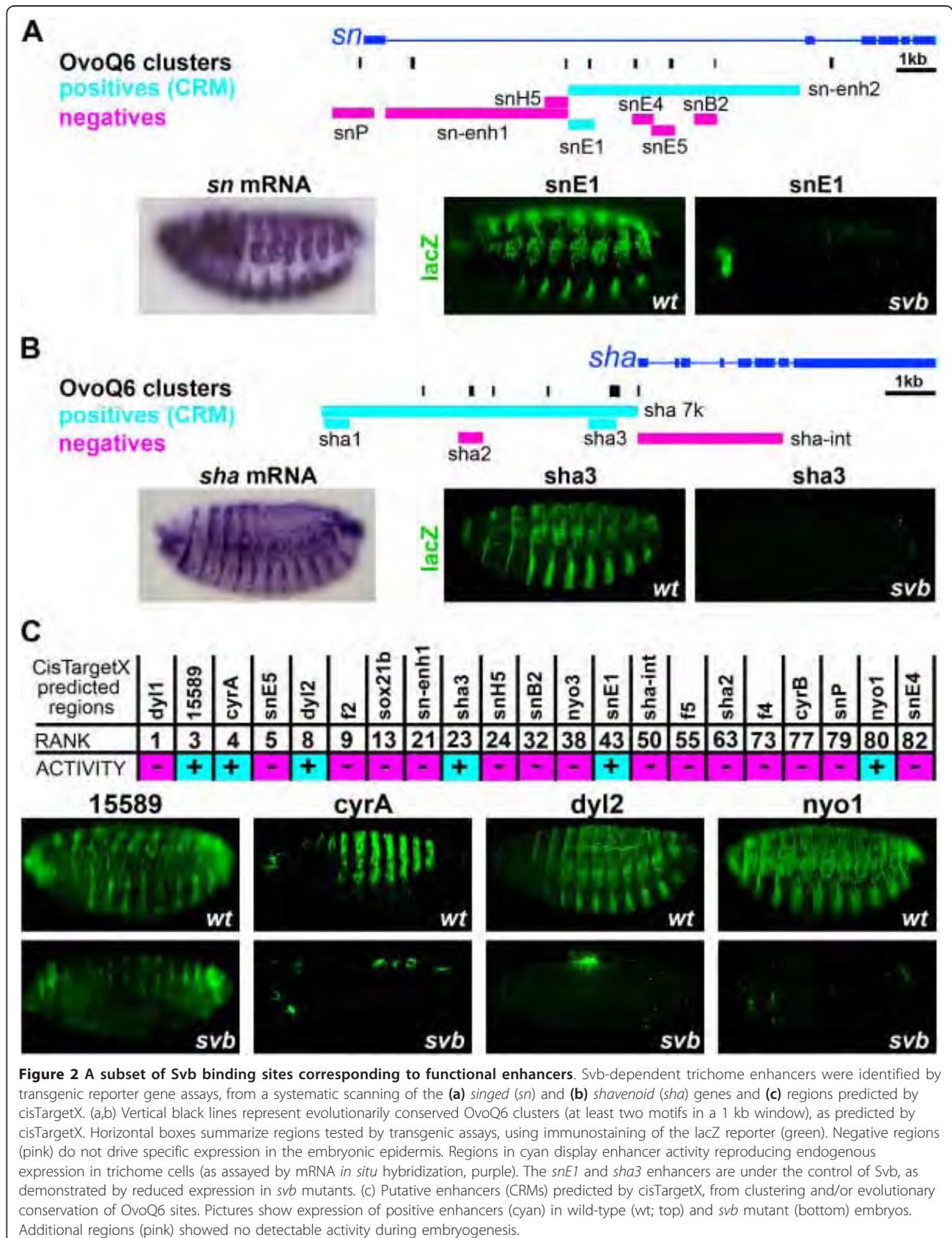
**Figure 1** Enrichment in binding sites defines an evolutionarily conserved signature of *svb* downstream genes. (a) Expression of *svb* mRNA determines the epidermal cells that form trichomes, visible on the larval cuticle. *In situ* hybridization shows mRNA expression of two *svb* downstream genes, *shavenoid* (*sha/koj*) and *CG15589*, in wild-type (wt; top) and *svb* mutant embryos (bottom). (b) Receiver operating characteristic curve showing significant enrichment in putative Svb binding sites (OvoQ6 position weight matrix) among the 39 Svb downstream genes (y-axis) compared to a randomized set of 1,000 *Drosophila* genes (x-axis) using cisTargetX. The blue curve shows the detection of Svb downstream genes, the red curve a random distribution, and the green curve shows a 2 sigma interval from random.

control. cisTargetX aims at detecting motifs enriched among a group of co-expressed genes - for example, to predict direct targets of a TF [40]. It exploits a library of >3,000 motifs, including TF BSs and ultra-conserved DNA words [41], each motif being ranked with a score representative of both clustering and evolutionary conservation [40]. When applied to Svb targets, four of the top five motifs match the consensus CnGTT (Figure 1b; Figure S1C in Additional file 1), characteristic of the Ovo/Svb BS CnGTTa as defined *in vitro* [35]. From the 39 input genes, CisTargetX determined an optimal subset of 16 Svb direct targets, having the highest scores for the OvoQ6 motif (Figure 1b; Figure S1C in Additional file 1) [35,36]. OvoQ6 was specific to Svb targets since it was not detected in control epidermal genes (Figure S1C in Additional file 1). In contrast, motifs matching the BS of TFs involved in general epidermis differentiation, such as Grainy head [42] or Vrille/c-EBP [43], were highly ranked

in Svb-independent genes (Figure S1C in Additional file 1) while lowly ranked in Svb downstream genes. Hence, OvoQ6 motifs appear to be a signature of a subset of genes activated by Svb, a result consistent with their direct regulation.

#### Distribution of Svb binding site clusters poorly correlates with enhancer activity

We then examined the genomic distribution of OvoQ6 motifs within Svb target loci showing significant enrichment compared to random *Drosophila* genes. We found that each target gene contained evolutionarily conserved OvoQ6 scattered throughout intergenic and intronic regions (Figure 2a, b), instead of OvoQ6 clusters enriched locally (even using relaxed conditions of at least two sites per kilobase). To delineate which regions mediate epidermal expression, we generated a series of transgenic reporters that systematically scan two Svb downstream genes.



We focused on *singed* since it encodes Fascin, a conserved regulator of actin organization [44], and *shavenoid*, which encodes a pioneer protein but displays an extreme trichome phenotype upon inactivation [31]. Although most regions with OvoQ6 sites did not show embryonic expression, we identified three sequences, one in *singed* (*snE1*) and two in *shavenoid* (*sha1* and *sha3*), that drove expression in the epidermis, specifically in trichome cells (Figure 2a, b). Unexpectedly, one of the three sequences, *sha1*, displays a single recognizable OvoQ6 motif (see below) in *D. melanogaster*, as well as in sibling species. The activity of all three regions was lost when introduced into a *svb* null mutant background, showing that they are functional Svb target enhancers (Figure 2a, b). cisTargetX predicts the location of putative enhancers within each gene [40] and two out of three enhancers defined *in vivo* matched these predictions, in one case (*sha3*) at the highest rank for this gene (Figure 2c). We therefore investigated whether evolutionarily conserved OvoQ6 sites were sufficient to predict trichome enhancers and assayed 18 additional regions (Figure 2c) taken from the top 100 predictions. Transgenic reporter assays identified four novel sequences from *CG15589*, *cypher*, *dusky-like* and *neyo* driving expression in the epidermis. We verified in each case that they were specifically expressed in all (*dyl2*, *nyo1*) or subsets (*15589*, *cyrA*) of trichome cells where Svb is active. Consistently, these four enhancers depended on Svb since they displayed a strong reduction in their expression in the absence of *svb* (Figure 2c). Hence, analysis of Svb downstream targets shows that they are enriched in OvoQ6 BSs, a feature well conserved across *Drosophila* species. However, putative trichome enhancers predicted from evolutionary conservation and clustering of OvoQ6 sites were validated at a rate of only 28% (6/21; Figure 2c), most tested regions being devoid of activity in embryos, suggesting that other criteria distinguish enhancers from negative regions.

We noticed that OvoQ6 clusters failed to predict a number of active enhancers. This was the case for *sha1* (Figure 2) and *Emin*, an epidermal enhancer previously identified in the gene *miniature* [31]. Examination with Cluster-Buster [45] and Swan [46] did not detect supplementary OvoQ6 in *sha1* or *Emin* sequences (even in *D. melanogaster* only), explaining why these enhancers, containing a single Svb BS, are not included in *in silico* predictions. Six additional enhancers identified during initial stages of this study using alternative prediction criteria (Figure S1C in Additional file 1) were not highly ranked by cisTargetX because they lack BS clustering and/or evolutionary conservation. These data therefore show that BS clustering is not an absolute requisite for Svb regulation (Figure 2c), suggesting that additional sites are required to discriminate between enhancers and inactive regions.

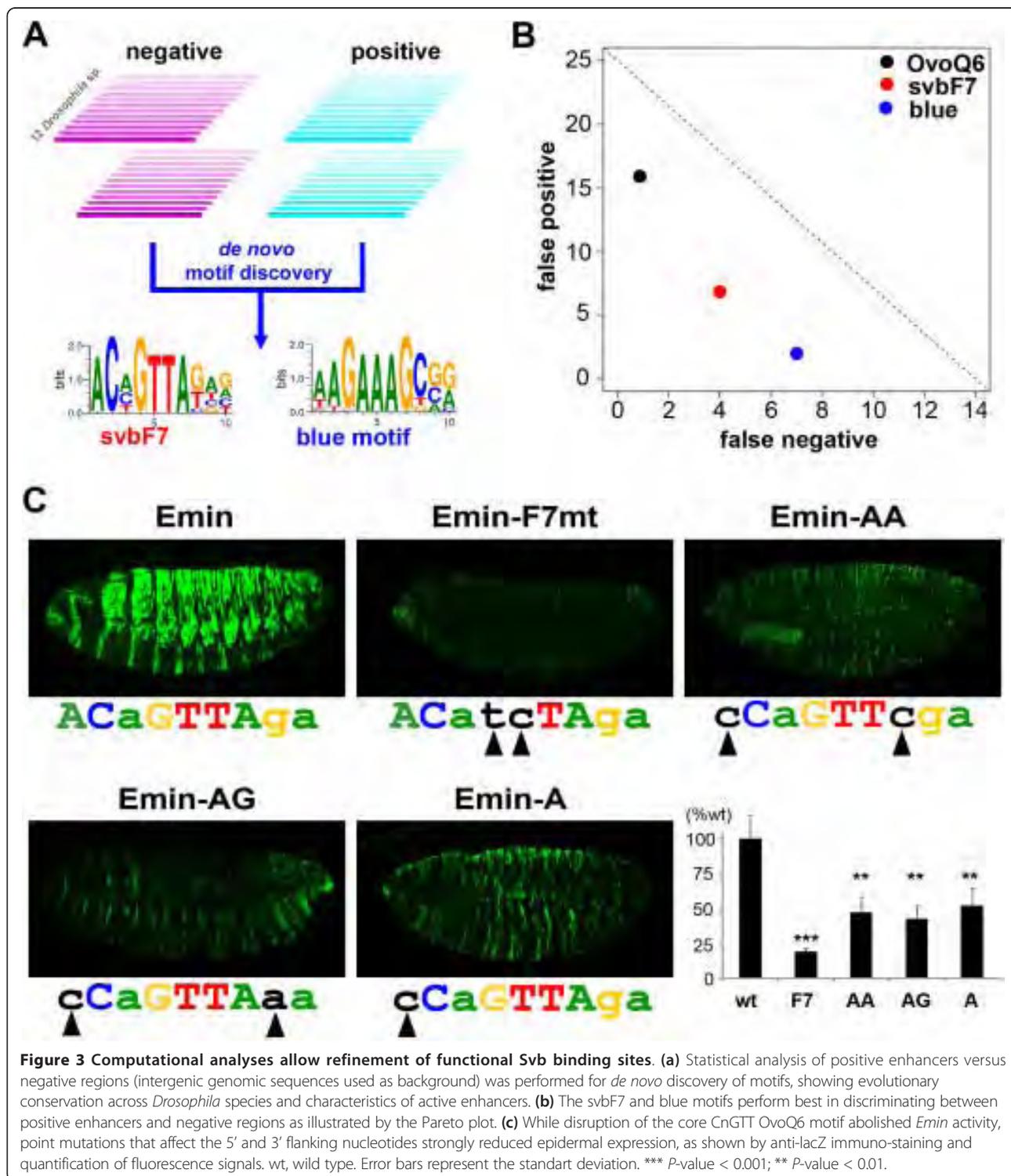
### **De novo motif discovery identifies a specific signature of Svb binding sites active *in vivo***

To search for putative active Svb binding sites, we compared the two sets of experimentally tested regions - that is, the 14 enhancers (positive) and 25 inactive regions (negative) - using *Imogene*, an algorithm designed for *de novo* motif discovery [47]. Briefly, we systematically searched, *ab initio*, for 10 bp motifs that are evolutionarily conserved across Drosophilidae and display a distribution within each region statistically different from background sequences. We then evaluated how well each motif discriminated between enhancers and inactive regions and ranked these *de novo* motifs accordingly (Figure 3a). Strikingly, the most discriminative motif overlaps OvoQ6 (CnGTTa), with a similar core consensus but extending to adjacent nucleotides (ACHGTTAK). A second discriminative motif (WAGAAAGCSR), called the blue motif, was also found, and is discussed below.

The ACHGTTAK motif, hereafter called svbF7, was sufficient to detect 10 out of 14 enhancers (Figure 3b). The proportion of svbF7-positive enhancers reached 13/14, when relaxing the penalty imposed for poor conservation [47]. In contrast, svbF7 was found in only 6/25 negative regions (Figure 3b), even when lowering the threshold (data not shown). Once added to the cisTargetX library, svbF7 is the most significant motif found in the set of 39 Svb downstream genes (Figure S1C, D in Additional file 1). It also increased the accuracy of enhancer predictions, with 3 additional positives (*32159*, *Emin* and *EminB*) while 9 negatives were removed from the top 100 cisTargetX regions (Figure S1C in Additional file 1). Hence, svbF7 performs better than OvoQ6 or any other related motifs [48] (Figure 3b; Figure S1D in Additional file 1). To evaluate whether this slight extension of the Svb BS was relevant for activity, we substituted nucleotides flanking the core CnGTTa in the single svbF7 of *Emin* - that is, we altered the svbF7 motif without disrupting the OvoQ6 consensus sequence (Figure 3c). When assayed *in vivo*, different patterns of flanking substitutions, including a single point mutation of the 5' A residue, were sufficient to strongly reduce *Emin* expression (Figure 3c). This demonstrates the functional importance of flanking nucleotides within the svbF7 motif for CRM activity. Hence, our computational analysis of Svb-dependent enhancers has discovered a refined nucleotide sequence required for *in vivo* regulation.

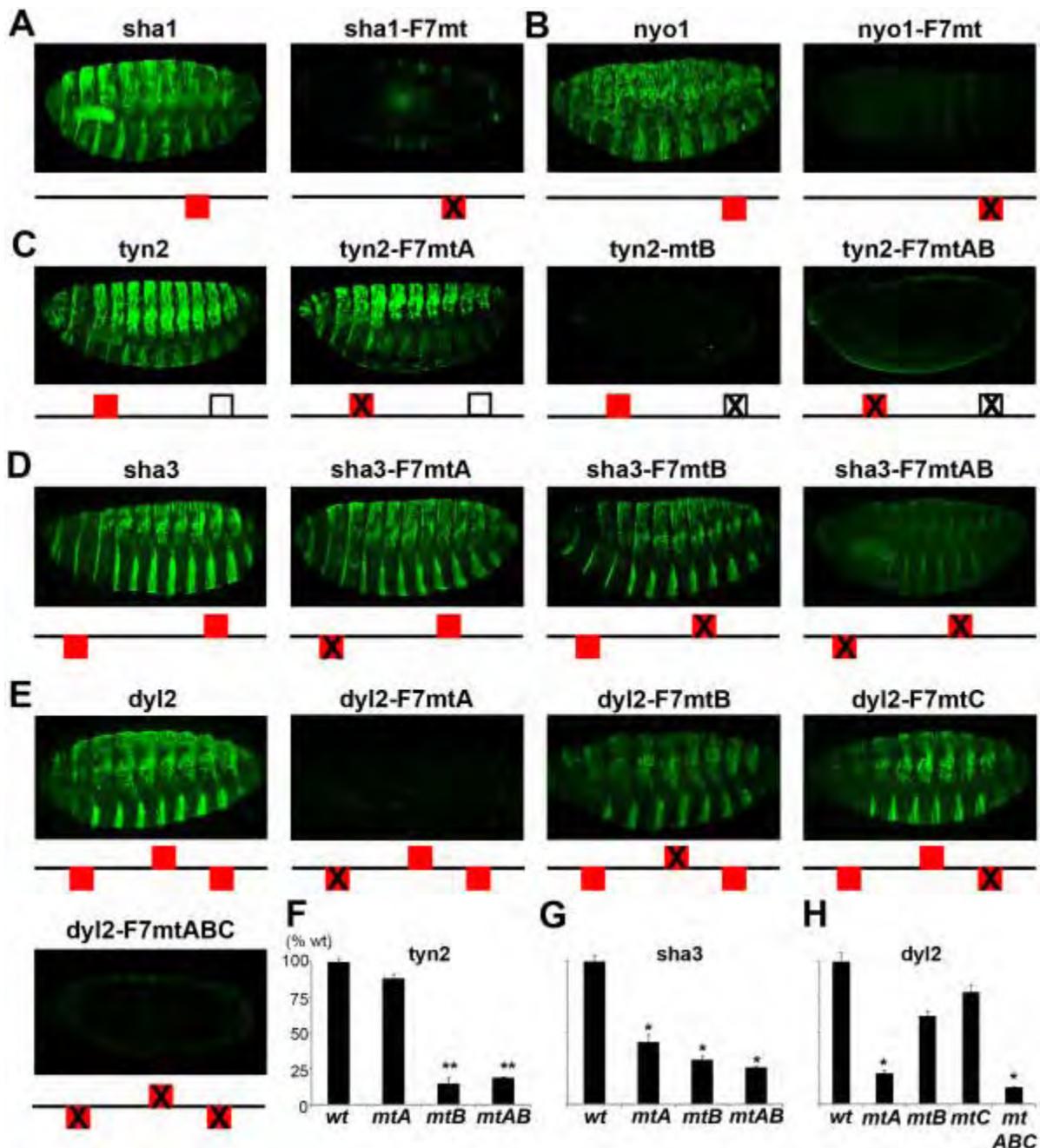
### **Trichome enhancers use different combinations of cis-regulatory motifs**

Having shown the role of svbF7 in *Emin*, we investigated its functional significance in other enhancers. We focused on enhancers containing from one to three predicted SvbF7 sites, to address the importance of single versus



clustered BSs for trichome cell expression. As observed for *Emin*, disruption of the single svbF7 site abolished the activity of both *sha1* and *nyo1* (Figure 4a, b). The mutation of svbF7 also decreased the activity of *tyn2*, albeit weakly and only in ventral cells (Figure 4c). In this

enhancer, however, we detected a second putative site that appears less conserved across species. Its inactivation strongly reduced expression (Figure 4), showing that this site mainly contributes to *tyn2* activity. For *sha3* and *dyl2*, which contain two and three svbF7 sites, respectively,



**Figure 4** *In vivo* role of svbF7 motifs in Svb-dependent enhancers. Anti-LacZ staining (green) shows modifications of reporter gene expression resulting from individual and simultaneous inactivation of svbF7 motifs in (a) *sha1*, (b) *nyo1*, (c) *tyn2*, (d) *sha3* and (e) *dyl2* enhancers. Red boxes schematize evolutionarily conserved svbF7 motifs; the open black box, a site that does not appear conserved across Drosophilidae. (f-h) Quantification of residual activity following individual disruption of svbF7 sites. \*\**P*-value < 0.01; \**P*-value < 0.05. wt, wild type. Error bars represent the standard deviation.

simultaneous inactivation of these sites abrogated expression (Figure 4d, e). The individual disruption of svbF7 sites nonetheless led to varying defects. The two svbF7 sites of *sha3* are partly redundant, their individual knockout showing similar and limited impacts when compared

to their simultaneous knockout (Figure 4d-g). In contrast, a single svbF7 site plays a major role in *dyl2* activity, whereas the two others contribute marginally to expression pattern or levels (Figure 4e, h). Hence, the disruption of svbF7 leads to reduced expression for all enhancers that



inactivation abolishing expression (Figure 5c), similar to the simultaneous mutation of both svbF7 sites (Figure 4d). In addition, we noticed that one important region for *Emin* expression (10mt; Figure 5a) matches an 8mer (TTATGCAA), previously predicted as a regulatory element from discovery of ultra-conserved DNA words in the genome of distant *Drosophila* species [41]. Although not sufficient by itself to discriminate between active enhancers and negative regions (data not shown), this motif, which we call the yellow motif, was nevertheless retrieved in six additional trichome enhancers (Figure 5b). To further assay *in vivo* the role of the yellow motif, we generated mutant versions of the *17058* and *nyo1* enhancers that disrupt their yellow motifs. As observed for *Emin*, mutation of the yellow motif led to a strong decrease in the expression driven by both *nyo1* and *17058* (Figure 5d), showing that the yellow motif represents a functional *cis*-regulatory element in a subset of enhancers.

Taken together, these data support that svbF7 is a main feature of Svb targets, this motif being shared by the vast majority (13/14) of active enhancers. Our analyses have discovered two additional *cis*-regulatory elements, the blue and yellow motifs, present in overlapping subsets of trichome enhancers (9/14 and 7/14, respectively). While the three motifs are present in various patterns and combinations (Figure 5B; Figure S2 in Additional file 1), functional assays demonstrated that each of them contributes to the *in vivo* activity of this sample of trichome enhancers.

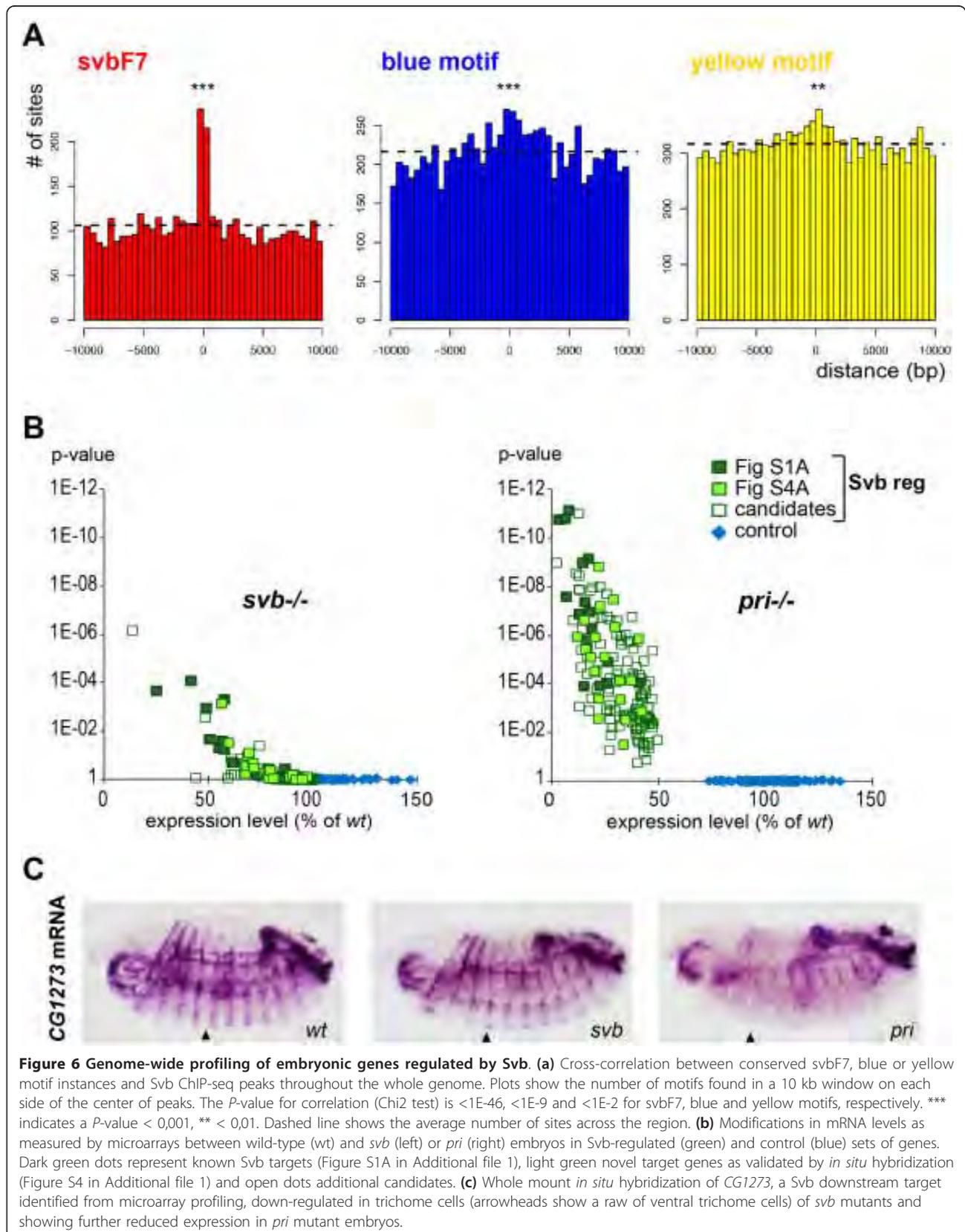
#### Genome-wide prediction of Shavenbaby target enhancers

To address whether these *cis*-regulatory motifs were a relevant signature of the genome-wide set of enhancers regulated by Svb, we undertook ChIP-seq to obtain an extensive mapping of Svb binding sites in epidermal cells. To improve specificity, we used a Svb::GFP transgene driven in ventral and dorsal trichome cells by two complementary *svb* *cis*-regulatory regions [34], likely at levels comparable to endogenous since it rescues *svb* mutant phenotypes [49]. ChIP-seq data indicated that Svb was bound to almost 6,000 genomic sites, a large number of binding events being a feature shared by several *Drosophila* TFs [6,8,15]. Analysis of ChIP peaks with *i-cisTarget* [50] showed that svbF7 and OvoQ6 are the most enriched motifs. A strong cross-correlation between conserved svbF7 and the center of ChIP peaks confirmed the importance of this motif (Figure 6a). As observed in our pilot analysis of enhancers, we did not detect high svbF7 clustering, multiple svbF7 motifs being rarely found within genome-wide ChIP peaks. Blue motifs (and to a lesser extent yellow motifs) also displayed a significant but weaker correlation with Svb peaks, consistent with wider genomic distribution (Figure 6a).

With the large number of Svb bound regions detected by ChIP-seq, it was unlikely that all of them were

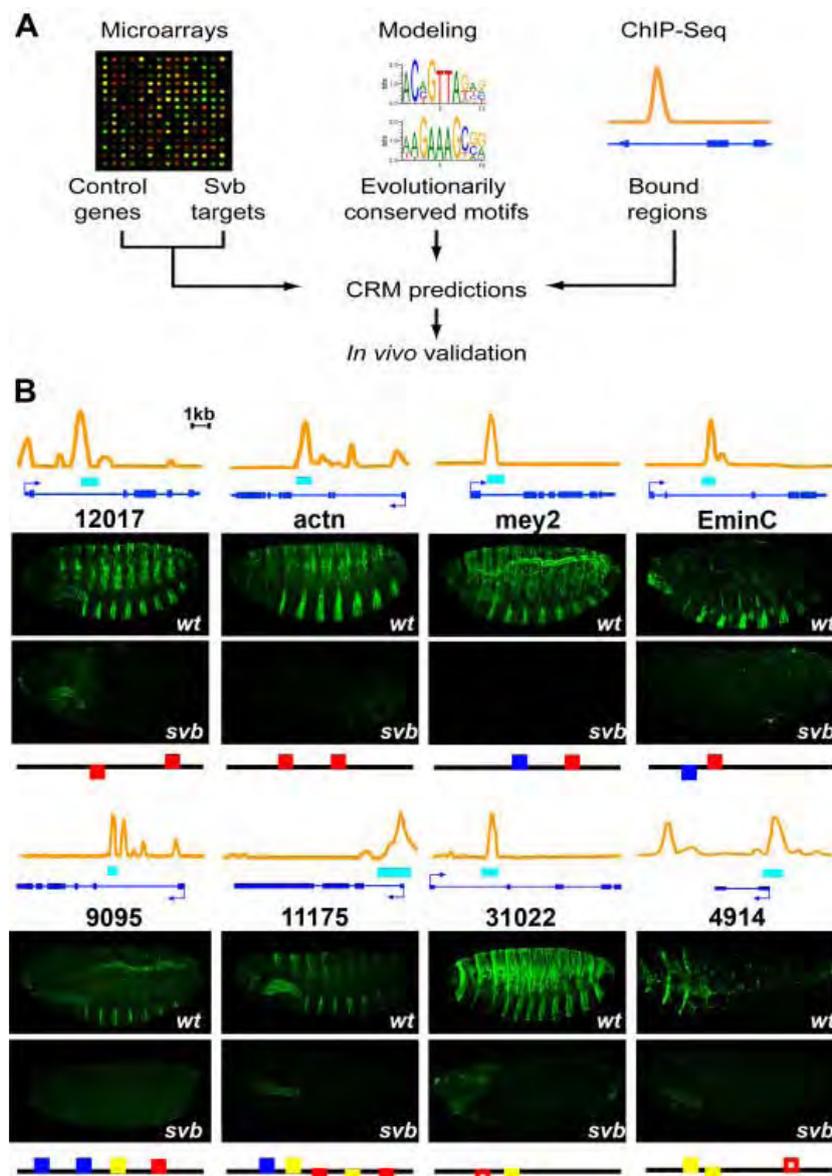
functional in the regulation of target genes [5,15]. Therefore, in order to identify the entire set of genes regulated by Svb, we performed microarray profiling, comparing wild-type to mutant embryos. In mRNA samples prepared from *svb* whole embryos, we often detected only a modest reduction in the levels of validated targets (Figure 6b; Figure S3 in Additional file 1), challenging unambiguous identification of Svb downstream genes. In the absence of *pri*, Svb behaves as a dominant repressor [34] and consistently we observed a stronger decrease in the levels of known Svb targets in *pri* mutants (Figure 6b, c; Figure S3 in Additional file 1), therefore providing an additional criterion to identify genes regulated by Svb. Henceforth, we selected the genes down-regulated in *svb* mutants and that also displayed a further (more than two-fold) reduction in their expression in *pri* mutants, as benchmarked for known Svb targets. This defined a set of 150 genes encompassing 16/39 Svb targets validated *in vivo* (Figure S1A in Additional file 1), as well as 42 additional epidermal candidates (Figure S3 in Additional file 1). Among these, we examined 23 genes by *in situ* hybridization and confirmed that 21 of them required Svb to be expressed in trichome cells (Figure 6b, c; Figure S4 in Additional file 1). These results therefore show that microarray profiling has defined a representative set of genes activated by Svb in trichome cells.

Focusing on this genomic set of Svb-regulated genes, we found 172 peaks associated with 85 genes (Figure S3 in Additional file 1), including 11 out of 14 active enhancers (Figure S7 in Additional file 1). Within the whole set of relevant Svb-bound regions, we retrieved the characteristic features of *cis*-regulatory motifs as defined previously. Although retrieved in many Svb-bound regions (Figure 6a; Figure S5 in Additional file 1), the enrichment of yellow motifs within ChIP peaks associated with Svb-regulated genes does not reliably reach a significant threshold, consistent with a broad genomic distribution [41]. In contrast, we found clear association of svbF7 motifs and to a lesser extent of blue motifs (Figure S5 in Additional file 1). Importantly, these motifs were not detected in peaks associated with a control set of genes independent of Svb (Figure S5 in Additional file 1), strongly supporting that they are hallmarks of Svb-target enhancers. As an independent way to evaluate this conclusion, we used *ab initio* analysis of ChIP peaks using PeakMotif [51]. This identified the motif ACAGTTA, which is characteristic of peaks associated with Svb downstream genes and extensively matches svbF7 (Figure S6 in Additional file 1). A second sequence (TGAAAAG), partly matching the blue motif, was also detected in about 50% of peaks, again only in Svb-regulated genes and not among control genes (Figure S6 in Additional file 1).



Hence, we interpret these results to imply that *svbF7*, and to a lesser extent the blue and/or yellow motif, would allow prediction of the location of additional trichome enhancers (Figure 7a). To evaluate this, we tested ChIPed regions containing *svbF7* alone (12017, 14395), *svbF7* in association with either the blue motif (*mey2*, *EminC*, *actn*, 12017-2) or the yellow motif (31022, 4914), or all three motifs together (9095, 11175) (Figure 7b; Figure S7 in Additional file 1). We found that 8/10

(80%) of these regions act as *Svb*-dependent enhancers when assayed *in vivo* (Figure 7b). Indeed, they drove robust expression, specifically in trichome cells, and their activity was reduced in *svb* mutant embryos (Figure 7b). Moreover, these data confirm that trichome enhancers are generally built from different combinations of the three *cis*-regulatory motifs. For example, only a subset of newly predicted trichome enhancers relies on the blue motif, since *mey2*, *EminC*, 9095 and



**Figure 7 Identification of *Svb* direct targets and their trichome enhancers using computational and *in vivo* experimental approaches.** (a) Flow diagram summarizing the pipeline used for enhancer prediction and validation. (b) Motif distribution coupled to ChIP-seq allows prediction of location of enhancers in *Svb* downstream targets. Graphs show ChIP intensity at the time of trichome formation (12 to 14 h of embryogenesis). Active enhancers are drawn as cyan rectangles. Pictures show reporter gene expression driven by corresponding regions in wild-type (*wt*) and *svb* mutant embryos, as revealed by anti-lacZ immunostaining (green). The composition, orientation and respective positioning of *svbF7* (red), blue and yellow motifs is schematized by filled (evolutionarily conserved) and open (not traceable across species) boxes.

11175 contain conserved blue motifs whereas 12017, 31022 and 4914 do not (Figure 7b; Figure S7 in Additional file 1). In the case of the *actn* enhancer, there are four partly degenerate blue motifs in the sequence from *D. melanogaster* and sibling species, while it is not retrieved in more distant species, suggesting a turnover of *cis*-regulatory motifs (Figure S8 in Additional file 1). However, aside from a couple of fast evolving enhancers, we found in many cases remarkable conservation of svbF7, blue and yellow motif patterns within individual enhancers across distantly related *Drosophila* species (Figure 8; Figure S8 in Additional file 1).

Therefore, the regulatory signatures derived from modeling and experimental dissection of a subset of enhancers helps in understanding how the Svb TF selects the genomic set of its direct targets. Furthermore, they collectively allow efficient identification of CRMs that specify the program of trichome-specific expression in response to Svb.

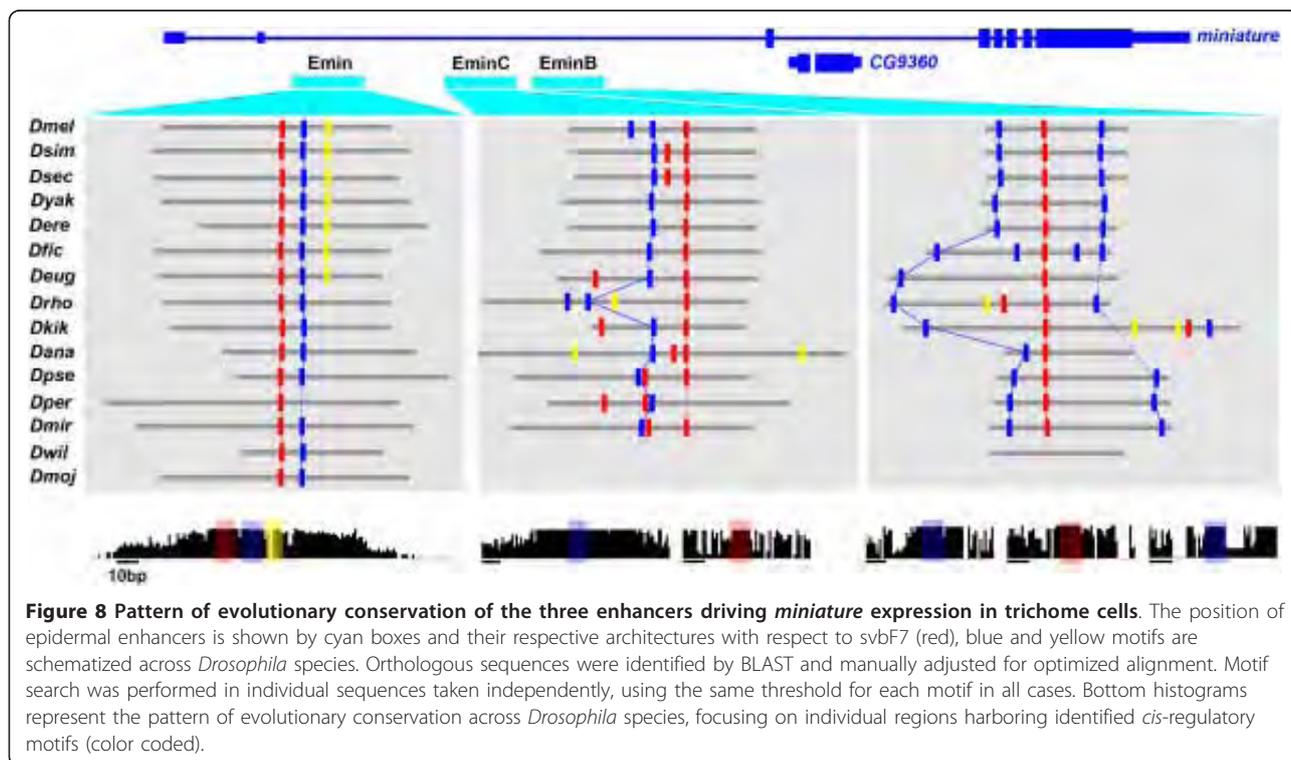
## Discussion

It is well established that the Shavenbaby TF determines trichome fate [29,32,52]; however, little was known on the repertoire of its direct target genes and mechanistic insights into the functional organization of trichome enhancers were lacking. Combining functional dissection, computational modeling and genome-wide profiling, we provide here a molecular map of the ultimate repertoire

of genes and *cis*-regulatory elements implementing the network of trichome differentiation.

## Physical elements of the GRN governing trichome formation

Our results identify a high-confidence set of more than 150 genes activated by Svb in trichome cells. We confirmed 60 of these, showing complete or partial down-regulation in the absence of active Svb protein. While most genes are expressed in all trichome cells, some are restricted to trichome subsets, suggesting that they can contribute to the diversity of trichome shape and organization observed along the body [52]. Functional annotation (Gene Ontology and manual curation) indicates that Svb controls terminal players of trichome differentiation. In addition to novel factors of F-actin organization [31,39], extracellular matrix remodeling [31,33], cuticle formation [31,38] and pigmentation [31], we identify enzymes involved in oxidation-reduction, proteolysis and cell trafficking, further extending the repertoire of cellular functions involved in the terminal differentiation of trichome cells. Hence, a major role of Svb in trichome formation is to directly activate the expression of a battery of cell morphogenesis effectors. In support of this, ChIP-seq peaks are present in >70% of these Svb-dependent effector genes. Experimental assays further validated 22 functional enhancers driving the expression of genes encoding factors involved in



cytoskeletal or extracellular matrix reorganization, sugar binding, proteolysis and additional enzymes.

Recent work has established that apparently redundant, or shadow, enhancers ensure robust expression of TFs [53,54]. For example, the transcription of *svb* itself involves separate enhancers that buffer the trichome pattern against variations in the genetic background and external conditions [53]. It has been proposed that shadow enhancers are required to drive acute expression of some key developmental regulators [55]. We define within both *shavenoid* and *miniature* separable enhancers (*sha1*, *sha3*, *Emin*, *EminB*, *EminC*) that mediate Svb regulation. These data indicate that apparently redundant enhancers may not be limited to regulatory factors operating at high hierarchic positions in gene networks. Instead, we provide evidence that several 'blue collar' effector genes display a similar regulatory architecture, suggesting that multiple enhancers represent an overlooked feature of the successive tiers of gene networks.

#### Binding site clustering as a general signature of active enhancers?

Early acting enhancers often comprise multiple BSs for a given TF [56,57]. For example, conserved BS clusters have identified target enhancers of Dorsal [13] or Bicoid [58] and feature functional Twist-bound regions [15]. Of note, most algorithms developed for enhancer detection extensively use motif clustering as an important predictor [59]. We found a clear enrichment in putative Svb BSs (OvoQ6 motif) in its downstream genes; however, only a small proportion of these motifs mediate *in vivo* regulation. There is very limited, if any, clustering of Svb BSs in ChIP peaks associated with Svb target genes, and even genome-wide. Within the trichome enhancers we validated experimentally, 13 out of 22 display a single Svb site. Furthermore, for the enhancers *tyn2*, *sha3* and *dyl2*, which contain two to three Svb BSs, the inactivation of individual sites has often limited consequences, as also reported for other TFs [60]. Even if some sites have been missed by computational approaches, the presence of multiple BSs within a short region is not a deterministic feature of active Svb-dependent enhancers.

These findings highlight a paradoxical discrepancy between the enrichment of putative BSs accumulated in Svb downstream genes and the limited number of those acting as *cis*-regulatory elements. Is there a role for this evolutionary accumulation of Svb-like motifs in Svb targets? For example, these sites with presumably weaker affinity (at least *in vivo*) can increase the local concentration of the TF facilitating regulation through a few BSs stably bound *in vivo*, as it has been suggested on thermodynamics grounds [61] or to explain the existence of thousands of binding events that are transcriptionally inactive [5,15].

#### Trichome enhancers rely on diverse combinations of *cis*-regulatory motifs

We found that the motif bound by Svb *in vivo* is more constrained than the consensus defined from *in vitro* [35] or one-hybrid approaches [48]. This shows that slight sequence differences, not detected *in vitro*, can play a key role within genomic context [62], such as revealing the influence of co-factors [63].

In addition, other motifs influence which Svb BSs are functional as regulatory elements, a notion well in line with recent results on the *in vivo* specificity of Hox factors [64]. Our statistical approaches identified a more widely spread 'blue' motif. Importantly, only half of the enhancers comprise blue motifs, indicating that there are several ways to build Svb-responsive enhancers. Indeed, the systematic dissection of *Emin* disclosed an additional motif (TTATGCAA) ultra-conserved across Drosophilidae [41] and contributing to its activity. This 'yellow' motif is retrieved in half of the trichome enhancers, with or without blue motifs. It is, however, barely specifically enriched in Svb-bound regions and therefore was not predicted by our computational analyses (positives versus negative regions), showing the importance of unbiased functional dissection to disclose the full spectrum of *cis*-regulatory elements. Indeed, the disruption of either blue or yellow motifs strongly affects enhancer function in all tested cases, providing experimental evidence of their *cis*-regulatory activity.

Trichome enhancers thus display various combinations of motifs, from those containing only Svb BSs (5/22), Svb plus yellow (4/22), Svb plus blue (6/22) or all three together (7/22). These different motif compositions do not appear to correlate with distinct subclasses of gene function (DM, unpublished data). Furthermore, multiple enhancers from the same gene can harbor distinct combinations, as exemplified by *shavenoid* and to a lesser extent by *miniature* (Figure 8; Figure S6 in Additional file 1). Several studies have shown that motif composition may correlate with a given spatio-temporal pattern - for example, for neurogenic or muscular GRNs [11,16]. Since most trichome enhancers are often active in the very same population of cells, with highly similar dynamics, it is surprising to observe such diversity in their motif compositions. There are four enhancers restricted to dorsal trichome cells, but again they accommodate different motif compositions, with *EminB* and *4702B*, which contain blue motifs, versus *cyrA* and *31559*, which do not. These data thus indicate that trichome enhancers display diverse distributions of functional motifs, supporting that distinct *cis*-regulatory architectures drive highly similar spatio-temporal expression.

### Flexibility in *cis*-regulatory motifs among enhancers versus across species

Although highly constrained sequences, such as the interferon- $\beta$  enhanceosome, do not seem widely spread [20], developmental enhancers may yet require some 'grammar' for motif positioning [23] - for example, with an optimal pair-wise spacing of motifs [64] that could reflect the cooperative binding of TFs. For trichome enhancers we did not detect any obvious bias in the number or respective arrangement of the *cis*-regulatory motifs they rely on (Figure S2 in Additional file 1). Likewise, recent results from the analysis of *Drosophila* cardiac enhancers support that similar expression patterns can be generated from divergent compositions and positioning of motifs [10,65].

That several different inputs lead to similar enhancer outputs does not, however, formally rule out the existence of constraints, even though they are not detected by 'horizontal' comparison of different enhancers within the same species. An independent way to evaluate this possibility is to look at the evolution of individual regulatory regions throughout species [15,21]. Across Drosophilidae, trichome enhancers often display similar numbers and organization of *cis*-regulatory motifs (Figure 8; Figure S6 in Additional file 1). Furthermore, besides turnover of some motifs, svbF7, blue and yellow motifs are often embedded within short-sized islands of high evolutionary conservation, when compared to neighboring sequences (Figure 8). Similar strong evolutionary conservation was also noticed for the binding site of Twist [62] and its partner TFs [15], although these studies did not examine evolution of the detailed pattern of motif positioning. These data therefore suggest that despite diverse arrangements of motifs, patterns of evolutionary conservation likely represent the signature of functional constraints that locally shape the architecture of individual enhancers.

### Materials and methods

#### Fly strains and transgenic constructs

We used *btd*, *svb*<sup>1</sup> or *svb*<sup>R9</sup> [30,31] and *pri*<sup>1</sup> [34] stocks kept over green fluorescent protein (GFP) balancers. To delineate the epidermal enhancer of *sn* and *sha*, transgenic lines were initially generated using P-element-mediated transformation (Fly Facility) and at least three independent insertions were analyzed for each construct. We then switched to the PhiC31 system (Bestgene, Chino Hills, CA, USA) to quantify effects of mutations, with all constructs integrated at the same location (*zh-86F*), except for *sha1*, *sha3* and *snE1*, for which mutant versions were assayed in P-elements for homogeneity (Additional file 3). Genomic regions were amplified and cloned into pCasper or pAttB lacZ derivatives. QuikChange II XL site-directed mutagenesis (Agilent Technologies, Santa Clara, CA, USA) was used to introduce point mutations in enhancers, or

CCGCCGGCGG stretches for linker scanning of *Emin*. All constructs were verified by sequencing. Details (genomic position) of the CRM are given in Additional file 3.

#### Embryo staining

Dig- or biotin-labeled antisense RNA probes were used for *in situ* hybridization following standard protocols and embryos imaged using a Nikon Eclipse90i microscope. For immunodetection of lacZ reporter expression, 10- to 14-h embryos were stained using anti- $\beta$ -galactosidase (1/1,000; Cappel, MP Biomedical, Solon, OH USA) and Alexafluor488 (Molecular Probes, Life Technologies, Carlsbad, CA, USA). Pictures were taken with a Leica SP2 confocal microscope, using the same settings to allow quantitative comparisons.

#### Microarrays

We hand selected 13- to 15-h *svb*<sup>R9</sup> or *pri*<sup>1</sup> embryos using GFP balancers. We subjected 200 embryos to trizol (Invitrogen, Life Technologies, Carlsbad, CA, USA) extraction and RNA quality was monitored using Agilent Chip. Five independent samples of each genotype were used for microarrays (Affymetrix, Santa Clara, CA, USA; IGBMC, Strasbourg, France). Data extraction and normalization were performed using Affymetrix software and statistical analyses with R. A more than two-fold difference in expression levels between mutant genotypes was the most efficient criterion to retrieve Svb downstream genes (with a false discovery rate of 0.01 for *pri*). The top 150 genes down-regulated in both *pri* and *svb* mutants defined the set of Svb-regulated genes. One-hundred genes showing irrelevant variation of their expression (*P*-value > 0.8, false discovery rate >0.99) were used as a negative control set. The data discussed in this article have been deposited in NCBI's Gene Expression Omnibus [66] and are accessible through GEO series accession number GSE48997. Details are given in Additional file 3.

#### ChIP-seq

A *svb* rescue construct (RSQ8) [34] was used for ChIP-seq experiments. It expresses a Svb-GFP protein under the control of two *svb* enhancers (medial and proximal) driving specific expression in epidermal trichome cells. Stocks were expanded to fill three population cages. Adults were allowed to lay eggs for 2 hours on apple juice plates covered with yeast. Embryos deposited on the plates were aged for 12 h at 25°C. Chromatin was collected from approximately 100 mg of whole embryos for each replicate chromatin collection. ChIP was done with an anti-GFP antibody as described [8]. Data presented are from two independent replicates. Peaks were called for single replicates using MACS *P* < 0.00001 for downstream computational analyses. MACS was used to call loose criteria peaks for two replicates of RSQ8 12- to 14-h embryos. Those

peaks were then used for an IDR (Irreproducible Discovery Rate) analysis (IDR = 0.02). DNA sequencing libraries were generated with Nextera DNA Sequencing Library kits (VWR, Radnor, PA, USA). Details are given in Additional files 3 and 4. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [66] and are accessible through GEO series accession number GSE48791.

### Motif detection and genome analysis

Detection of motifs enriched in Svb-dependent and Svb-independent epidermal genes was performed using cis-TargetX [40]. For *de novo* motif discovery, genomic sequences of enhancer and negative regions were processed through a C++ program and statistical operations performed within the R software, as described [47]. To compute the cross-correlation between conserved motif instances and Svb ChIP-Seq data, we defined a 10 kb window centered around each ChIP peak, collected distances of each motif to the peak center and plotted these values using a 500 bp bin. In the cases of Svb-regulated and control genes, each ChIP peak was associated with the nearest transcription start site. Further details are available in Additional files 3 and 4.

### Additional material

**Additional file 1: Supplementary figures.**

**Additional file 2: Legends to the supplementary figures.**

**Additional file 3: Supplementary information (details of experimental procedure, constructs, and so on).**

**Additional file 4: bed files (BS prediction, Chip-seq).**

### List of abbreviations used

bp: base pair; BS: binding site; ChIP: chromatin immunoprecipitation; CRM: cis-regulatory module; GFP: green fluorescent protein; GRN: gene regulatory network; TF: transcription factor.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SA, FP and SP designed the experiments, MS, HR and VH statistical analyses; DM, MS, IF, RS, JZ, IG, YL, PF, SA and SP performed the experiments and are listed according to their contributions. SP and FP wrote the paper and all authors commented on the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We are grateful to the Bloomington *Drosophila* Stock Center and *Drosophila* Genomic Resource Center for providing us with flies and molecular clones. We are indebted to B Ronsin (Toulouse RIO Imaging), P Valenti and O Bohner for excellent technical assistance. We also thank C Hermann, N Negre, E Preger-Ben Noon, A Vincent and members of the SP/FP lab for critical reading of the manuscript. This work was supported by grants from ANR Blanc 'Netoshape', Association pour la Recherche contre le Cancer (n°3832, n°1111, SFI20101201669 and fellowships to IF and DM) and University Paul Sabatier (cisdecode). SA was supported by grants from FWO (G.0704.11N and G.0640.13) and HFSP (RGY0070/2011).

### Authors' details

<sup>1</sup>Centre de Biologie du Développement, Université de Toulouse, 118 route de Narbonne, Toulouse, F-31062, France. <sup>2</sup>CNRS UMR5547, 118 route de Narbonne, Toulouse, F-31062, France. <sup>3</sup>Laboratoire de Physique Statistique, CNRS, Université Pierre & Marie Curie, Université Denis Diderot, ENS, 24, rue Lhomond, F-75231 Paris, France. <sup>4</sup>Current address: Department of Biology, McGill University, 845 Sherbrooke St W, Montreal, QC H3A 0G4, Canada. <sup>5</sup>Institute for Genomics and Systems Biology, Department of Human Genetics, The University of Chicago, 57th St Room, 10100 Chicago, IL 60637, USA. <sup>6</sup>Université de Toulouse, INSA, 118 route de Narbonne, Toulouse, F-31062, France. <sup>7</sup>Institut de Mathématiques, CNRS UMR5219, 118 route de Narbonne, Toulouse, F-31062, France. <sup>8</sup>Current address: Howard Hughes Medical Institute, Janelia Farm Research Campus, 19700 Helix Dr., Ashburn, VA 20147, USA. <sup>9</sup>Laboratory of Computational Biology, Center for Human Genetics KU Leuven, O&N I Herestraat 49 - 3000 Leuven, Belgium.

Received: 7 March 2013 Revised: 10 June 2013

Accepted: 23 August 2013 Published: 23 August 2013

### References

1. Stathopoulos A, Levine M: Genomic regulatory networks and animal development. *Dev Cell* 2005, **9**:449-462.
2. Ptashne M: Regulation of transcription: from lambda to eukaryotes. *Trends Biochem Sci* 2005, **30**:275-279.
3. Rister J, Desplan C: Deciphering the genome's regulatory code: the many languages of DNA. *Bioessays* 2010, **32**:381-384.
4. Yanez-Cuna JO, Kvon EZ, Stark A: Deciphering the transcriptional cis-regulatory code. *Trends Genet* 2013, **29**:11-22.
5. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP, Eisen MB, Biggin MD: Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* 2008, **6**:e27.
6. Slattery M, Negre N, White KP: Interpreting the regulatory genome: the genomics of transcription factor function in *Drosophila melanogaster*. *Brief Funct Genomics* 2012, **11**:336-346.
7. MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keranen SV, Knowles DW, Stapleton M, Bickel P, Biggin MD, Eisen MB: Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* 2009, **10**:R80.
8. Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R, Li Z, Ishii H, Spokony RF, Chen J, Hwang L, Cheng C, Auburn RP, Davis MB, Domanus M, Shah PK, Morrison CA, Zieba J, Suchy S, Senderowicz L, Victorsen A, Bild NA, Grundstad AJ, Hanley D, MacAlpine DM, Mannervik M, et al: A cis-regulatory map of the *Drosophila* genome. *Nature* 2011, **471**:527-531.
9. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U: Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 2008, **451**:535-540.
10. Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EE: A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 2012, **148**:473-486.
11. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE: Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 2009, **462**:65-70.
12. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M: Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev* 2007, **21**:385-390.
13. Markstein M, Markstein P, Markstein V, Levine MS: Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci USA* 2002, **99**:763-768.
14. Sinha S, Schroeder MD, Unnerstall U, Gaul U, Siggia ED: Cross-species comparison significantly improves genome-wide prediction of cis-regulatory modules in *Drosophila*. *BMC Bioinformatics* 2004, **5**:129.
15. He Q, Bardet AF, Patton B, Purvis J, Johnston J, Paulson A, Gogol M, Stark A, Zeitlinger J: High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nat Genet* 2011, **43**:414-420.

16. Markstein M, Zinzen R, Markstein P, Yee KP, Erives A, Stathopoulos A, Levine M: **A regulatory code for neurogenic gene expression in the *Drosophila* embryo.** *Development* 2004, **131**:2387-2394.
17. Khoueiry P, Rothbacher U, Ohtsuka Y, Daian F, Frangulian E, Roure A, Dubchak I, Lemaire P: **A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites.** *Curr Biol* 2010, **20**:792-802.
18. Erives A, Levine M: **Coordinate enhancers share common organizational features in the *Drosophila* genome.** *Proc Natl Acad Sci USA* 2004, **101**:3851-3856.
19. Levine M: **Transcriptional enhancers in animal development and evolution.** *Curr Biol* 2010, **20**:R754-763.
20. Spitz F, Furlong EE: **Transcription factors: from enhancer binding to developmental control.** *Nat Rev Genet* 2012, **13**:613-626.
21. Crocker J, Erives A: **A closer look at the eve stripe 2 enhancers of *Drosophila* and *Themira*.** *PLoS Genet* 2008, **4**:e1000276.
22. Papatsenko D, Goltsev Y, Levine M: **Organization of developmental enhancers in the *Drosophila* embryo.** *Nucleic Acids Res* 2009, **37**:5665-5677.
23. Swanson CI, Evans NC, Barolo S: **Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer.** *Dev Cell* 2010, **18**:359-370.
24. Rowan S, Siggers T, Lachke SA, Yue Y, Bulyk ML, Maas RL: **Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity.** *Genes Dev* 2010, **24**:980-985.
25. Papatsenko D, Levine M: **A rationale for the enhanceosome and other evolutionarily constrained enhancers.** *Curr Biol* 2007, **17**:R955-957.
26. Crocker J, Tamori Y, Erives A: **Evolution acts on enhancer organization to fine-tune gradient threshold readouts.** *PLoS Biol* 2008, **6**:e263.
27. Etchberger JF, Lorch A, Sleumer MC, Zapf R, Jones SJ, Marra MA, Holt RA, Moerman DG, Hobert O: **The molecular signature and cis-regulatory architecture of a *C. elegans* gustatory neuron.** *Genes Dev* 2007, **21**:1653-1674.
28. Laurencon A, Dubrulle R, Efimenko E, Grenier G, Bissett R, Cortier E, Rolland V, Swoboda P, Durand B: **Identification of novel regulatory factor X (RFX) target genes by comparative genomics in *Drosophila* species.** *Genome Biol* 2007, **8**:R195.
29. Payre F, Vincent A, Carreno S: **ovo/svb integrates Wingless and DER pathways to control epidermis differentiation.** *Nature* 1999, **400**:271-275.
30. Sucena E, Delon I, Jones I, Payre F, Stern DL: **Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism.** *Nature* 2003, **424**:935-938.
31. Chanut-Delalande H, Fernandes I, Roch F, Payre F, Plaza S: **Shavenbaby couples patterning to epidermal cell shape control.** *PLoS Biol* 2006, **4**:e290.
32. Chanut-Delalande H, Ferrer P, Payre F, Plaza S: **Effectors of tridimensional cell morphogenesis and their evolution.** *Semin Cell Dev Biol* 2012, **23**:341-349.
33. 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.
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. 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.
36. Lu J, Oliver B: ***Drosophila* OVO regulates ovarian tumor transcription by binding unusually near the transcription start site.** *Development* 2001, **128**:1671-1686.
37. Brown CD, Johnson DS, Sidow A: **Functional architecture and evolution of transcriptional elements that drive gene coexpression.** *Science* 2007, **317**:1557-1560.
38. Andrew DJ, Baker BS: **Expression of the *Drosophila* secreted cuticle protein 73 (dsc73) requires Shavenbaby.** *Dev Dyn* 2008, **237**:1198-1206.
39. Bejsovec A, Chao AT: **crinkled reveals a new role for Wingless signaling in *Drosophila* denticle formation.** *Development* 2012, **139**:690-698.
40. Aerts S, Quan XJ, Claeys A, Naval Sanchez M, Tate P, Yan J, Hassan BA: **Robust target gene discovery through transcriptome perturbations and genome-wide enhancer predictions in *Drosophila* uncovers a regulatory basis for sensory specification.** *PLoS Biol* 2010, **8**:e1000435.
41. Elemento O, Tavazoie S: **Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach.** *Genome Biol* 2005, **6**:R18.
42. Mace KA, Pearson JC, McGinnis W: **An epidermal barrier wound repair pathway in *Drosophila* is mediated by grainy head.** *Science* 2005, **308**:381-385.
43. Szuplewski S, Kottler B, Terracol R: **The *Drosophila* bZIP transcription factor Vrille is involved in hair and cell growth.** *Development* 2003, **130**:3651-3662.
44. Jayo A, Parsons M: **Fascin: a key regulator of cytoskeletal dynamics.** *Int J Biochem Cell Biol* 2012, **42**:1614-1617.
45. Frith MC, Li MC, Weng Z: **Cluster-Buster. Finding dense clusters of motifs in DNA sequences.** *Nucleic Acids Res* 2003, **31**:3666-3668.
46. Kim J, Cunningham R, James B, Wyder S, Gibson JD, Niehuis O, Zdobnov EM, Robertson HM, Robinson GE, Werren JH, Sinha S: **Functional characterization of transcription factor motifs using cross-species comparison across large evolutionary distances.** *PLoS Comput Biol* 2010, **6**:e1000652.
47. 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 USA* 2010, **107**:14615-14620.
48. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, Brasefield JA, Zhu C, Asriyan Y, Lapointe DS, Sinha S, Wolfe SA, Brodsky MH: **FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system.** *Nucleic Acids Res* 2011, **39**:D111-117.
49. Frankel N, Erezylmaz DF, McGregor AP, Wang S, Payre F, Stern DL: **Morphological evolution caused by many subtle-effect substitutions in regulatory DNA.** *Nature* 2011, **474**:598-603.
50. Herrmann C, Van de Sande B, Potier D, Aerts S: **i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules.** *Nucleic Acids Res* 2012, **40**:e114.
51. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J: **RSAT 2011: regulatory sequence analysis tools.** *Nucleic Acids Res* 2011, **39**:W86-91.
52. Payre F: **Genetic control of epidermis differentiation in *Drosophila*.** *Int J Dev Biol* 2004, **48**:207-215.
53. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL: **Phenotypic robustness conferred by apparently redundant transcriptional enhancers.** *Nature* 2010, **466**:490-493.
54. Perry MW, Boettiger AN, Bothma JP, Levine M: **Shadow enhancers foster robustness of *Drosophila* gastrulation.** *Curr Biol* 2010, **20**:1562-1567.
55. Hong JW, Hendrix DA, Levine MS: **Shadow enhancers as a source of evolutionary novelty.** *Science* 2008, **321**:1314.
56. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, Eisen MB: **Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome.** *Proc Natl Acad Sci USA* 2002, **99**:757-762.
57. Schroeder MD, Pearce M, Fak J, Fan H, Unnerstall U, Emberly E, Rajewsky N, Siggia ED, Gaul U: **Transcriptional control in the segmentation gene network of *Drosophila*.** *PLoS Biol* 2004, **2**:E271.
58. Chen H, Xu Z, Mei C, Yu D, Small S: **A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes.** *Cell* 2012, **149**:618-629.
59. Aerts S: **Computational strategies for the genome-wide identification of cis-regulatory elements and transcriptional targets.** *Curr Top Dev Biol* 2012, **98**:121-145.
60. Halfon MS, Zhu Q, Brennan ER, Zhou Y: **Erroneous attribution of relevant transcription factor binding sites despite successful prediction of cis-regulatory modules.** *BMC Genomics* 2011, **12**:578.
61. Walter J, Biggin MD: **DNA binding specificity of two homeodomain proteins in vitro and in *Drosophila* embryos.** *Proc Natl Acad Sci USA* 1996, **93**:2680-2685.
62. Ozdemir A, Fisher-Aylor KI, Pepke S, Samanta M, Dunipace L, McCue K, Zeng L, Ogawa N, Wold BJ, Stathopoulos A: **High resolution mapping of Twist to DNA in *Drosophila* embryos: Efficient functional analysis and evolutionary conservation.** *Genome Res* 2011, **21**:566-577.
63. Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, Mann RS: **Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins.** *Cell* 2011, **147**:1270-1282.
64. Sorge S, Ha N, Polychronidou M, Friedrich J, Bezdan D, Kaspar P, Schaefer MH, Ossowski S, Henz SR, Mundorf J, Ratzer J, Papagiannouli F,

Lohmann I: **The cis-regulatory code of Hox function in *Drosophila*.** *EMBO J* 2012, **31**:3323-3333.

65. Jin H, Stojnic R, Adryan B, Ozdemir A, Stathopoulos A, Frasch M: **Genome-wide screens for in vivo tinman binding sites identify cardiac enhancers with diverse functional architectures.** *PLoS Genet* 2013, **9**:e1003195.
66. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A: **NCBI GEO: archive for functional genomics data sets—update.** *Nucleic Acids Res* 2013, **41**: D991-995.

doi:10.1186/gb-2013-14-8-r86

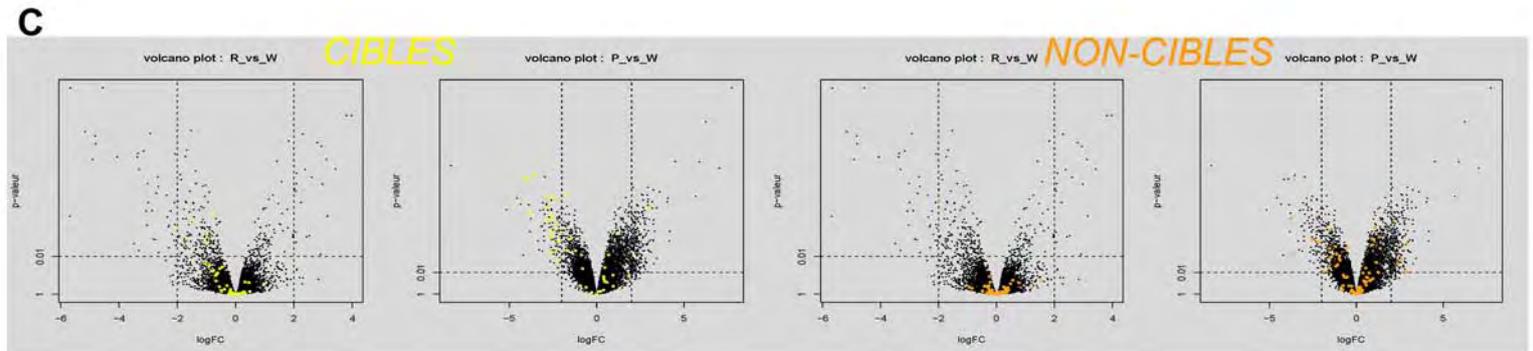
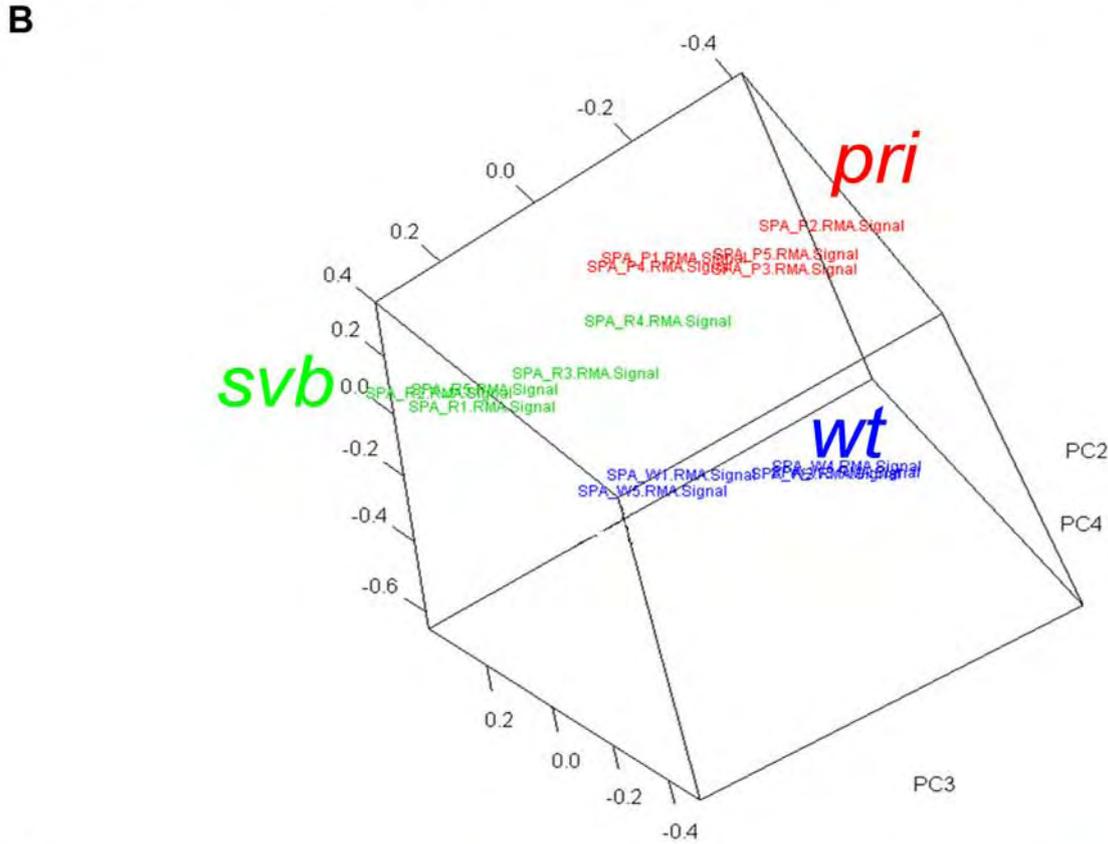
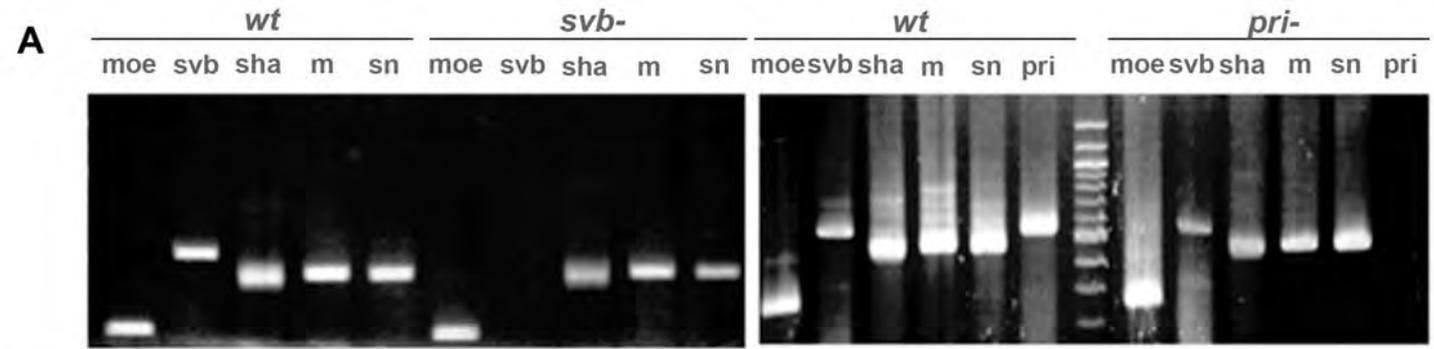
**Cite this article as:** Menoret *et al.*: Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization. *Genome Biology* 2013 **14**:R86.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)





**Figure 17. Qualité des échantillons utilisés pour le transcriptome.**

**A. Diminution d'expression des effecteurs de *Svb* dans les embryons *svb* et *pri*.**

Après extraction des ARNs totaux, un aliquot est testé par RT-PCR pour des effecteurs de *Svb* : *sha*, *m*, *sn*. La *moesine* est utilisé comme contrôle négatif. *svb* n'est pas exprimé dans les embryons triés *svb*- et l'expression de ses effecteurs est diminuée. *pri* n'est pas exprimé dans les embryons triés *pri*-, et l'expression des effecteurs de *Svb* est diminuée.

**B. Analyse en composante principale (ACP).**

Une ACP sans dimension définie permet de regrouper à l'aveugle les échantillons par génotype.

**C. Dispersion des effecteurs de *Svb* (cibles) et gènes indépendants de *Svb* (non cibles).**

Volcano plot montrant le log(FC) et la p-value de tous les gènes de *Drosophila melanogaster*. Parmi eux certains effecteurs de *Svb* (jaunes) sont dérégulés dans les embryons *svb*. La condition *pri* permet d'obtenir plus d'effecteurs dérégulés. Les non cibles (orange) ne sont pas dérégulés.

## RESULTATS ADDITIONNELS ET DISCUSSION

Bien que cet article de Genome Biology représente un corpus de données assez conséquent, le format contraint de la publication ne permet cependant pas d'inclure le jeu complet des différentes données qui y sont naturellement associées. Il nous a semblé opportun de compléter cet article par des détails (ou choix) méthodologiques et des résultats complémentaires, ainsi que de discuter plus librement leur portée et leur limite.

### Réalisation et validation du transcriptome.

Pour établir une liste « exhaustive » des gènes cibles de Svb au cours de la différenciation de l'épiderme, nous avons réalisé une approche globale d'analyse des variations du transcriptome en comparant des embryons de différents contextes génétiques : sauvage, mutant pour *svb* et mutant pour *pri*.

La connaissance de la cinétique d'activation des cibles identifiées de Svb a permis de définir précisément la fenêtre temporelle de développement (13-15h) à analyser. Pour ces expériences, j'ai extrait les ARNm d'embryons sauvages, ou homozygotes mutants sélectionnés à la main grâce à un balancer GFP (200 embryons par réplicat), en effectuant 5 réplicats indépendants. J'ai validé la qualité des échantillons par RT-PCR en montrant que les ARN extraits d'embryons *svb* ou *pri* présentent bien l'absence des ARNm *svb* et *pri*, respectivement (Fig 17A). J'ai également montré une diminution d'expression des cibles de Svb : *m*, *dyl*, *sha* ou *sn* dans les mutants *svb* et *pri* (Fig 17A). Les ARNs ont ensuite été traités à Strasbourg (IGBMC), rétro-transcrits en ADNc biotinylés et hybridés sur puces à ADN Affymetrix. Nous avons analysé les données normalisées par le logiciel MAS 5.0.

Afin d'évaluer la qualité des données microarrays, notre premier test fut une analyse en composante principale (ACP), permettant de tester "à l'aveugle" la qualité des réplicats.



L'analyse en composantes principales (ACP) est une analyse statistique multidimensionnelle de réduction de dimension. Elle résume les variables initiales par un ensemble réduit de variables dites principales, de plus grande variance, combinaisons linéaires des variables initiales et deux à deux dé-corrélées. Elle fournit également des représentations graphiques ou nuages de points en petite dimension (2 ou 3) respectant "au mieux" les distances entre individus de l'échantillon. Ces analyses ont été réalisées avec le logiciel R et grâce à l'aide précieuse d'Ignacio Gonzalez (Institut de Mathématiques de Toulouse). L'ACP montre que, dans l'espace des trois premières variables principales qui expliquent 83% de la variance totale (figure 17B), les cinq réplicats des trois différents génotypes se regroupent de manière très homogène. Ceci illustre bien l'hypothèse selon laquelle les variations des expressions des gènes dépendent directement et globalement de l'activité des régulateurs *Svb* et *Pri*.

*Premières validations basées sur les variations d'expression des cibles connues de Svb*

Le deuxième test de nos données de puces à ADN a été la mesure des niveaux d'expression des ARNm *svb* et *pri*. Chaque gène présente bien une diminution drastique de son expression dans la condition mutante correspondante (Fig 18A). De plus, l'absence de *svb* n'influence pas le niveau d'expression de *pri*, et celle de *pri* n'a pas d'effet sur l'expression de *svb* (Fig 18A). Ces résultats corroborent ceux établis précédemment par hybridation *in situ* (Kondo, Hashimoto et al. 2007), validant que les deux gènes sont transcriptionnellement indépendants l'un de l'autre.

Les gènes cibles de *Svb* *mey*, *sn* et *m* présentent une diminution significative de leur expression en condition mutante *svb* ou *pri* par rapport à la condition sauvage (Fig 17C et 18A). Nous avons cependant observé des différences notables entre les résultats de transcriptome et d'hybridation *in situ* concernant d'autres gènes cibles connus de *Svb*. Par exemple, alors que l'absence de *svb* entraîne une perte du signal d'hybridation *in situ* de l'ARNm *dyl* dans l'épiderme, cette variation n'est pas détectée dans les puces. Cependant, on



observe une diminution significative de l'expression de *dyl* en absence de *pri* (Fig 18A). Pour *sox21b* ou *CG10175*, nous n'observons pas de modification de leur expression ni en condition mutante *svb*, ni pour *pri* (Fig 18B). Ces résultats soulignent sans doute la limitation due à l'analyse d'échantillons extraits d'embryons entiers. Pour des gènes exprimés dans différents tissus embryonnaires, la diminution de leur expression uniquement dans les cellules à trichomes (probablement quelques % des cellules de l'embryon) peut ne pas être suffisante pour induire une variation quantitative significative du niveau global d'ARNm, telle qu'évaluée à partir de l'embryon entier.

Nous observons que les gènes *mey*, *tyn*, *m* présentent une diminution d'expression plus forte dans des embryons *pri* par rapport aux embryons *svb*. Les gènes cibles de Svb *ChLD3*, *kkv*, *CG31559*, *CG4702*, *f*, *tyn* ou *dyl* ne présentent pas de diminution significative de leur expression en condition *svb* mais, comme pour *dyl*, présentent une baisse significative en absence de *pri* (Fig 18B). En effet, en absence des peptides Pri, Svb reste sous forme répresseur actif (Kondo, Plaza et al. 2010). L'expression dans les cellules à trichome de cette forme « dominante négative » peut permettre d'expliquer les résultats des puces, c'est à dire une plus forte baisse d'expression en présence du répresseur actif (mutants *pri*) qu'en absence de la forme activateur (mutants *svb*). Le comportement de ce premier ensemble de gènes cibles connus de Svb permet en quelque sorte d'effectuer un « benchmarking » des données de puce, montrant la nécessité d'un traitement des valeurs brutes d'expression afin de définir les critères optimaux pour identifier de nouvelles cibles potentielles.

*Des seuils sont indispensables pour écarter des candidats biologiquement non-relevants.*

Certains candidats présentent une forte variation de Fold Change (FC), et semblent donc être de nouvelles cibles intéressantes (Fig 18C, flèches violettes). Cependant la plupart de ces gènes présentent un niveau d'expression sauvage très faible, et une expression encore plus faible dans une condition mutante (souvent inférieure à 1), comme le montrent les

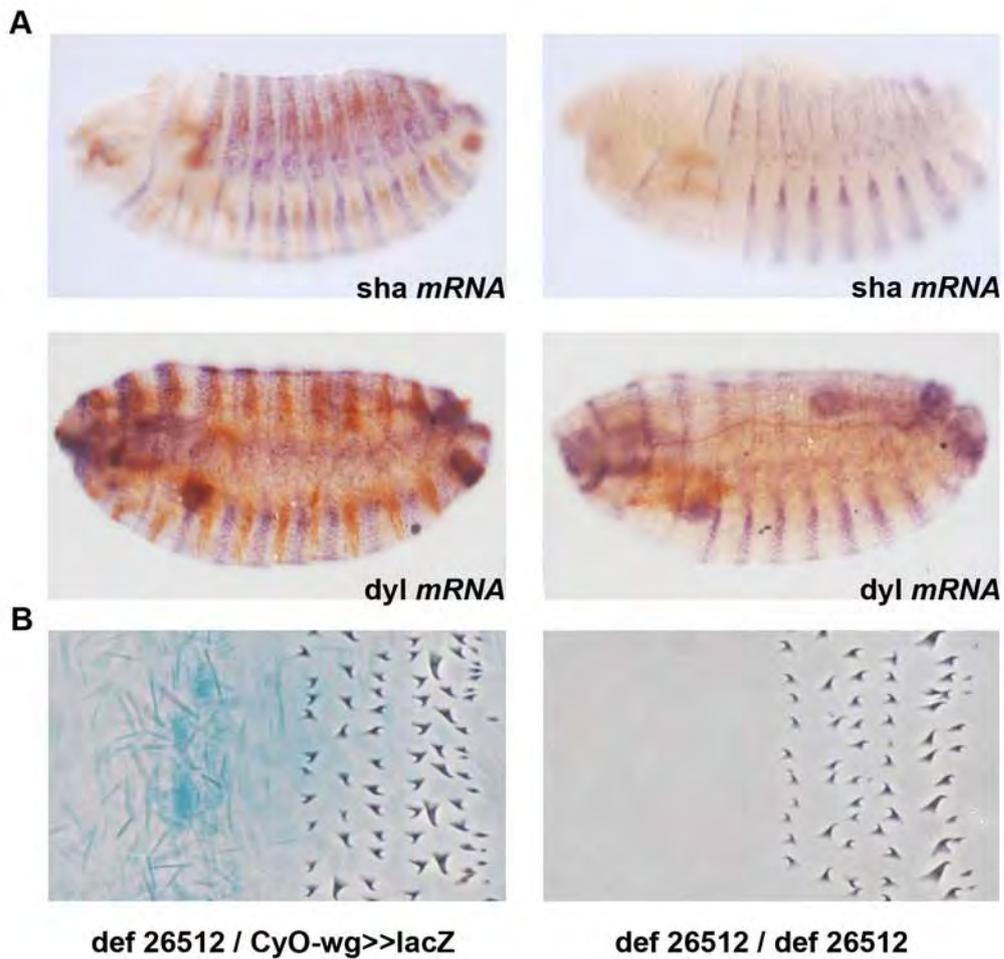


quantifications pour *CG12506* comparées à celles de *sha* (Fig 18D). *sha* est l'une des cibles les mieux caractérisées pour Svb, et son expression est drastiquement diminuée en hybridation *in situ* en absence de *svb*. Ce gène a également l'avantage d'être exprimé spécifiquement dans l'épiderme, et présente le niveau d'expression le plus faible parmi l'ensemble des effecteurs de Svb, donnant ainsi une idée du niveau d'expression que l'on peut visualiser dans l'épiderme par hybridation *in situ*. J'ai donc sélectionné le niveau d'expression de *sha* comme limite inférieure à imposer à l'ensemble des effecteurs potentiels de Svb. L'utilisation de ce seuil permet de s'affranchir de gènes candidats dont il sera difficile de vérifier si les variations d'expression détectées dans les puces présentent un sens biologique dans notre système.

L'attribution d'un seuil ne sera utilisée que pour définir des gènes dont l'expression est diminuée dans une condition mutante par rapport à la condition sauvage. En effet, il est possible que des gènes peu ou pas exprimés dans la situation sauvage soient effectivement sur-exprimés en contexte mutant.

#### *Des critères optimisés pour rechercher des cibles directes de Svb*

L'analyse de la grammaire associée aux cibles de Svb nécessitait de travailler avec une liste de candidats contenant un maximum de cibles directes potentielles. Au vu des analyses précédentes, il était clair qu'une significativité importante pour la modulation observée dans le transcriptome *svb* n'était pas un gage d'enrichissement en gènes cibles connus, montrant les limites de l'utilisation du transcriptome *svb*. Nous lui avons donc associé le transcriptome *pri*, car les cibles connues présentent une diminution d'expression dans le transcriptome *pri* au moins 2 fois supérieure à celle observée dans le transcriptome *svb*. Ces critères ont permis de sélectionner les candidats présentant des modulations similaires à celles des cibles connues. Un filtre à 0.01 a ensuite été choisi pour la pvalue du transcriptome *pri*, ce critère permettant d'obtenir une liste optimisée contenant un maximum de cibles connues : 14 parmi 150 candidats. Cette liste contient 42 gènes au patron



**Figure 19. Le FT Ken&Barbie ne semble pas participer à la morphogénèse des denticules.**

**A. Implication de *Ken&Barbie* sur l'expression des gènes cibles.**

Hybridation *in situ* des effecteurs *sha* et *dyl* dans des embryons déficients pour *ken&barbie*. Une immunocoloration lacZ (marron) permet de génotyper les embryons hétérozygotes et homozygotes. La déficience homozygote n'affecte pas l'expression des deux gènes cibles .

**B. Implication de *Ken&Barbie* sur la mise en place des denticules.**

Cuticule de larves déficientes pour *ken&barbie*. Une coloration beta-gal (bleue) permet de génotyper les larves hétérozygotes et homozygotes. La déficience n'affecte pas la morphogénèse des denticules des larves homozygotes.

d'expression épidermique (soit 28%). 23 d'entre eux ont été testés par hybridation *in situ*, et 21 dépendent de Svb pour être exprimés dans l'épiderme, soit une efficacité de 91%.

### Vers l'identification des cofacteurs de Svb

Par l'utilisation de l'ensemble des gènes cibles de Svb nous avons pu redéfinir la matrice PWM associée à Svb, cette matrice étant plus contrainte que la matrice OvoQ6 définie *in vitro* (Lu, Andrews et al. 1998; Lee and Garfinkel 2000; Bielinska, Lu et al. 2005). Nous montrons par ailleurs que les bases additionnelles de la nouvelle matrice sont effectivement importantes pour l'activité du CRM. Cette différence pourrait être la marque de l'action de cofacteurs associés à Svb pour activer la transcription de ses gènes cibles.

Nous avons également identifié une seconde matrice enrichie dans les CRMs dépendants de Svb, dite Blue Motif. En analysant cette matrice avec des algorithmes de comparaison de séquence (MEME), nous avons identifié des FTs candidats présentant une PWM proche du Blue Motif. Le premier candidat est le FT à doigt de zinc *Ken and Barbie* (*ken*). Des adultes mutants pour *ken* présentent des phénotypes de malformation des terminalia (généralia et analia) (Lukacsovich, Yuge et al. 2003). Au cours de l'embryogenèse, *ken* est exprimé dans les cellules épidermiques aux stades 13-16 (BDGP et (Arbouzova, Bach et al. 2006)). Le FT Ken reconnaît la séquence GAGAAAG, donc le corps GAAAG est indispensable à la fixation. *ken* interagit avec la voie JAK/STAT et régule certains effecteurs de cette voie (Arbouzova, Bach et al. 2006). Cependant l'expression des cibles de Svb (*sha*, *dyl*, *sn*) pour lesquelles le CRM présente un ou plusieurs Blue Motif n'est pas altérée dans des embryons homozygotes pour une déficience qui délète entièrement le gène *ken* (Fig 19A). Cette même déficience n'a pas d'effet sur la morphogénèse des denticules (Fig 19B).

De plus, l'expression d'un dsRNA dirigé contre *ken* dans les rangées 2 et 3 des cellules à denticules de l'épiderme embryonnaire (via le driver *patched-Gal4* (*ptc-Gal4*))



n'altère pas le patron des trichomes (Ahmad Alsawadi). Ensemble, ces données suggèrent donc que *ken* ne participe pas au processus de morphogénèse épidermique.

#### *Un candidat attractif pour lier le Blue Motif*

Un second candidat possible est le FT Blimp-1. Durant l'embryogenèse, Blimp-1 est exprimé dans l'épiderme et les trachées (Yvan Latapie et BDGP). Au cours de la métamorphose, il est bien établi que l'expression de Blimp-1 dépend de l'activation de la voie de l'hormone stéroïde 20-hydroxyecdysone (Chavoshi, Moussian et al. 2010). En effet, le FT Blimp-1 se fixe sur les régions de contrôle de l'expression du gène  $\beta$ ftz-f1, via le motif GAAAGCGAAAG (Agawa, Sarhan et al. 2007). Blimp-1 réprime l'expression de  $\beta$ ftz-f1, codant lui-même un FT, pour réguler temporellement la réponse transcriptionnelle à l'ecdysone (Agawa, Sarhan et al. 2007). Dans l'embryon, l'absence de Blimp-1 affecte la différenciation des trachées (Ng, Yu et al. 2006). L'expression d'un dsRNA dirigé contre Blimp-1 dans le domaine *ptc* entraîne un défaut de morphologie des denticules des rangées 2 et 3. De plus, on observe dans ces cellules une répression de l'activité du CRM Emin (Ahmad Alsawadi, données non publiées), un résultat en accord avec le rôle du Blue Motif pour l'activité d'Emin. Ces données soutiennent donc que Blimp-1 puisse être un (le) cofacteur de Svb qui reconnaît le Blue Motif. Cette hypothèse sera testée par immuno-précipitation du fragment Emin *wt* ou muté pour le Blue Motif. On peut imaginer que la fixation du cofacteur sur le Blue Motif nécessite au préalable la fixation de Svb sur son propre site, un modèle compatible avec le ChIP-seq qui montre une fixation *in vivo* de Svb dès les stades 11-12 (avant le début de l'expression épidermique, qui intervient seulement 2-4h plus tard). L'immuno-précipitation de chromatine sera également effectuée avec le fragment Emin muté pour SvbF7.

Blimp-1 est donc une cible de la voie ecdysone. Son rôle dans l'activité du CRM Emin suggère que cette voie participe à la régulation des effecteurs de Svb. Jérôme Bohère, lors de



son stage de Master, a montré que l'expression de *pri* dépend de la voie ecdysone. Ces données suggèrent que Blimp-1 pourrait, en parallèle avec *pri*, participer à la mise en place des trichomes, ceci via une co-régulation des effecteurs de Svb.

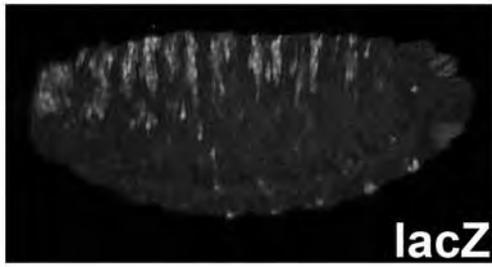
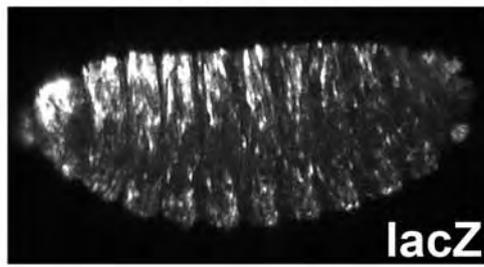
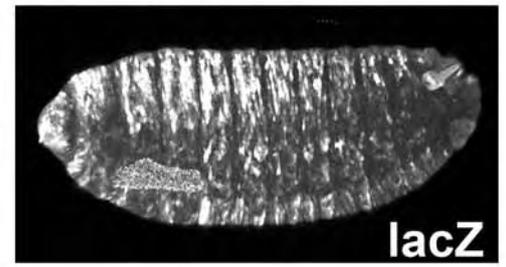
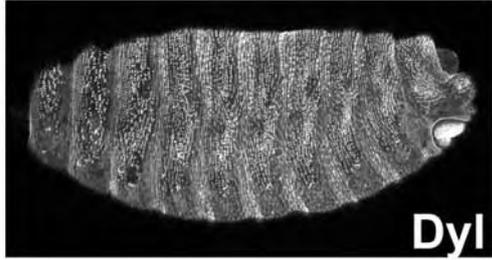
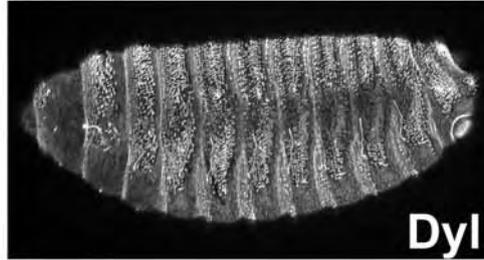
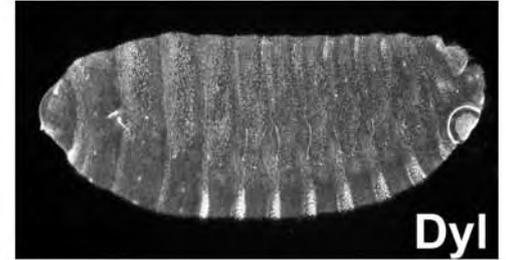
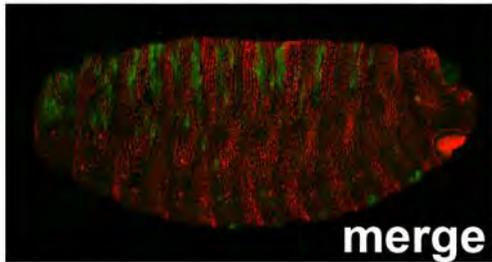
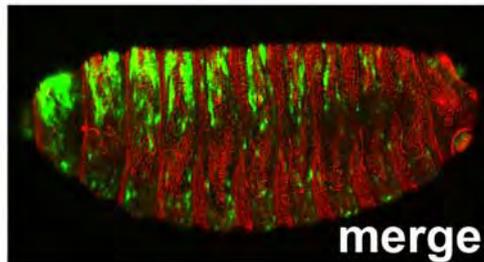
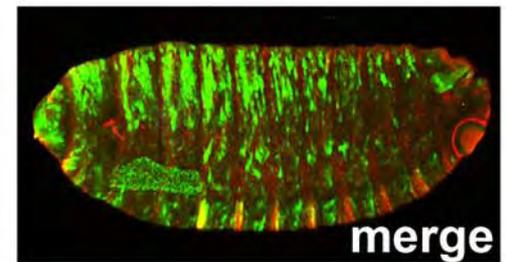
#### *A la recherche d'un facteur se fixant sur le Yellow Motif*

Un second motif identifié est le Yellow Motif, retrouvé dans un sous-ensemble des CRMs dépendants de Svb. Ce motif TTATGGAA a été initialement découvert par analyse systématique des profils de conservation évolutive de motifs de séquence (Elemento and Tavazoie 2005). Ce motif est donc significativement sur-conservé dans les génomes des drosophiles. On ignore cependant son rôle ; de même la nature du(es) facteur(s) qui pourrai(en)t s'y fixer reste inconnue. L'analyse des CRMs contenant le Yellow Motif par l'algorithme MEME ne permet pas de dégager des FT candidats. Cependant, la mutation de ce site dans le CRM Emin entraîne une perte partielle d'activité. De même la mutation des Yellow Motifs des CRM *nyo1* et *17058* affecte leur expression, montrant une contribution au moins partielle de ce motif dans l'activité des CRMs dépendants de Svb. La séquence du Yellow Motif semble toutefois proche de motifs reconnus par les facteurs OCT, mais nous n'avons pas identifié de candidats OCT épidermiques. L'identification du facteur qui reconnaît et lie le Yellow Motif reste donc un challenge, mais permettrait de mieux définir le réseau génique responsable de la mise en place des trichomes.

### **Organisation fonctionnelle des CRMs dépendants de Svb**

#### *Les sites SvbF7 et Blue Motif ne sont pas suffisants pour diriger une expression épidermique*

Afin de déterminer si l'association de sites Svb et Blue Motif est suffisante pour définir un CRM fonctionnel, nous avons construit des éléments synthétiques contenant différentes combinaisons de sites. La stratégie générale a consisté à partir d'une région génomique d'un gène cible de Svb, *shavenoid*, bien caractérisée pour son absence d'activité

**sha2****sha2-2R****sha2-2R3B****lacZ****lacZ****lacZ****Dyl****Dyl****Dyl****merge****merge****merge**

**Figure 20. La combinaison SvbF7 + Blue Motif ne suffit pas à induire une expression épidermique.** Immunocoloration des CRMs synthétiques sha2, sha2-2R et sha2-2R3B (lacZ en vert) et de la protéine Dyl (rouge). L'ajout de sites Svb et Blue Motif au CRM négatif sha2 entraîne une activité épidermique qui n'est pas restreinte aux cellules à denticules.

dans les cellules à trichomes, et chercher à « réveiller » son activité par l'introduction de différentes combinaisons des motifs que nous avons identifiés.

L'insertion de 2 sites Svb ne suffit pas à conférer une activité à la région négative *sha2* (Fig 20). Par contre le fragment *sha2-2R3B* qui contient 2 sites Svb et 3 sites Blue Motif dirige une expression épidermique dorsale et ventrale. Cependant l'expression dans les cellules ventrales n'est pas restreinte aux cellules à trichomes (Fig 20), ce qui suggère la présence d'autres régulateurs qui restent à identifier pour restreindre le patron d'expression de certains CRMs dans les cellules à denticules.

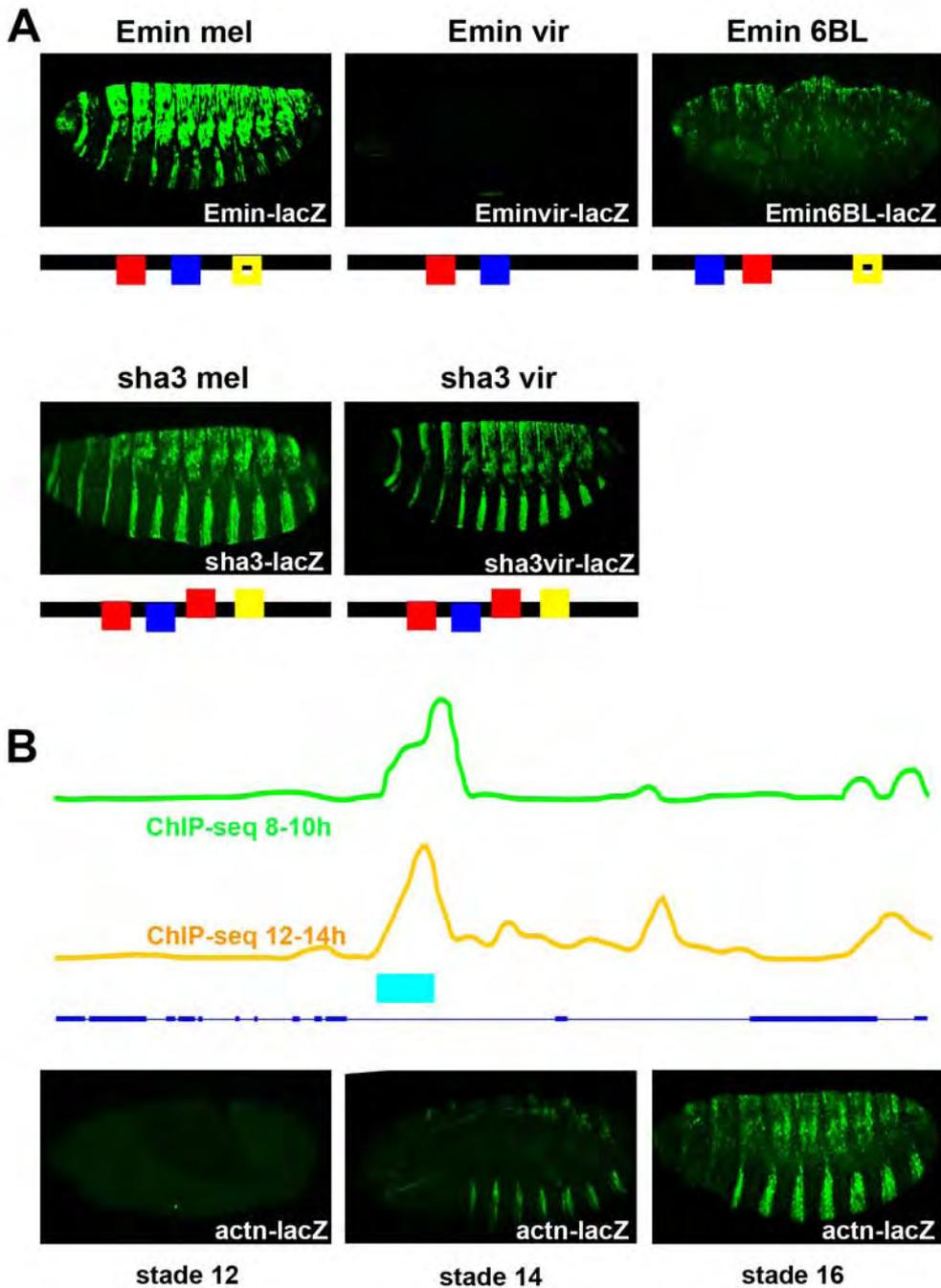
Le décryptage de la grammaire utilisée par Svb (motif Svb +/- Blue Motif +/- Yellow Motif) au sein de différents CRMs connus a permis ensuite de prédire l'activité de nouveaux CRMs répondant à ces critères. Pour affiner cette recherche, nous avons croisé les prédictions bio-informatiques avec des données de ChIP-seq pour Svb, ainsi qu'avec nos données de transcriptome d'embryons mutants pour *svb* ou *pri*. Nous avons ainsi défini des régions ChIPées associées à des gènes cibles de Svb et présentant une combinaison de sites. Parmi ces nouvelles régions testées, les régions présentant un pic de ChIP de faible intensité se sont avérées être négatives, alors que la majorité des autres régions dont le pic de ChIP est plus intense dirigent une activité dans l'épiderme qui dépend de Svb. Ces résultats soulignent le fait que des ChIP-seq ressortent un nombre important de régions avec une intensité moyenne ou faible qui ne sont pas fonctionnelles *in vivo* (Li, MacArthur et al. 2008). A l'inverse, nous avons observé des pics de ChIP très intenses non-associés à des sites reconnus par Svb, ou alors seulement à un site OvoQ6. Au moins certaines de ces régions pourraient correspondre aux sites HOT, récemment identifiés (Farley and Levine 2012). Ils semblent accumuler différents FTs, sans forcément dépendre d'une interaction spécifique ADN/protéine. Il est à noter que certaines cibles de Svb ne présentent pas de régions répondant à ces critères, ce qui suggère qu'une autre grammaire complètement différente



pourrait être utilisée pour activer ces effecteurs, ou bien qu'il ne s'agit simplement pas de cibles directes.

### *Shadow ou CRM classiques ?*

Nous avons établi que les gènes cibles de Svb possèdent souvent plusieurs régions régulatrices indépendantes, capables de reproduire tout ou partie du patron d'expression endogène. En effet, certains CRMs reproduisent parfaitement le patron d'expression du gène associé, alors que d'autres présentent une activité uniquement dans une sous-région du patron du gène. C'est le cas pour les trois CRMs associés à *miniature*. EminB et EminC dirigent respectivement des expressions dorsales et ventrales, alors qu'Emin reproduit l'ensemble du patron. De même, le nouveau CRM identifié associé au gène *CG9095* reproduit un patron uniquement dans les cellules ventrales de l'épiderme (Fig 7 publication) alors que le gène *CG9095* présente un patron complet également au niveau dorsal (Fig 23), ce qui suggère qu'une autre région est nécessaire pour diriger l'ensemble de l'expression du gène. On peut observer dans l'environnement du locus d'autres pics de CHIP associés à un ou plusieurs Yellow Motif. Ces régions pourraient diriger une expression qui « compléterait » celle du premier CRM identifié. Le nouveau CRM associé au gène *PH4alphaEFB*, qui contient un site Svb non conservé ainsi qu'un Yellow Motif, reproduit entièrement le patron du gène (Fig 7 publication). La conservation évolutive du/des sites(s) Svb n'est donc pas strictement indispensable pour définir les CRMs reconnus par Svb *in vivo*. Le nouveau CRM 11175 reproduit le patron d'expression du gène *CG11175* (Fig 7 publication). Le CRM 4914 présente un patron particulier uniquement dans le thorax, mais qui dépend quand même de Svb (Fig 7 publication). On observe dans l'environnement immédiat du gène un second pic de CHIP de forte intensité et qui contient deux sites SvbF7 non-conservés. Cette région pourrait correspondre à un second CRM, qui compléterait le patron d'expression dirigé par le premier CRM identifié. De façon générale, l'expression de nombreux effecteurs de Svb



**Figure 21. Vers un modèle contraint et une nouvelle fonction du FT Svb.**

**A. Une conservation chez *Drosophila virilis* pour certains CRMs**

Immunocolorations des CRMs Emin ou sha3 orthologues chez *virilis* ou ayant subi une modification de leur organisation (Emin6BL). Le CRM sha3 virilis dirige une activité épidermique similaire à sha3 *melanogaster*, alors que Emin *virilis* n'a pas d'activité. La modification de l'organisation de Emin *melanogaster* entraîne une perte partielle d'activité de la région.

**B. Svb, un FT pionnier ?**

Profils de ChIP-seq et immunocolorations du CRM actn à différents stades. Le CRM n'a pas d'activité au stade 12 alors que la protéine Svb est déjà fixée sur le site. Une activité épidermique est décelée au stade 14 puis est accentuée au stade 16.

dépend de plusieurs CRMs. Bien que leur localisation soit relativement proche du démarrage de transcription, ils apparaissent dotés de propriétés similaires aux shadow enhancers. On peut donc imaginer que les shadow enhancers contribuent à l'expression correcte de différents effecteurs cellulaires de la différenciation terminale, et pas seulement à des "master genes" qui régulent des étapes précoces du développement tel qu'initialement proposé.

### *Enhanceosome ou billboard ?*

L'ensemble de nos données soutient l'idée selon laquelle il n'existe pas de grammaire stricte, ou en tout cas unique, pour les éléments régulant l'expression des cibles de Svb. Certains ont besoin de plusieurs CRMs, alors que d'autres voient leur expression dirigée par une seule région. Au sein même des CRMs validés *in vivo*, on observe une variété de combinaisons possibles, certains ayant uniquement des sites Svb (12017, actn), d'autres présentent une combinaison Svb + Blue Motif (mey2, EminC), d'autres sont Svb + Yellow Motif (31022 et 4914) et d'autres Svb + Yellow Motif + Blue Motif (9095 et 11175).

Même si l'architecture fonctionnelle, la grammaire, de CRMs dépendants de Svb présente une flexibilité importante entre CRMs, la plupart de ces régions présentent cependant une forte conservation évolutive. On retrouve en général une conservation de l'organisation des différents motifs à travers les espèces de drosophile dont le génome est séquencé. On peut noter que les régions entourant les sites SvbF7, Blue & Yellow Motifs (25bp de part et d'autre) sont relativement plus fortement conservées, par rapport au reste de la séquence de chacun des éléments. Cette observation va dans le sens d'une conservation "verticale" de chacun des éléments entre différentes espèces. Afin de tester l'activité transcriptionnelle des régions potentiellement orthologues des CRMs, j'ai cloné des régions correspondantes de *Drosophila virilis* en amont d'un rapporteur lacZ. La région correspondant à sha3 de *virilis* dirige une activité épidermique similaire à celle de sha3 *melanogaster* (Fig 21A). Le CRM sha3 de *virilis* présente une organisation similaire au CRM



de *melanogaster*. Pour la région correspondant à Emin chez *virilis*, qui a perdu le Yellow Motif présent dans Emin *melanogaster*, nous n'observons cependant pas d'expression notable du rapporteur (Fig 21A). Ainsi la région génomique correspondante à un CRM validé chez *melanogaster* ne présente pas forcément la même activité chez *virilis*, bien que les sites Svbf7 et bleus soient conservés. La taille inférieure du fragment testé chez *virilis* par rapport au fragment Emin *melanogaster*, ainsi que la perte du Yellow motif peuvent cependant suggérer que le fragment testé est trop court, et qu'une insertion a éloigné le Yellow Motif des autres sites. L'analyse des régions adjacentes serait intéressante pour vérifier si Yellow Motif est présent et, si oui, tester *in vivo* l'activité de cette région plus grande.

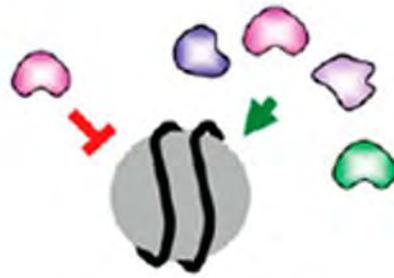
Au vu des résultats des CRM synthétiques nous avons voulu tester si l'organisation des sites d'un CRM est importante pour son activité. J'ai donc construit un nouveau CRM Emin, dans lequel j'ai changé de place le Blue Motif, déplacé de l'autre côté du site Svbf7 par rapport au CRM sauvage. Ce CRM appelé Emin6BL présente une forte altération de son activité, réminiscente de l'effet des mutations des bases flanquantes du site Svbf7 (Fig 21A). Il semblerait donc que l'organisation des sites les uns par rapport aux autres soit importante pour conférer une activité complète, ce qui va dans le sens, malgré la flexibilité horizontale d'un CRM à l'autre, de l'existence d'une contrainte verticale forte pour le même CRM à travers l'évolution des insectes.

### **Svb, facteur pionnier de la transcription**

La plupart des cibles de Svb présentent un patron d'expression qui commence aux stades 13-14. Cependant des ChIP-seq effectués aux stades 11-12 (8-10h de développement), montrent que dans la majorité des cas la protéine Svb est déjà fixée sur les CRMs (Fig 21B). Le test de l'activité de ces CRMs au stade 12 montre qu'ils ne dirigent pas encore d'activité, mais que celle-ci débute uniquement à partir du stade 14 (Fig 21B). Il semblerait donc qu'au

**A**

FTs classiques



fixation coopérative  
pour ouvrir la chromatine  
fixation simultanée  
d'autres FTs

FTs pionniers



fixation indépendante  
de la chromatine  
fixation avant  
les FTs classiques

**B**

Rôle passif

augmentation de la rapidité  
de fixation des autres FTs



fixation du FT pionnier : enhancer inactif

↓ Induction



fixation des autres FTs : enhancer actif

Rôles actifs

ouverture de la chromatine



↓ fixation d'autres FTs



### Figure 22. Caractéristiques des FTs pionniers.

Les FTs pionniers sont représentés par un triangle vert, les FTs classiques par les formes roses, bleues, violettes et vertes. L'ADN est représenté linéairement (panel B gauche), ou à l'état de nucléosomes (rond gris).

#### A. Différences entre FTs pionniers et FTs classiques

Alors que la fixation des FTs classiques nécessite un état chromatinien décondensé et se fait de façon simultanée, la fixation d'un FT pionnier ne dépend pas de l'état chromatinien et se fait avant la fixation d'autres FTs.

#### B. Les rôles des FTs pionniers.

La fixation d'un FT pionnier favorise un recrutement rapide d'autres FTs (passif) et entraîne une ouverture de la chromatine permettant la fixation des FTs classiques (actif).

D'après Zaret Carroll 2011

stade 12 Svb soit déjà fixé sur ses sites, mais à l'état répresseur. L'action de Pri aux stades 13-14 induisant la maturation activatrice de la protéine Svb (Kondo, Plaza et al. 2010) pourrait se faire directement au niveau du CRM, ce qui faciliterait une activation rapide/efficace et synchrone des effecteurs. Ces résultats suggèrent que Svb fasse partie des FTs dits pionniers, dont le rôle est de faciliter l'accès à la chromatine à d'autres FTs (Zaret and Carroll 2011) (Fig 22). Une approche possible pour tester cette hypothèse serait de comparer par ChIP-seq, dans les cellules S2, les marques chromatiniennes associées aux CRMs en présence de Svb sous sa forme répresseur ou sous sa forme d'activateur. On pourrait ainsi tester si la modification de Svb par les peptides Pri induit une modification des marques sur le CRM, participant ainsi à son activation.



## CONCLUSION

Cette étude nous a donc permis de découvrir le code utilisé par Svb pour reconnaître ses cibles transcriptionnelles et induire leur expression dans les cellules à trichomes. Nous avons établi qu'il existe des signatures de motifs permettant de prédire les CRMs associés aux effecteurs de Svb, mais qu'ils ne présentent pas d'organisation stéréotypée en terme de nombre, orientation et combinaison, une notion qui commence à émerger aussi suite à l'analyse d'autres processus de différenciation (Jin, Stojnic et al. 2013). Le modèle que nous avons établi, associé à d'autres approches globales comme CHIP-seq et transcriptome, permet de déceler à l'échelle du génome entier de nouveaux CRMs associés aux gènes cibles de Svb. Bien que les différents CRMs identifiés présentent chacun une organisation différente, une forte conservation de cette organisation se retrouve pour chacun d'entre eux pour les différentes espèces de drosophile. Il s'agit donc d'un des premiers déchiffrages du code de régulation utilisé par un FT de différenciation terminale pour reconnaître ses gènes cibles. De manière intéressante, il apparaît intermédiaire entre les modèles soit de l'enhanceosome (grammaire figée) soit billboard (grammaire flexible), généralement décrits dans la littérature.



## Partie II :

**Identification de nouvelles fonctions cellulaires  
en aval de Shavenbaby  
pour la morphogénèse des cellules à trichomes**



## INTRODUCTION

L'approche de microarrays m'a permis d'identifier une large liste de nouveaux gènes potentiellement régulés par Svb. J'ai poursuivi l'analyse de leur régulation par des expériences fonctionnelles, comparant de manière systématique leur expression épidermique en fonction du statut d'activité de Svb, dans des embryons de différents génotypes. En parallèle, un crible génétique réalisé dans l'équipe par Hélène Chanut vise à identifier de nouveaux acteurs de la morphogénèse. Collectivement, nos approches révèlent l'importance de la voie de polarité planaire (PCP) pour la morphogénèse des trichomes.

En combinant des analyses phénotypiques des trichomes différenciés et des cellules épidermiques en cours de différenciation, nous avons caractérisé le rôle des protéines de PCP appartenant aux modules GLOBAL et CORE dans la mise en place des extensions. Ces résultats mettent en lumière la participation de nouveaux acteurs insoupçonnés à la PCP dans l'épiderme embryonnaire. De plus ils montrent que Svb contribue au contrôle de la PCP via la régulation de certains membres de cette voie, établissant pour la première fois un rôle de Svb dans la PCP et élargissant le répertoire de ses cibles.



## RESULTATS

### Validation par hybridation in situ de 30 nouveaux effecteurs de Svb.

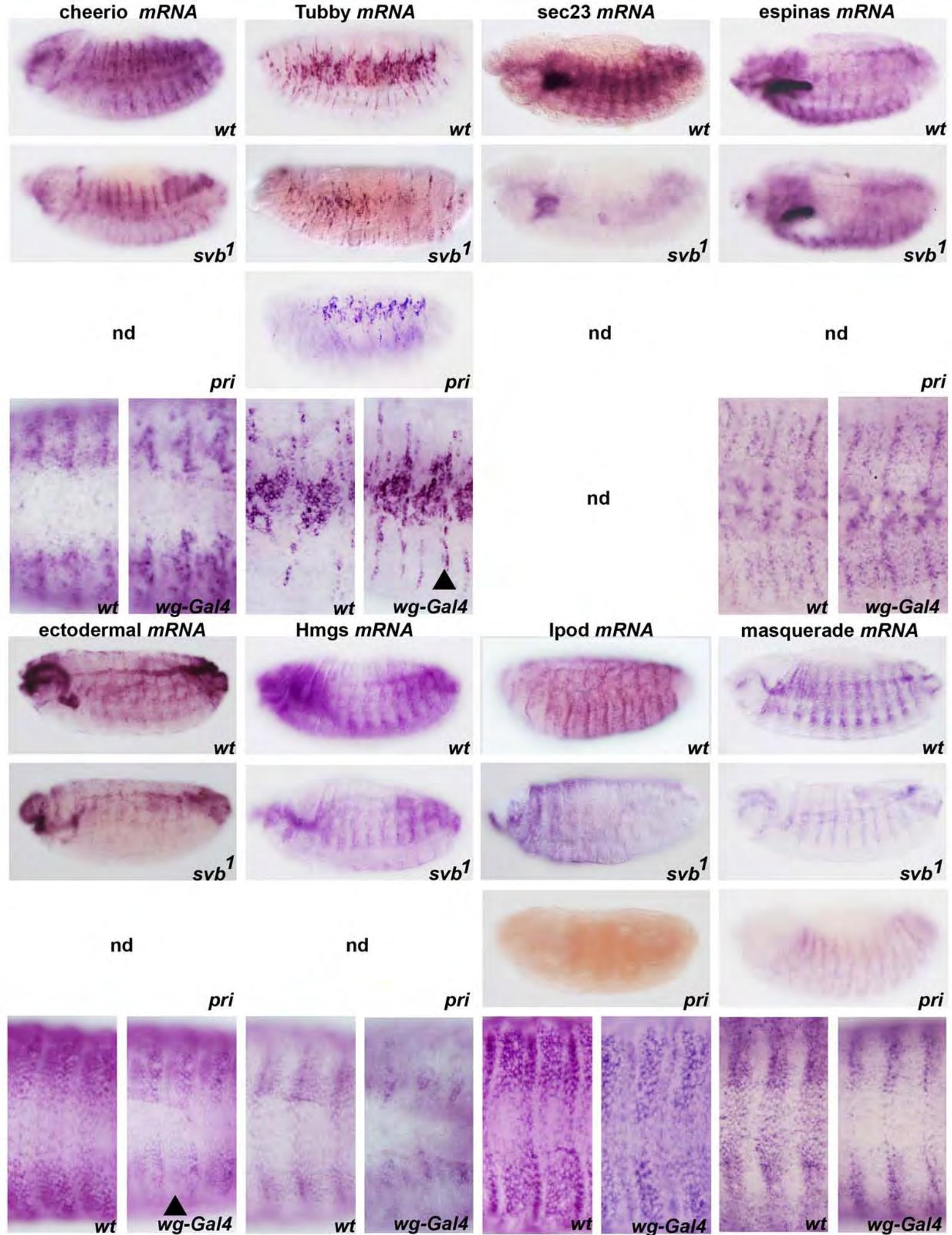
*Listes des gènes dont l'expression est diminuée en absence de svb ou pri.*

Afin d'identifier de nouveaux effecteurs de Svb, le transcriptome *svb* a été analysé. Des critères « relâchés » ( $pvalue < 0.05$ ,  $\logFC < 0.5$ ) établissent une liste de 149 candidats dont 9 cibles connues. Cette liste ne présente pas de FTs, soutenant que le FT Svb est probablement le régulateur le plus en aval de l'ensemble du réseau génique dirigeant la morphogénèse des cellules à trichomes. Plusieurs nouveaux candidats présentent toutefois des patrons d'expression non-épidermique, ajoutant un niveau de complexité dans l'analyse de cette liste de candidats. La mutation *svb*<sup>R9</sup> utilisée pour ce transcriptome correspond à une grande délétion du locus *ovo/svb* et délète 4 autres gènes voisins dont un facteur de transcription (CG6789, *rugose*, CG15465, CG32767). Ces candidats non-épidermiques peuvent donc être des cibles de l'un des autres gènes affectés par la délétion.

Pour identifier les gènes dont l'expression dépend de Pri, dans les cellules à trichomes mais aussi d'autres tissus, j'ai choisi des critères relâchés pour le transcriptome *pri* ( $\logFC < -1$  et  $pvalue < 0,01$ ) grâce auxquels j'obtiens une liste de plus de 350 candidats dont 19 cibles connues de Svb. De façon générale, les résultats du transcriptome *pri* pourront également donner accès à de nouvelles fonctions de Pri dans le développement, fonctions que nous identifierons par un enrichissement en gènes participant à ce processus et présentant une dérégulation.

*Validation par une approche qualitative : hybridation in situ*

3 gènes (*quail*, CG9196, CG18507) ne présentent pas de diminution d'expression détectable par hybridation *in situ* en absence de *svb*. Ils montrent cependant une expression significativement diminuée en absence de *pri*. Il peut donc s'agir d'effecteurs dont les



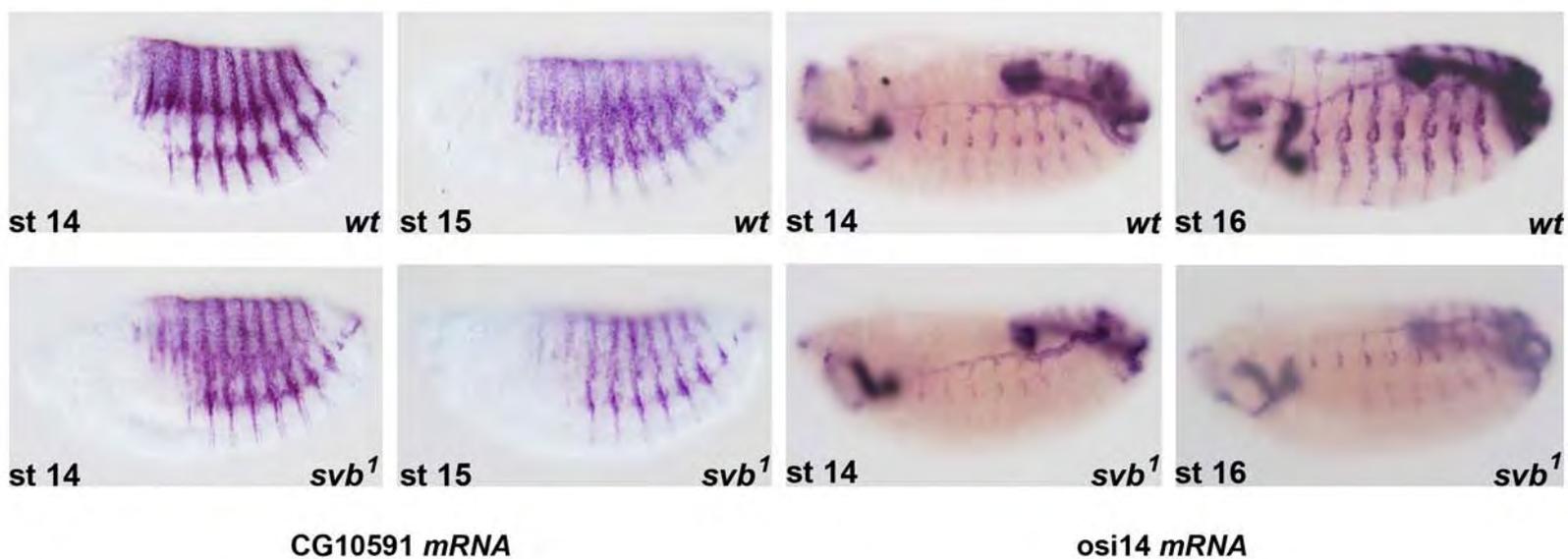
**Figure 24. De nouveaux effecteurs aux fonctions connues.**

Hybridation *in situ* de 8 nouveaux effecteurs dans des embryons *wt*, *svb* et *pri*. Le driver *wg>Gal4* est utilisé pour tester l'effet d'une expression ectopique de *svb* sur leur expression (triangle noir).

variations ne sont pas visibles par hybridation *in situ* bien qu'ils soient régulés par Svb, ou bien qu'il s'agit de gènes dont l'expression dépend de Pri mais pas de Svb.

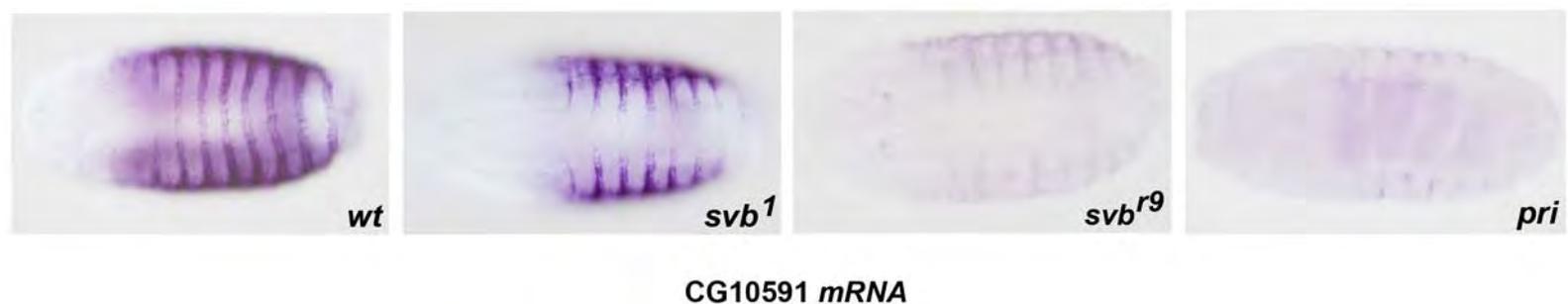
Afin de cibler les candidats les plus prometteurs, j'ai recherché pour chaque liste de candidats (*svb-* et *pri-*) les gènes : i) dont l'expression dans l'épiderme est déjà répertoriée, ii) présentant les plus fortes diminutions d'expression. J'ai ainsi analysé 75 gènes candidats par hybridation *in situ* systématique en embryons mutants *svb*. Pour m'assurer du rôle de Svb pour l'expression de ces gènes, j'ai choisi de changer de fond génétique en utilisant la mutation *svb1*. *svb1* est une mutation ponctuelle qui affecte la fonction de Svb spécifiquement dans l'épiderme. Les tests d'expression des nouveaux effecteurs dans les mutants *svb1* permettent de valider les modulations d'expression quantifiées dans un autre fond génétique plus complexe. Les candidats présentant une modification d'expression ont ensuite été testés en condition mutante *pri* et/ou sur-expression de Svb. J'ai ainsi identifié 30 nouveaux effecteurs de Svb, la plupart ayant une fonction encore inconnue (Fig 23), et d'autres pour lesquels nous disposons de quelques informations bibliographiques (Fig 24). Pour 4 de ces nouveaux effecteurs (CG9095, CG13822, CG4914, CG12017), un CRM associé et dépendant de Svb a pu être identifié (Fig 7 publication). CG1273, CG4666, CG7173, CG9514, CG12017 et CG13822 présentent une expression diminuée en absence de Svb, et une expression ectopique forte quand on exprime Svb dans 2 rangées de cellules lisses (driver *wg-Gal4*) (Fig 23), montrant que Svb est nécessaire et suffisant pour induire leur expression dans l'épiderme. L'expression du gène CG33056 n'est pas induite par expression ectopique de Svb (Fig 23). Il peut s'agir d'un effecteur pour lequel Svb est nécessaire mais non-suffisant pour induire son expression, montrant la diversité de régulation des cibles de Svb.

L'enrichissement en gènes dont la fonction reste inconnue souligne la diversité des mécanismes impliqués dans la morphogénèse épidermique. De façon générale, les nouveaux effecteurs ne présentent pas de patron épidermique aussi stéréotypé que les effecteurs déjà



**Figure 25. Des variations différentes selon le contexte génétique.**

Hybridation *in situ* pour les nouveaux effecteurs de Svb : *CG10591* et *Osi14*. Alors que la mutation *svb* a un effet partiel sur l'expression de *CG10591*, l'expression de *Osi14* est totalement abolie dans l'épiderme dans le même contexte.



**Figure 26. Un exemple typique de modulation d'expression : *CG10591***

Hybridation *in situ* pour le nouvel effecteur *CG10591*. La mutation nulle *svbR9* a plus d'effet sur l'expression du gène que la mutation *svb1*. L'expression du gène est totalement perdue en absence de Pri.

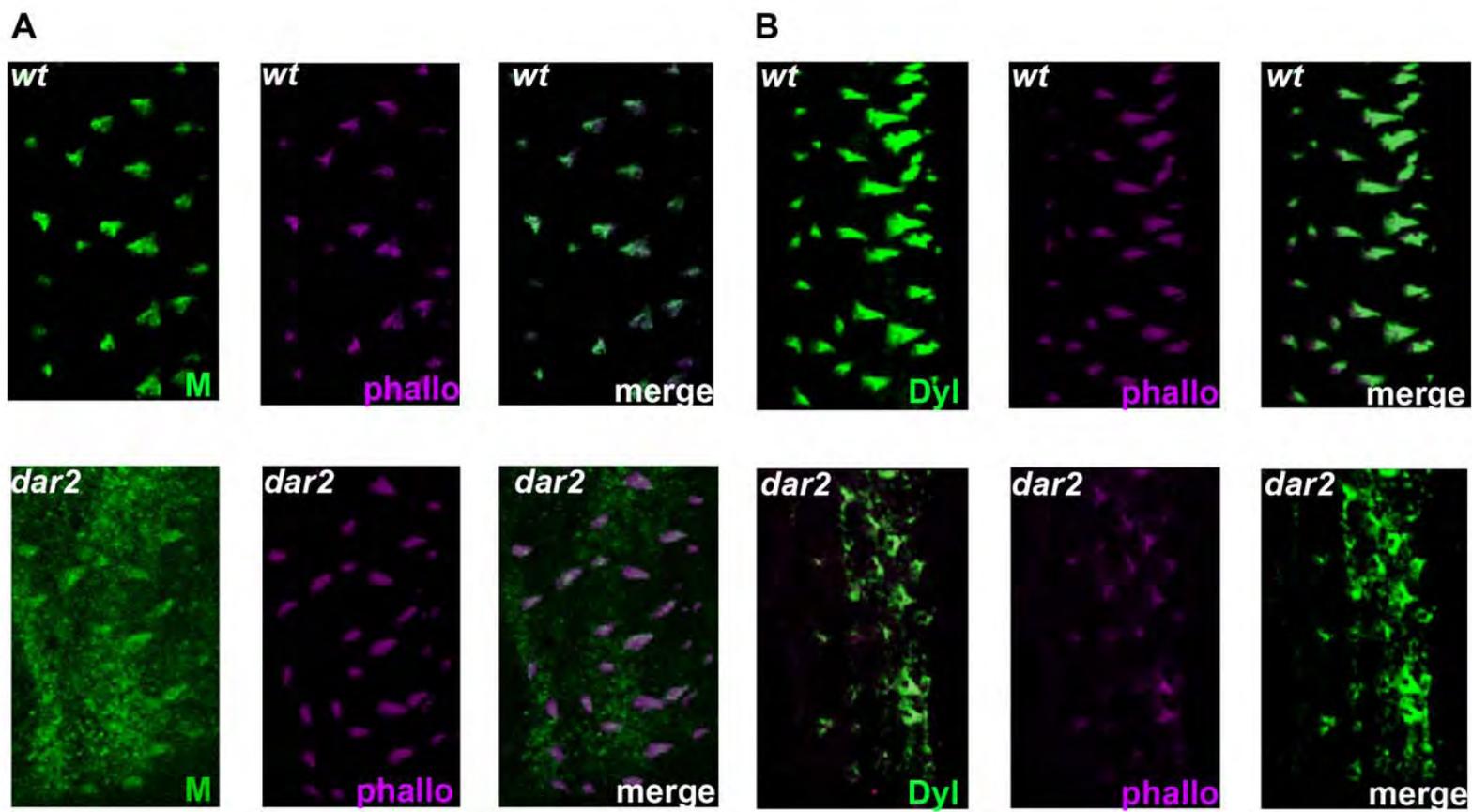
connus. Par exemple, *Espinas* présente une forte expression dans le système nerveux et les glandes salivaires. *ectodermal*, *Hmgs* et *masquerade* sont aussi fortement exprimés dans l'intestin antérieur (Fig 24). Ces patrons secondaires complexifient les analyses quand les modulations d'expression sont partielles.

Les variations d'expression observées en hybridation *in situ* peuvent également être classées selon leur amplitude. Il existe des gènes comme *Osi14* pour lesquels la perte d'expression est totale en absence de *Svb*, et des gènes comme *CG10591* pour lesquels la diminution d'expression est partielle (Fig 25). La diminution d'expression du gène *CG10591* est partielle dans le ventre dans le mutant *svb1* et quasiment totale dans le mutant *svbR9* (Fig 26). Ces données montrent aussi l'intérêt, et sans doute la limite, d'utiliser différents allèles mutants d'un même gène pour mieux comprendre sa fonction. L'utilisation du fond génétique *svb1* peut ainsi également être source de faux négatifs.

Les 30 nouveaux gènes cibles de *Svb* peuvent donc présenter des niveaux de régulation différents. L'expression de *svb* peut être nécessaire et suffisante pour l'expression de certains effecteurs, et seulement nécessaire pour d'autres. L'effet de l'absence de *svb* sur leur expression peut également être partiel. Ces résultats soutiennent l'hypothèse d'une régulation directe et exclusive par *Svb* de certains effecteurs, et d'une régulation couplée avec d'autres facteurs (voire indirecte) pour les autres cibles.

### **De nouvelles fonctions dépendantes de *Svb* pour la morphogénèse.**

Parmi les nouveaux effecteurs on peut souligner la présence de *cheerio*, codant pour la protéine de liaison à l'actine appelée Filamin (Fig 24). Cette nouvelle cible s'ajoute donc aux acteurs faisant partie des protéines de liaison à l'actine, son rôle précis restant cependant à



**Figure 27. Sec23, un composant de la voie de sécrétion nécessaire pour la localisation des ZPs.**

Immunocoloration des protéines M et Dyl (vert) et des faisceaux d'actine (violet) dans des embryons mutants *sec23* (*dar2*).

**A. La localisation de M est affectée en absence de *sec23*.**

Alors que M est localisée dans la région médiane de la denticule dans des embryons *wt*, la mutation de *sec23* entraîne une délocalisation de M vers la base des denticules.

**B. La localisation de Dyl est affectée en absence de *sec23*.**

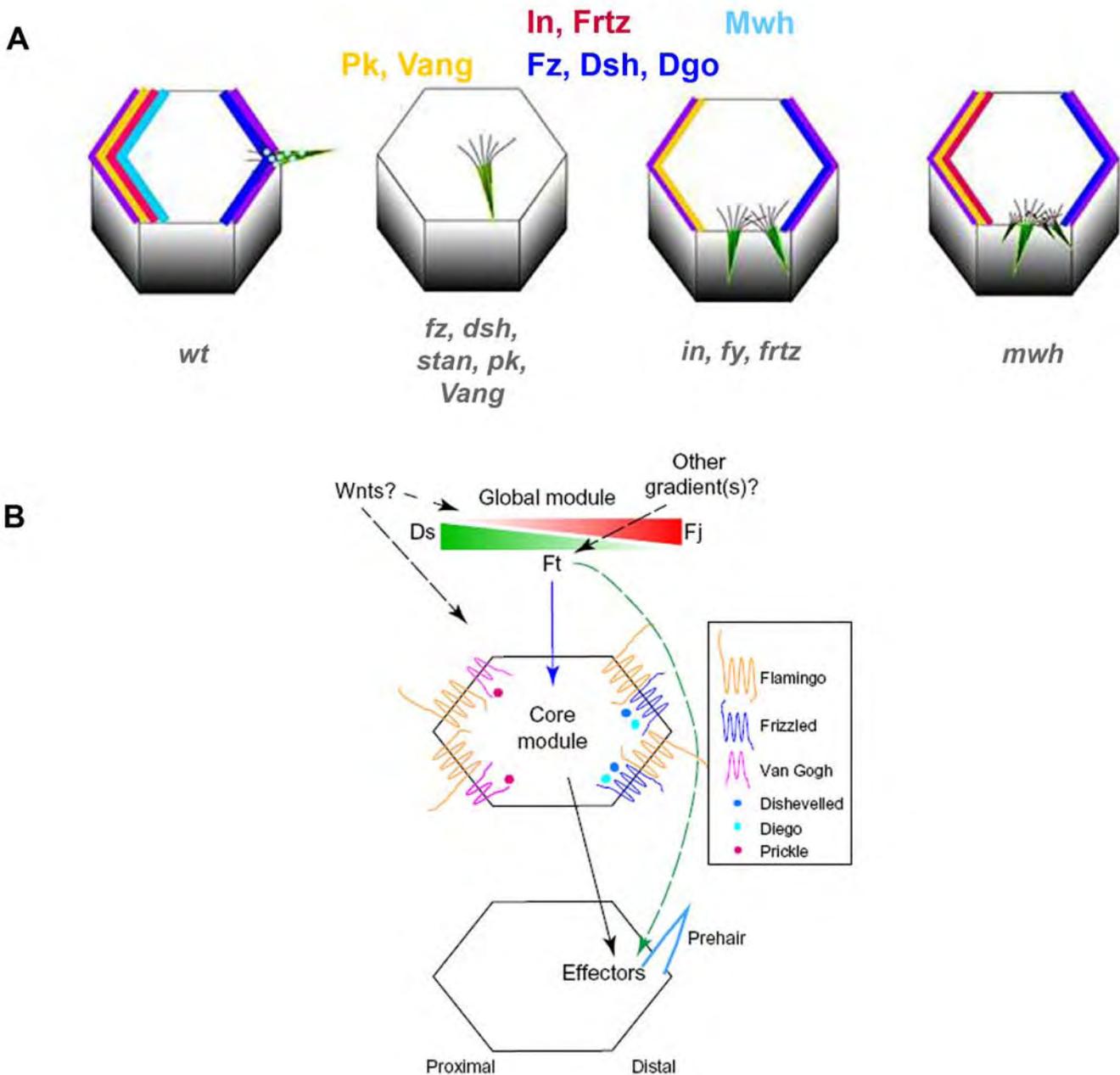
Alors que Dyl est localisée dans la région la plus apicale des denticules dans des embryons *wt*, l'absence de *sec23* entraîne une délocalisation de Dyl vers la base des denticules.

identifier. Un nouvel effecteur de Svb appartenant à une classe connue est *Tubby* (*Tb*) appelé aussi *TweedleA*, codant un composant structural de la cuticule (Fig 24).

Des acteurs de la voie de sécrétion sont également surreprésentés dans le transcriptome *pri*. *sec23* et *sec13* sont des composants des vésicules COPII, *sec64* est associé au réticulum endoplasmique et *epsilon-COP* associé aux vésicules COPI. La diminution d'expression de *sec23* en absence de Svb a été validée par hybridation *in situ* (Fig 24). Dans des embryons mutants *sec23*, les protéines Dyl et M sont délocalisées de leur sous-région spécifique dans les denticules (Fig 27A, B). Des analyses génétiques réalisées par Hélène montrent l'implication d'autres membres de la voie de sécrétion dans la morphogénèse épidermique.

Un nouvel effecteur de Svb qui m'a particulièrement intéressé est le gène *masquerade* (*mas*) (Fig 24), codant une peptidase atypique. En effet, cette protéine est clivée en 2 régions N-terminale et C-terminale, qui sont alors localisées dans différents tissus de l'embryon (Murugasu-Oei, Balakrishnan et al. 1996). La région C-terminale se localise dans le tissu musculaire, au niveau des sites d'attachement des muscles, alors que la région N-terminale se localise spécifiquement dans l'épiderme et les trachées, où elle participerait à la stabilisation des interactions cellule/matrice. De façon intéressante, on observe en immunofluorescence un enrichissement de la protéine Mas au niveau des denticules (Hélène Chanut, données non publiées), ce qui suggère effectivement un rôle de la protéine dans les interactions entre matrice extracellulaire et cuticule. Des analyses sont en cours au laboratoire afin de tester l'implication de *mas* dans la morphogénèse des trichomes.

Parmi les nouveaux processus impliqués dans la morphogénèse que j'identifie grâce au transcriptome, se trouve le mécanisme de polarité planaire (PCP). La polarité planaire



**Figure 28. La PCP dans l'aile chez *Drosophila melanogaster*.**

**A. La PCP dépend de la localisation asymétrique de ses acteurs.**

La localisation asymétrique des acteurs de la PCP conduit à la formation d'une extension dans la région distale de la cellule (wt). La mutation d'un ou plusieurs acteurs entraîne des défauts de localisation de l'extension.

D'après Adler 2012.

**B. Deux modules participent à la PCP.**

Le module Fat, Ds (rouge), Fz (vert) donne une polarité à l'ensemble du tissu. Les protéines trans-membranaires Fz (bleu foncé), Vang (rose) et Stan (Flamingo, orange) interagissent pour entrainer une localisation asymétrique des composants cytoplasmiques : Dsh (bleu foncé), Dgo (bleu clair) et Pk (rose), entrainant la localisation asymétrique d'effecteurs tissu-spécifiques qui restreignent la formation de l'extension dans la partie distale de la cellule.

D'après Peng, Axelrod 2012.

définit l'implantation d'une extension au sein de la surface apicale des cellules à trichome (vertex postérieur), ainsi que l'orientation antéro-postérieure de l'extension. J'identifie ainsi plusieurs acteurs de la PCP que je vais détailler dans la suite.

## Un nouveau rôle de *Svb* dans la définition du site d'implantation des trichomes

### *Etat de l'art sur la polarité planaire (PCP)*

L'analyse du transcriptome *pri* m'a permis d'identifier de nouvelles cibles du FT *Svb*. Parmi ces gènes, on retrouve certains acteurs de la voie de signalisation de polarité planaire (PCP), qui détermine l'implantation et l'orientation d'une extension à la face apicale d'une cellule. Ce processus est principalement étudié dans deux modèles chez la drosophile : l'aile et l'œil. L'aile est composée d'une mono-couche de cellules portant toutes à leur face apicale une extension distale (Fig 28A). Ce tissu à organisation simple fut à l'origine de la découverte des acteurs de la PCP, souvent conservés dans d'autres tissus et organismes (Adler 2012).

La PCP est une voie de signalisation Wg non-canonique composée de 2 modules. Le premier module dit GLOBAL est composé de cadhérines atypiques appelées Dachsous (Ds) et Fat (ft) interagissant respectivement comme récepteur et ligand entre deux cellules voisines. Ce module définit globalement à travers le tissu un gradient permettant d'orienter l'activité du second module, dit CORE qui contient Frizzled (Fz)/Stan (Fig 15). Selon les tissus, les activités des modules Ft/Ds et Fz/Stan peuvent être parallèles ou interdépendantes. Des interactions hétérologues entre Fz et Vang, et homologues entre 2 molécules Stan, entre 2 cellules voisines permettent le recrutement asymétrique de composants cytoplasmiques : Dishevelled (Dsh), Diego (Dgo) et Prickle (Pk) (Fig 28B) (Adler 2012; Peng and Axelrod 2012). Ces composants cytoplasmiques régulent eux-mêmes la localisation asymétrique d'effecteurs spécifiques, comme des molécules de liaison à l'actine, ce qui permet de délimiter le site d'initiation de l'extension dans la partie distale de la cellule de l'aile.

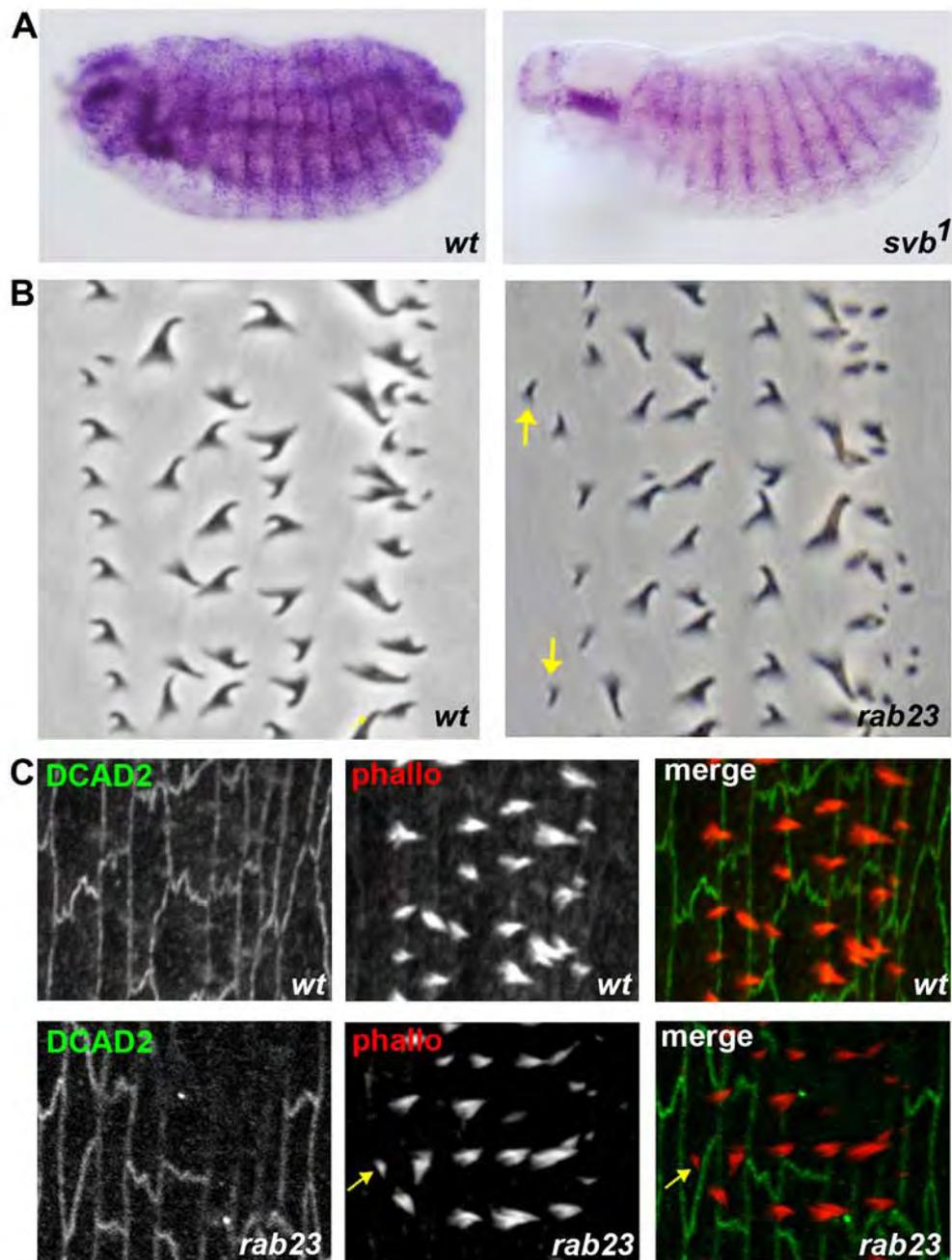


Figure 29. Le nouvel effecteur *Rab23* participe à la morphogénèse épidermique.

**A. *Rab23* est un nouvel effecteur de *Svb*.**

Hybridation *in situ* de l'ARNm *Rab23* sur embryons *wt* et *svb*. L'expression de *Rab23* est diminuée dans l'épiderme en absence de *svb*.

**B. Des larves *rab23* présentent des denticules ectopiques.**

Cuticule d'embryon *wt* ou mutant *rab23*. La mutation nulle de *rab23* entraîne la formation de denticules ectopiques (flèches jaunes).

**C. Les denticules ectopiques sont localisées dans la région antérieure des cellules de la rangée 1.**

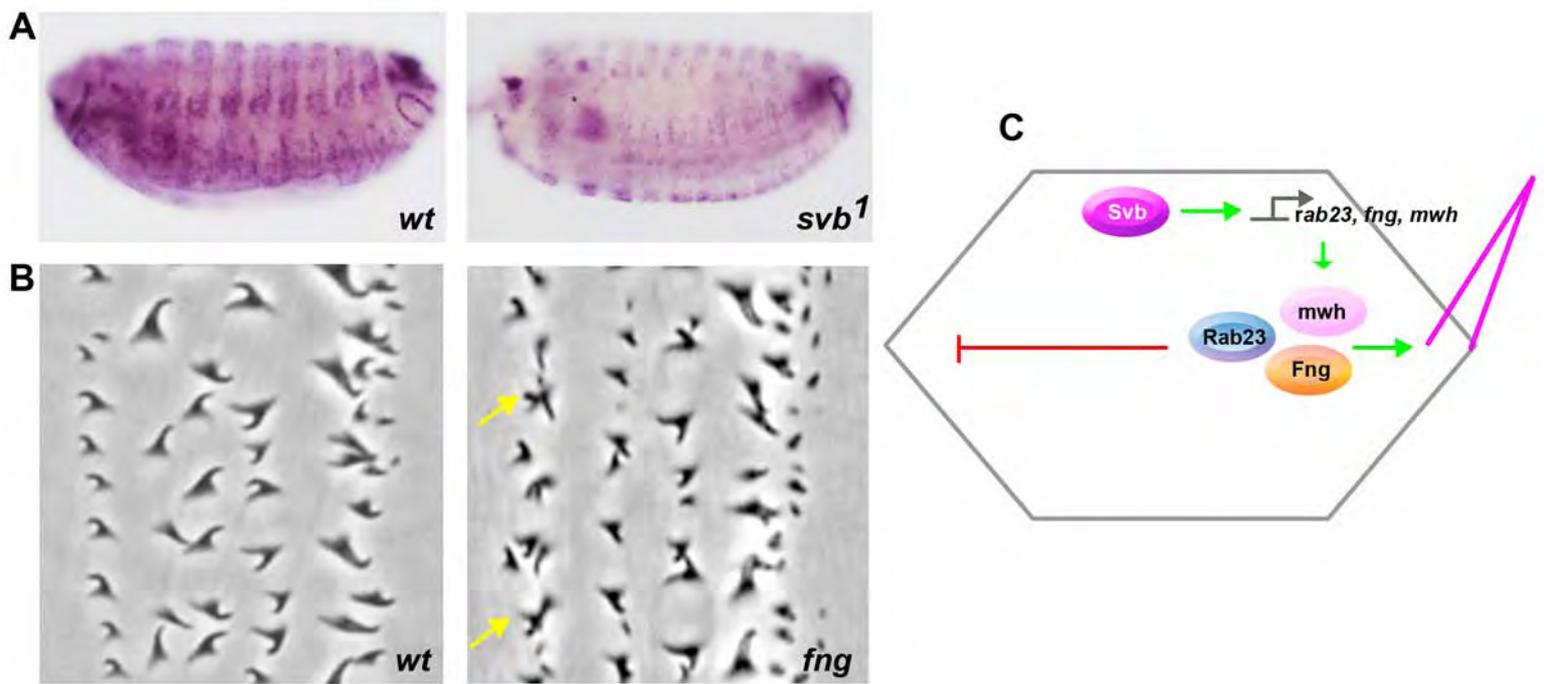
Immunocoloration de la DE-cadhérine (DCAD2, vert) et des faisceaux d'actine (phallo, rouge) en embryons *wt* ou *rab23*. L'extension ectopique est visualisée par une flèche jaune.

La mutation de certaines petites GTPases impliquées dans le trafic vésiculaire entraîne des défauts d'orientation des extensions des cellules de l'aile (Mottola, Classen et al. 2010; Pataki, Matusek et al. 2010; Purvanov, Koval et al. 2010). Des mutations affectant des protéines de liaison à l'actine entraînent des défauts de localisation asymétrique des protéines du module Fz/Stan dans l'aile (Blair, Tomlinson et al. 2006; Ren, Charlton et al. 2007). L'isolement de mutations conduisant à des défauts de PCP a permis l'identification d'acteurs contribuant de manière différentielle à la PCP. Même si de nombreux travaux ont disséqué la voie PCP dans l'aile, d'autres tissus requérant ce processus n'ont pas été étudiés aussi finement. Ainsi il est établi que la mutation d'acteurs de la PCP perturbe la formation des soies sensorielles de l'abdomen, du thorax et des appendices adultes (Adler 2012). De même l'organisation stéréotypée des trichomes suggère l'implication de la PCP dans la différenciation de l'épiderme embryonnaire et larvaire. En effet, plusieurs études montrent un rôle pour les modules Fz/Stan et ft/ds, qui semblent ainsi agir en parallèle pour les denticules du troisième stade larvaire (Repiso, Saavedra et al. 2010; Donoughe and DiNardo 2011). Il reste cependant à découvrir si les acteurs identifiés dans l'aile sont également impliqués dans l'épiderme, et ceci aux différents stades développementaux.

#### *Des acteurs de la PCP : nouveaux effecteurs de Svb*

Nous avons été intéressés par l'observation que l'expression des gènes *prickle* (*pk*), *Rab23* et *fringe* (*fng*) est fortement diminuée dans le transcriptome *pri*. L'hybridation *in situ* m'a permis de valider *in vivo* cette down-régulation pour *Rab23* (Fig 29A) et *fng* (Fig 30A), qui présentent une baisse significative d'expression dans l'épiderme d'embryons mutants *svb* ou *pri*. Concernant *pk*, j'observe par contre une expression dans les glandes salivaires qui n'est pas influencée par l'absence de Svb.

*Rab23* est une petite GTPase qui participe à la mise en place de la PCP dans différents tissus. En effet, en absence de *Rab23* on observe le même type de défauts dans



**Figure 30. Le rôle de Svb dans la restriction du site d'initiation des extensions.**

**A. Svb régule l'expression de *fng*.**

Hybridation *in situ* de l'ARNm *fng* sur embryons *wt* et *svb*. L'absence de *svb* entraîne une diminution de l'expression de *fng* dans l'épiderme.

**B. Des larves *fng* présentent des extensions ectopiques.**

Cuticules d'embryon *wt* ou *fng*. La mutation nulle de *fng* entraîne la formation de denticles ectopiques au niveau de la rangée 1 (flèches jaunes).

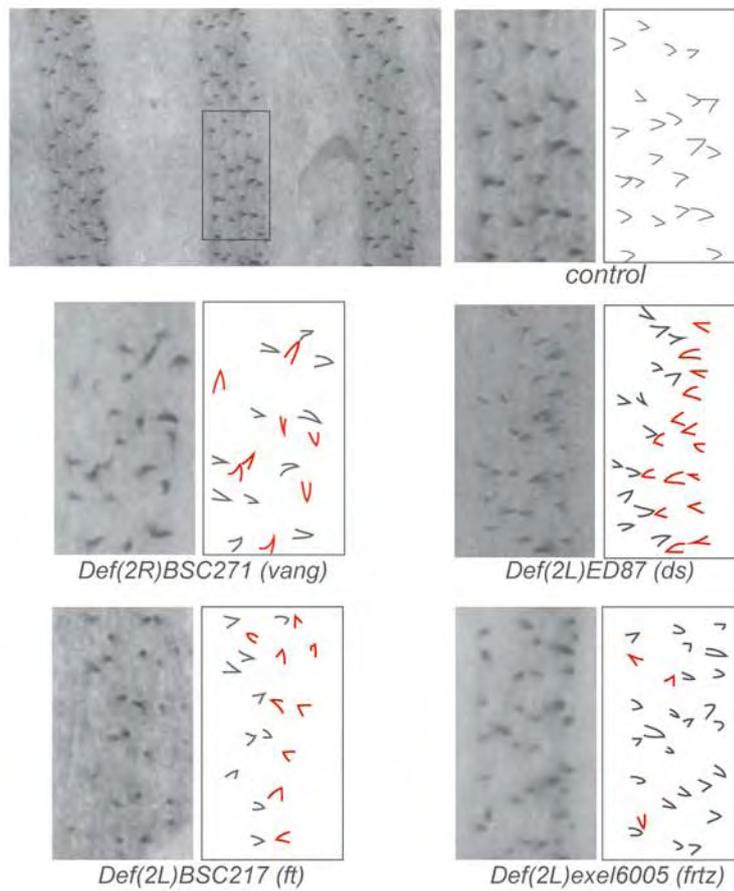
**C. Un nouveau rôle pour Svb.**

Modèle d'action de Svb sur les effecteurs Rab23, *fng* et Mwh. Svb induit l'expression de ces 3 gènes cibles dont le rôle est de restreindre la formation des extensions dans la partie postérieure de la cellule en réprimant la formation d'extensions dans les autres régions.

l'aile, la patte et la cuticule abdominale, à savoir une duplication de l'extension cytoplasmique à la face apicale des cellules. Des études *in vitro* ont montré une interaction spécifique entre Rab23 et Pk en cellules S2 (Pataki, Matusek et al. 2010). Comme le révèle l'analyse des préparations de cuticule, les embryons mutants pour *Rab23* présentent souvent des trichomes ectopiques (Fig 29B). Nous avons effectué des marquages fluorescents soulignant le contour cellulaire (DE-cadherine) et les faisceaux d'actine (phalloïdine). La résolution de ces observations menées en microscopie confocale permet ainsi de déterminer précisément le nombre d'extensions par cellule, leur implantation et leur orientation. Les extensions ectopiques observées en condition *rab23* *-/-* sont localisées dans les cellules de la rangée 1, et sont composées de faisceaux d'actine apparemment similaires aux extensions normales (Fig 29C).

*fng* est un modulateur de la voie Notch régulant l'interaction entre Notch et Serrate. En absence de *fng*, on observe des trichomes ectopiques de taille réduite, qui semblent correspondre au dédoublement des extensions (Fig 30B). Ces observations suggèrent un rôle de *fng* dans la PCP, pour définir ou/et limiter le site d'initiation de l'extension. Des analyses par immuno-marquage sont en cours pour analyser en détail ce phénotype et comprendre la fonction précise de Fng dans la PCP de l'épiderme embryonnaire.

Un autre acteur de la PCP avait été préalablement identifié comme cible de Svb : *multiple wing hair (mwh)*. *mwh* présente une forte diminution d'expression en absence de Svb, autant en microarrays qu'en hybridation *in situ*. La fonction de Mwh a surtout été étudiée dans l'aile adulte où Mwh est nécessaire pour restreindre la formation de faisceaux d'actine au site d'initiation (Yan, Huen et al. 2008). En son absence, de nombreux faisceaux se forment dans la cellule ce qui conduit à la multiplication des extensions (plus courtes) formées dans chaque cellule, un phénotype d'ailleurs à l'origine du nom *mwh*.



**Figure 31. De nouveaux acteurs de la morphogénèse : des membres de la polarité planaire (PCP).** Immunocoloration de la protéine Dyl sur des embryons déficients pour des gènes participant à la voie de signalisation PCP. Des défauts de polarité sont observés pour les déficiences qui recouvrent *Vang*, *ds* ou *fat*. Pour chaque déficience, un schéma représente les denticules présentant un défaut d'orientation (rouge).

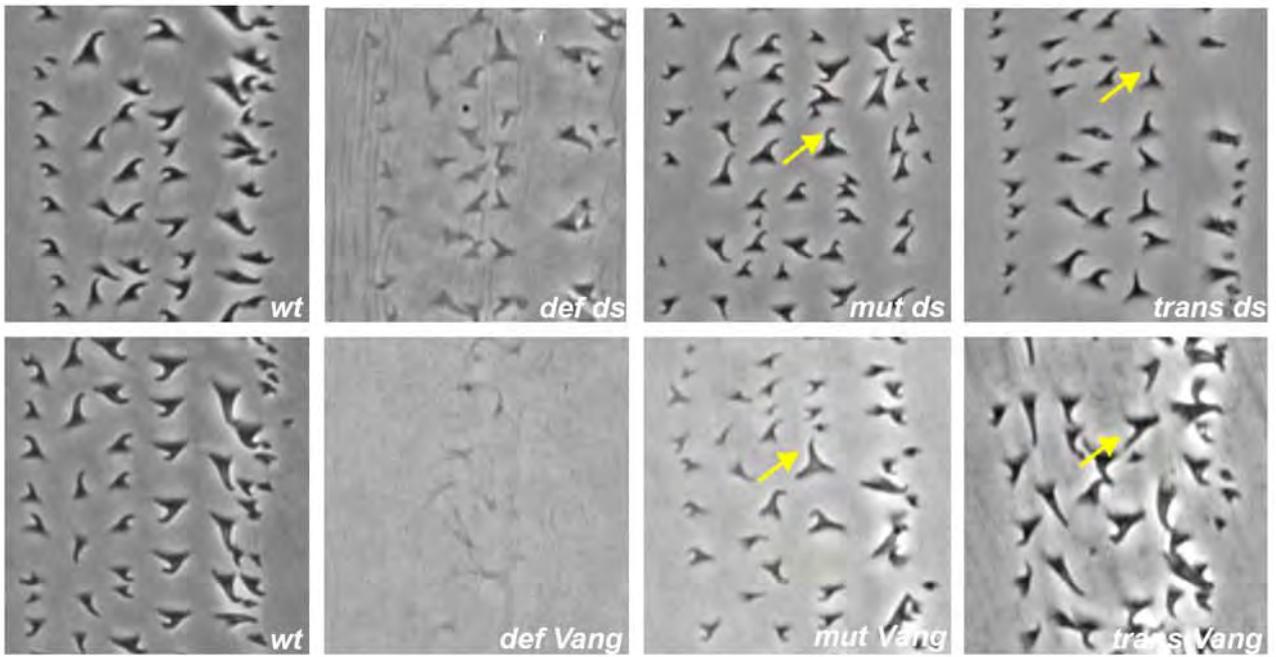
Le même type de phénotype est observé pour les mutants *Rab23* ou *fng* suggérant un rôle de ces protéines dans la restriction du site d'initiation de l'extension. Ainsi, nous identifions pour la première fois un rôle de *Svb* dans la restriction du nombre d'extensions par cellules, par l'activation des gènes *Rab23*, *fng* et *mwh* (Fig 30C).

**Collectivement, ces résultats montrent donc une régulation par *Svb* d'effecteurs tissu spécifique de la PCP, leur rôle étant de réguler l'initiation d'une extension dans la région postérieure de la cellule.**

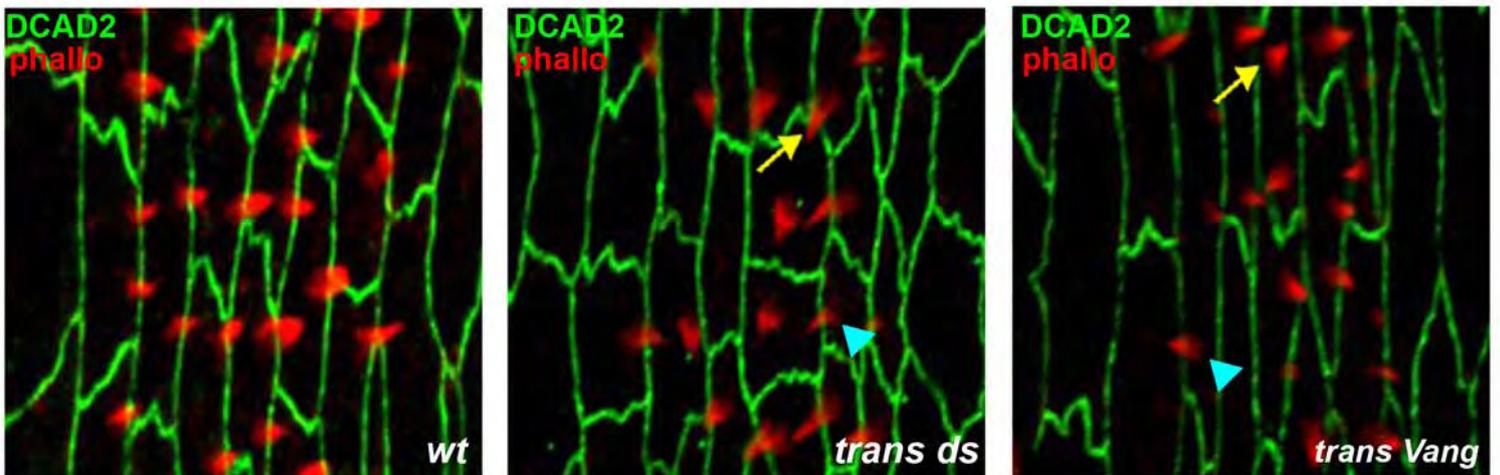
### Participation des modules GLOBAL (*fat/ds*) et CORE (*fz/stan*) dans la morphogénèse épidermique

En parallèle, Hélène a réalisé un crible génétique afin d'identifier de nouveaux régulateurs de la morphogénèse. Pour cela, elle a analysé la localisation de *Dyl* dans un ensemble de déficiences recouvrant collectivement l'ensemble du chromosome II. Parmi les déficiences candidates, certaines recouvrent des acteurs de la PCP, et présentent des défauts d'orientation des extensions. Ainsi l'absence de *Van Gogh* (*Vang*) conduit à des défauts qui affectent toutes les rangées de cellules à trichomes (Fig 31). Pour les embryons déficients pour *dachsous* (*ds*), on observe un défaut d'orientation progressif qui affecte essentiellement les rangées postérieures (Fig 31), au sein de chaque segment abdominal.

*ds* code une cadhérine atypique interagissant avec *Fat* dans le module GLOBAL, dont le rôle est de conférer une polarité globale à l'ensemble du tissu. *Vang* code une cadhérine du module CORE impliqué dans la définition précise de la polarité cellule par cellule. Alors que l'implication du module GLOBAL dans l'épiderme embryonnaire a déjà été décrit (Repiso, Saavedra et al. 2010; Brittle, Thomas et al. 2012), le rôle du module CORE dans l'épiderme restait controversé (Price, Roberts et al. 2006; Repiso, Saavedra et al. 2010). L'identification du rôle de *Vang* dans l'épiderme embryonnaire constitue donc une avancée importante. Afin



**Figure 32. Des défaut de polarité des denticules associés aux acteurs de la PCP *ds* et *Vang*.** Cuticules d'embryons déficients, mutants ou trans-hétérozygotes pour *ds* ou *Vang*. On observe des défauts importants de morphogénèse dans les déficiences et des défauts de polarité des denticules (flèches jaunes) dans les embryons mutants et trans-hétérozygotes.



**Figure 33. *Vang* et *ds* participent à l'orientation et l'implantation des denticules.** Immunocoloration de la DE-cadhérine (DCAD2 en vert) et des faisceaux d'actine (phallo en rouge) sur des embryons trans-hétérozygotes *ds* ou *Vang*. Les extensions présentent des défauts de polarité (flèches jaunes) et d'implantation (triangles bleus). Les défauts touchent toutes les rangées aléatoirement pour *Vang* alors qu'ils sont spécifiques des rangées postérieures pour *ds*.

de définir le rôle de ces nouveaux acteurs dans la morphogénèse des trichomes, j'ai réalisé des analyses pour les deux gènes présentant les défauts les plus forts : *ds* et *Vang*.

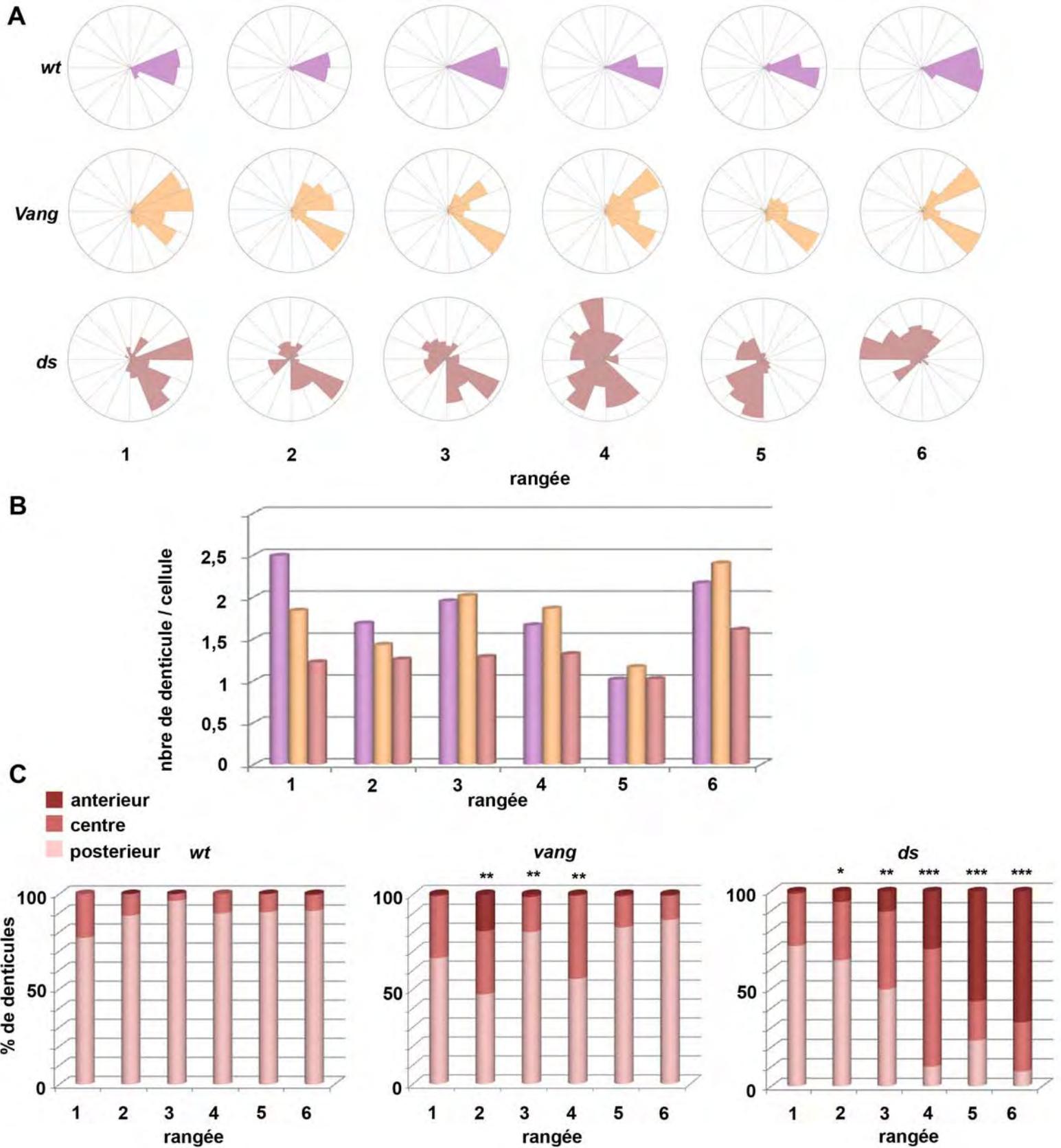
Des larves déficientes, mutantes, ou trans-hétérozygotes, pour *ds* présentent des défauts d'orientation des extensions spécifiquement dans les rangées les plus postérieures (Fig 32). Des larves *ds* de troisième stade présentent des défauts d'orientation affectant essentiellement les rangées 1 à 4 (Repiso, Saavedra et al. 2010). Ces résultats vont dans le sens d'une participation spécifique de *ds* dans certaines rangées. On remarque un phénotype additionnel de pigmentation pour les déficiences, qu'on ne retrouve pas dans les mutants ou trans-hétérozygotes, suggérant que ce défaut est dû à un autre gène que *ds* (Fig 32).

Pour *Vang*, la déficience entraîne une absence quasi-totale d'extensions, défaut qu'on ne retrouve pas dans les contextes mutant ou trans-hétérozygote. Dans ces 2 autres conditions, on observe des défauts d'orientation affectant aléatoirement les différentes rangées de denticules (Fig 32). Ces résultats impliquent clairement *Vang* dans la PCP épidermique et valident donc la participation du module CORE dans ce tissu.

#### *Rôle de Ds et Vang dans la PCP des cellules épidermiques embryonnaires ventrales*

Afin de déterminer pour chaque rangée le type de défaut associé, des immunomarquages des contours cellulaires et des faisceaux d'actine ont été effectués dans les différentes conditions (Fig 33). Pour chaque condition, j'ai ensuite quantifié différents paramètres.

L'**orientation** est estimée par la mesure de l'angle entre l'axe antéro-postérieur de la cellule et celui de l'extension. Les valeurs d'angle sont ensuite regroupées par cadran de  $22.5^\circ$ . Pour les défauts d'**implantation**, on attribue une position à l'extension au sein de 3 régions (antérieur, centre, postérieur) de la face apicale de chaque cellule. Enfin, le **nombre**



**Figure 34. Des phénotypes différents pour *Vang* et *ds*.**

**A. Défauts d'orientation des extensions.**

Alors que toutes les extensions pointent vers la région postérieure dans des embryons *wt*, les denticules des embryons *Vang* présentent une orientation aléatoire pour toutes les rangées. Les denticules des embryons *ds* présentent une réversion progressive de leur orientation.

**B. Défauts de nombre d'extensions par cellules.**

Le nombre d'extensions est diminué dans les rangées 1, 2, 3, 4 et 6 pour *ds* et rangées 1 et 2 pour *Vang*.

**C. Défauts d'implantation des extensions.**

Les denticules des rangées 2, 3 et 4 des embryons *Vang* présentent une centralisation de leur implantation dans la cellule, alors qu'on observe un décalage progressif de l'implantation des denticules des embryons *ds* vers la région antérieure de la cellule.

**d'extension par cellule** est compté. Pour chaque paramètre, les quantifications (une dizaine d'embryons par condition) sont séparées selon les rangées cellulaires.

### **Orientation**

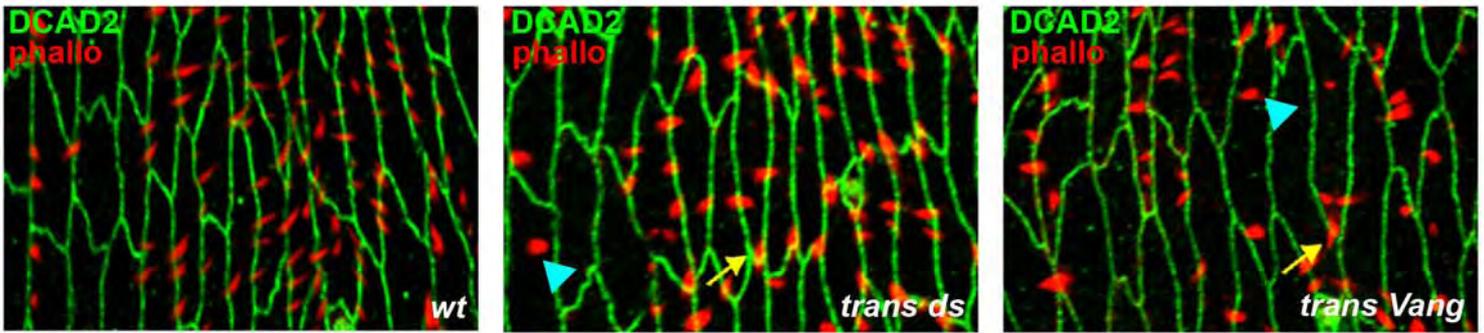
Dans des embryons *wt*, les extensions sont toutes orientées vers la partie postérieure de la cellule. En absence de *ds*, si les extensions des rangées 1, 2 et 3 pointent globalement vers la partie postérieure, l'orientation des rangées 4, 5 et 6 se retourne progressivement vers l'anérieur (Fig 33, 34A). Concernant *Vang*, les extensions présentent également des défauts d'orientation, mais sans inversion totale de la polarité (Fig 33, 34A). Ce phénotype différent affecte l'ensemble des rangées. Bien que les deux gènes soient impliqués dans l'orientation des denticules, la différence entre les deux phénotypes suggère un rôle de *Vang* sur l'ensemble des cellules à denticules et un rôle « rangée-spécifique » de *ds*.

### **Nombre**

Le nombre d'extensions par cellule est significativement diminué pour les embryons *ds*, dans toutes les rangées sauf la rangée 5 (Fig 34B). Des embryons *Vang* présentent le même défaut, mais limité aux rangées 1 et 2 (Fig 34B). *ds* et *Vang* participent donc à la définition du site d'initiation des extensions.

### **Implantation**

Plus de 80% des extensions sont normalement localisées dans la région postérieure dans les cellules sauvages (Fig 34C). L'absence de *ds* entraîne une modification progressive au fil des rangées de la localisation des extensions. Les extensions des rangées 3 et 4 se situent dans la partie centrale de la cellule, puis deviennent clairement localisées dans la partie antérieure pour les rangées 5 et 6 (Fig 33, 34C). Dans le cas de *Vang*, on observe un



**Figure 35. *Vang* et *ds* participent à l'orientation et l'implantation des trichomes.**

Immunocoloration de la DE-cadhérine (DCAD2 en vert) et des faisceaux d'actine (phallo en rouge) sur des embryons trans-hétérozygotes *ds* ou *Vang*. Les extensions présentent des défauts de polarité (flèches jaunes) et d'implantation (triangles bleus), touchant toutes les rangées aléatoirement.

déplacement vers le centre de la face apicale, pour les extensions principalement des rangées 2, 3 et 4 (Fig 33, 34C). Nos résultats sont donc en bon accord avec l'interprétation des phénotypes *ds* et *fat* (Marcinkevicius and Zallen 2013) dans l'épiderme embryonnaire.

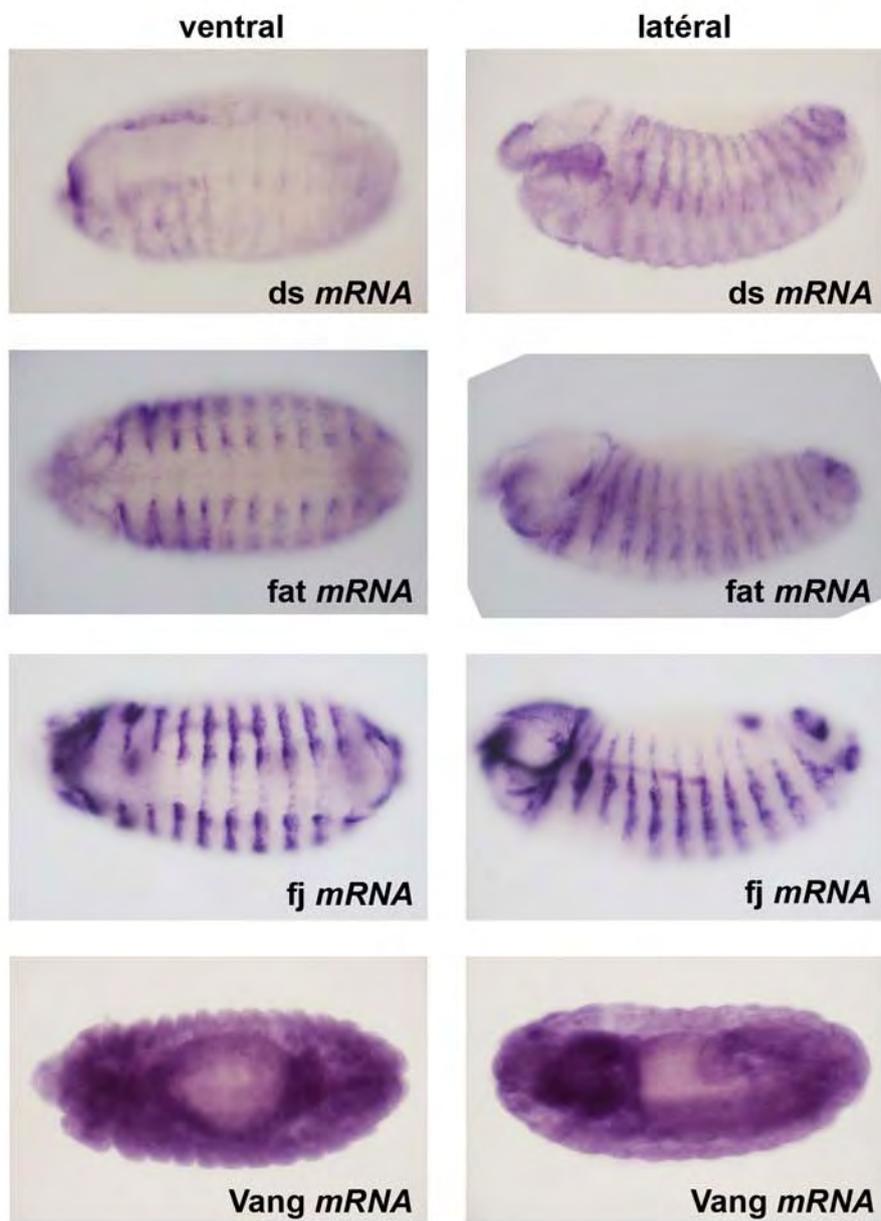
**Nos résultats montrent l'implication des modules GLOBAL et CORE de la PCP pour la morphogénèse des cellules à trichome en fin d'embryogénèse.**

#### *Un rôle de ds et Vang dans la PCP des cellules dorsales*

Nos analyses, comme l'ensemble des études publiées aujourd'hui concernant l'épiderme embryonnaire, ont porté sur les denticules, c'est-à-dire les trichomes de la région ventrale de l'embryon. Par contre, on ignore entièrement le rôle éventuel de la PCP, et des différents acteurs de la voie, pour les trichomes de la région dorsale. L'implantation des trichomes dorsaux est pourtant elle aussi stéréotypée, et propose un paradigme intéressant pour la PCP embryonnaire. Je me suis donc attachée à analyser l'impact de différents acteurs de la PCP sur le patron des trichomes de la région dorsale de l'embryon.

Léonie Védé dans l'équipe avait commencé à étudier le patron des trichomes dorsaux. Chaque segment abdominal est composé d'une rangée de cellules avec des extensions larges et robustes pointant plutôt vers l'antérieur. La rangée 2 est composée de cellules lisses, puis la région 3 est composée de 3 à 4 rangées de cellules présentant des extensions de taille intermédiaire dirigées vers le postérieur. Le début de région 4 (4a) est constitué de 3 rangées de cellules avec des extensions très fines, orientées vers le postérieur, immédiatement suivi par 3 à 4 rangées de cellules (4b) avec des extensions orientées cette fois vers l'antérieur. J'ai confirmé ces données par marquage des contours cellulaires et de l'actine (Fig 35).

Les extensions dorsales d'embryons *ds* présentent des défauts d'orientation différents des extensions ventrales. L'absence de *ds* affecte toutes les rangées, et les extensions ne



**Figure 36. Patron d'expression des acteurs de la PCP.**

Hybridation *in situ* des gènes de PCP *ds*, *Vang*, *fj* et *fat*. Pour chaque gène une vue ventrale et une vue latérale sont présentées. *ds*, *fj* et *fat* présentent un patron épidermique stéréotypé, alors que *Vang* est exprimé de façon ubiquitaire.

présentent pas de réversion totale de l'orientation (Fig 35). Les extensions *ds* sont également implantées aléatoirement selon l'axe A/P de la surface apicale de la cellule (Fig 35).

La mutation *Vang* entraîne un phénotype plus fort dans le dos que dans le ventre. En absence de *Vang*, les extensions dorsales sont orientées aléatoirement vers l'antérieur ou le postérieur, et sont implantées au centre de la cellule (Fig 35). Les quantifications sont en cours afin de déterminer pour chaque rangée le type de défaut de PCP, mais il apparaît clairement, et pour la première fois à notre connaissance, que les modules GLOBAL et CORE sont impliqués dans la PCP des cellules épidermiques dorsales.

*Les défauts observés pour ds et Vang sont-ils liés à leur domaine d'expression ?*

Nos travaux montrent donc le rôle important de *ds* et *Vang* pour la PCP de l'épiderme embryonnaire ventral et dorsal. Les phénotypes distincts suggèrent toutefois des rôles différents des deux modules dans la PCP embryonnaire, possiblement dû à une expression différentielle. Afin de tester cette hypothèse, j'ai donc défini par hybridation *in situ* le profil d'expression normal de *ds* et *Vang* au cours de l'embryogénèse. *Vang* présente un patron ubiquitaire, qui corrèle avec les défauts observés dans l'ensemble des cellules épidermiques (Fig 36). A l'inverse, *ds* présente un patron restreint dans les cellules à denticules (Fig 36). J'ai de plus analysé l'expression de *fat* et *ffj*, les partenaires de *ds* dans le module GLOBAL, qui sont également exprimés de façon restreinte dans l'épiderme (Fig 36). On voit donc une expression tissu-spécifique des acteurs du module GLOBAL, un résultat inattendu par rapport à la vision actuelle de son rôle sur la polarité de l'ensemble d'un tissu (Peng and Axelrod 2012).

Des hybridations *in situ* fluorescentes couplées à l'imagerie confocale sont en cours afin de définir si ces patrons sont restreints à certaines rangées de cellules à trichomes, et leur possible relation avec les phénotypes rangées-spécifiques observés pour *ds* et *fat*.



L'ensemble de nos résultats apporte donc une nouvelle vision la PCP des cellules épidermiques embryonnaires. On voit les rôles différents des modules GLOBAL et CORE au sein d'un même tissu, et entre deux régions différentes de l'épiderme embryonnaire. La mise en évidence d'un contrôle transcriptionnel de l'expression des différents membres de la voie PCP ouvre aussi de nouvelles perspectives.



## DISCUSSION

### De nouveaux processus régulés par Svb participent à la morphogénèse

#### *Svb régulerait au moins en partie la sécrétion des ZPs dans les extensions*

Parmi les gènes dont l'expression est diminuée en absence de Svb on note la présence de composants de la voie de sécrétion : *sec23*, *sec13* et *epsilon-cop*. En parallèle, Hélène a identifié *sec24like* (*sten*), un autre composant des vésicules COPII, et *sar1* comme indispensables à la distribution correcte de la protéine ZP Dusky-like. J'observe des défauts similaires pour les mutants *sec23*, c'est-à-dire une délocalisation des protéines ZP Dyl et M.

Des interactions physiques ont été détectées entre *sec23* et *Rab23* (Guruharsha, Rual et al. 2011), dont nous avons vu plus haut que sa transcription est aussi régulée par Svb pour être exprimée dans les cellules à trichomes. De même *Sec23* interagit avec *WASp*, un facteur régulant l'organisation de l'actine et dont l'expression est aussi régulée par Svb (Chanut-Delalande, Fernandes et al. 2006). Svb pourrait donc ainsi contrôler, ou modifier l'activité de la voie de sécrétion, pour le transport localisé de différents facteurs cellulaires vers les extensions en formation. Une étude est en cours dans l'équipe afin de mieux caractériser l'implication de la voie de sécrétion dans la morphogénèse des trichomes, notamment pour définir s'il s'agit de la voie classique de sécrétion, ou une voie spécifique pour la formation des denticules.

#### *Un nouveau rôle pour Svb, limiter la formation d'extensions ectopiques*

Grâce au recoupement des résultats de transcriptome *pri* et *svb* j'ai pu identifier de nouveaux acteurs de la morphogénèse épidermique, notamment des acteurs de la voie PCP. En effet l'expression épidermique de *Rab23*, *fng* et *mwh* est régulée par Svb : i) en absence de Svb on observe une diminution de leur expression dans l'épiderme, ii) l'expression ectopique de Svb dans les cellules lisses est suffisante pour induire une activation de ces gènes dans les mêmes rangées cellulaires. Dans les mutants *Rab23*, *fng* ou *mwh*, on observe des extensions



ectopiques. Ces résultats révèlent donc un rôle additionnel de Svb pour la restriction du nombre d'extensions par cellule, via la régulation transcriptionnelle de *Rab23*, *fng* et *mwh*.

Il a été proposé que Mwh agisse par une inhibition de la polymérisation d'actine (Strutt and Warrington 2008). En son absence, on observe de nombreux faisceaux de filaments d'actine, ce qui conduit à la formation de multiples extensions anormales par cellule. La régulation de la localisation de Mwh serait due aux protéines du module CORE (Strutt and Warrington 2008). On peut donc penser que le rôle de Svb dans la mise en place de la PCP dans l'épiderme embryonnaire serait de contrôler l'expression de *mwh*, et que la localisation de la protéine Mwh est ensuite régulée par d'autres acteurs de la PCP.

L'implication de *Rab23* dans la PCP a été mise en évidence dans l'aile. En son absence, des extensions supplémentaires sont présentes, mais la localisation des autres composants de la PCP n'est pas affectée, ce qui suggère que *Rab23* est un effecteur terminal de la voie de signalisation (Pataki, Matussek et al. 2010). L'analyse du phénotype *Rab23* dans l'épiderme embryonnaire amène des conclusions similaires, avec la production d'extensions ectopiques généralement dans la rangée 1 des cellules à denticules. *Rab23* fait partie de la famille des petites GTPases Rab, contrôlant le transport vésiculaire (Orlando and Guo 2009). Une mutation affectant le domaine GTPase de *Rab23* entraîne le même phénotype d'extensions supplémentaires qu'une mutation nulle, ce qui suggère que c'est bien via son implication dans le transport vésiculaire que *Rab23* participe à la PCP. Dans l'aile, la mutation de *Rab23* entraîne un défaut de distribution de Pk et Vang, soutenant l'hypothèse d'une implication de cette petite GTPase dans le transport des composants du module CORE de la PCP dont la localisation asymétrique dans la cellule définit le site d'initiation de l'extension.

Fng est une enzyme responsable de la glycosylation de Notch, ce qui conduit à une modulation de son affinité pour le ligand Delta (Rana and Haltiwanger 2011). Le rôle le mieux connu de Fng chez la drosophile est la définition de la bordure dorso-ventrale dans



différents tissus, comme le disque d'aile et d'œil, par la modulation de l'interaction Notch-Delta (Wu and Rao 1999). Une étude précédente avait identifié le rôle de Zipper (MyosineII) et Fng dans le contrôle du site d'implantation des denticules embryonnaires (Walters, Dilks et al. 2006). En effet, les embryons mutants pour *fng* présentent de petites denticules ectopiques, qui semblent provenir de la division des extensions normales. Ces denticules présentent aussi des défauts d'orientation, ce qui suggère une implication de *fng* dans la PCP. Le phénotype sera mieux décrit lors d'analyses par d'imagerie confocale du nombre d'extensions, ainsi que leur implantation et orientation. Nous pourrons alors conclure quant au rôle de *fng* dans la PCP des cellules épidermique.

Ces résultats issus d'analyses de transcriptomes nous ont permis d'identifier de nouvelles cibles de Svb. Couplé aux résultats de l'analyse génétique de l'ensemble du chromosome II de la drosophile, nos travaux éclairent d'un jour nouveau les rôles et mécanismes de la PCP.

### **Implication des modules GLOBAL et CORE ; rôles majeurs de ds et Vang.**

Dans l'œil et l'aile adulte, les cadhérines atypiques Ds et Fat établissent la polarité générale du tissu, en définissant le gradient du module CORE Fz/Stan. Vang est une protéine trans-membranaire, et fait partie du module CORE. La protéine est localisée de façon asymétrique au niveau de la membrane plasmique, et régule ainsi le recrutement asymétrique d'effecteurs cytoplasmiques comme Dishevelled et Diego. Dans d'autres tissus, le lien entre les deux modules peut être différent. Par exemple, dans l'abdomen adulte les deux modules agissent en parallèle et indépendamment (Casal, Lawrence et al. 2006; Adler 2012). Le rôle de Vang comme membre du module CORE est également bien décrit dans l'oreille interne des mammifères (Ezan and Montcouquiol 2013). Que ce soit dans la cochlée des mammifères ou l'aile de drosophile, le complexe formé par Vang, Pk et Stan inhibe la formation d'extensions dans la partie proximale de la cellule (Eaton 1997; Strutt and Strutt



2009). Les 6 protéines composant le module CORE sont d'ailleurs conservées entre la drosophile et les mammifères (Ezan and Montcouquiol 2013). Alors qu'aucun rôle du module CORE n'avait été précédemment décelé dans l'épiderme embryonnaire (Repiso, Saavedra et al. 2010), nos résultats montrent clairement qu'en absence de Vang les extensions présentent des défauts d'orientation et d'implantation, que ce soient les denticules ventrales ou les trichomes dorsaux.

De même, en absence de *ds* on observe un phénotype de réversion progressive de polarité et d'implantation des extensions, l'effet le plus fort étant observé pour les rangées 5 et 6. De façon intéressante, un rôle de *ds* dans la PCP épidermique a été démontré au troisième stade larvaire (Casal, Lawrence et al. 2006). Une sur-expression de *ds* entraîne des défauts cellulaires autonomes et non- autonomes de polarité. Ces résultats suggèrent que le gradient *fat/ds* est également impliqué dans la mise en place de la PCP dans l'épiderme lors du troisième stade larvaire (Repiso, Saavedra et al. 2010). Nos analyses effectuées durant la mise en place, au cours de l'embryogenèse, des trichomes des larves de premier stade, montrent également un rôle de *ds* à un stade plus précoce. Alors qu'au troisième stade larvaire les défauts touchent les rangées 1 à 4 (les denticules 5 et 6 sont normales en absence de *ds*), nos résultats montrent un phénotype progressif, avec les effets les plus forts dans les rangées postérieures. *Ds* semble donc agir de manière différente dans la PCP épidermique, en fonction des stades observés. Dans l'embryon, *ds* présente un patron d'expression particulier avec une expression renforcée dans l'épiderme au niveau des cellules à denticules. Nos données actuelles ne permettent pas de définir le répertoire exact de l'expression de *ds*, mais des études sont en cours afin de tester l'hypothèse selon laquelle *ds* serait majoritairement exprimé dans les cellules à denticules les plus postérieures, ce qui expliquerait le phénotype observé. Les effets « non-cellulaire autonome » observés au troisième stade larvaire (Repiso, Saavedra et al. 2010), pourraient contribuer à la réversion progressive de la polarité pour les rangées plus antérieures en absence de *Ds*.



Collectivement ces résultats démontrent un nouveau rôle de Svb dans la morphogénèse épidermique via la régulation de l'expression d'effecteurs de la PCP : *Rab23*, *fng* et *mwh*. Le rôle de ces protéines est de restreindre l'initiation de l'extension dans la région postérieure de la cellule. Nous avons également découvert l'implication des protéines Ds et Vang pour la formation des denticules embryonnaires. Ces gènes des modules GLOBAL et CORE de la voie PCP ne sont par contre pas régulés par Svb, alors qu'une partie au moins des effecteurs en aval dépend de Svb. Étant donné que certains autres acteurs de la PCP connus dans l'aile ne semblent pas impliqués dans l'épiderme embryonnaire (Hélène Chanut), il serait intéressant de tester l'ensemble des membres de cette voie afin de décrypter leurs variations inter-tissulaires et de définir de nouveaux acteurs de la morphogénèse des extensions épidermiques embryonnaires



## Partie III :

**Nouvelles fonctions embryonnaires des peptides Pri  
et identification des effecteurs cibles**



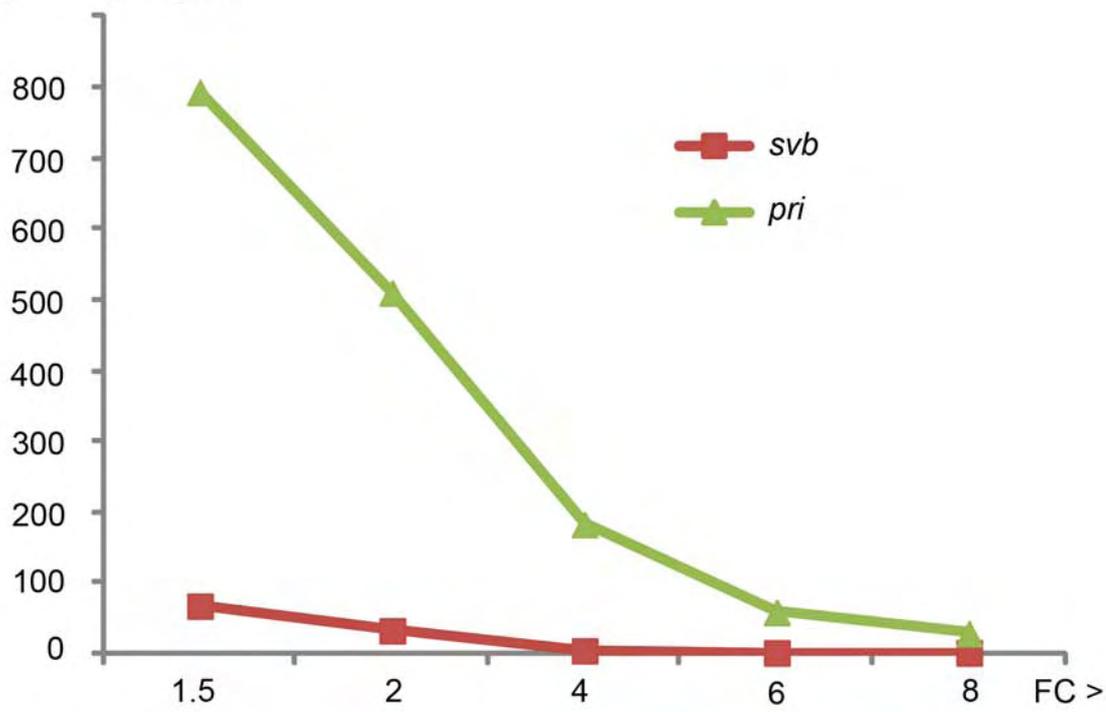
## INTRODUCTION

La motivation principale de l'analyse du transcriptome des embryons *pri* était de disposer d'une condition additionnelle pour faciliter l'identification des cibles du FT Shavenbaby, dont nous avons vu que l'activation dépend de l'activité des peptides Pri. Au-delà de cet objectif, finalement bien rempli par cette analyse, la bonne qualité de ces données nous a engagés à rechercher des gènes additionnels dont l'expression est modifiée (négativement ou positivement) dans les embryons *pri*, indépendamment de la question des cibles de Svb dans les cellules à trichomes.

Nous avons donc poursuivi une stratégie similaire : i) analyse des données de transcriptome, ii) définition d'une liste de candidats attractifs, iii) validation des données de puces par hybridation *in situ* dans des embryons de différents contextes mutants, iv) recherche de nouvelles voies/fonctions impliquées. Les expériences d'hybridation *in situ* ont bénéficié de l'aide précieuse d'Audrey Montigny.

Globalement ces approches ont été fructueuses et ont permis d'identifier de nombreux gènes dont l'expression dépend de l'activité des peptides Pri, indépendamment de leur rôle pour l'activation du facteur Svb. Ainsi, nous avons découvert différents ensembles de gènes down- ou up-régulés dans les embryons *pri*. De manière intéressante, ces données ont permis de découvrir de nouvelles fonctions embryonnaires de *pri*, notamment dans la différenciation du système respiratoire (les trachées) et de l'ensemble du tissu épidermique (en plus de son action spécifique dans les cellules à trichomes). Elles montrent l'importance des peptides Pri pour coordonner la temporalité d'expression d'un grand nombre d'effecteurs de la différenciation terminale, dans différents tissus embryonnaires, via des mécanismes transcriptionnels différents (dépendants ou non du FT Svb).

# gènes up-régulés



**Figure 37. De nombreux gènes up-régulés en absence de *pri*.**

Représentation graphique du nombre de gènes up-régulés en absence de *pri* (vert) ou *svb* (rouge) selon le seuil donné au Fold Change (FC).

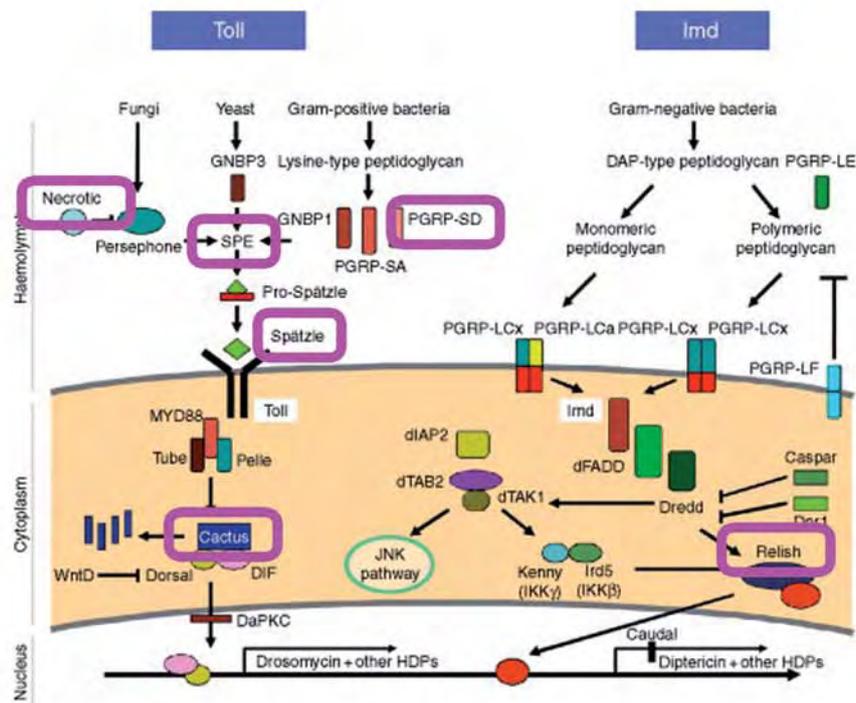
## RESULTATS

### Rôle de Pri dans la différenciation des trachées.

De nombreux gènes de la famille *Osiris* (24 gènes spécifiques des insectes), sont modulés en absence de *pri* ou *svb*. Les gènes *Osiris* sont en général exprimés dans 2 territoires : les trachées pour *Osi 20*, *Osi 8*, *Osi 19*, et *Osi 18*, ou l'épiderme et les trachées pour *Osi 14*. Organisés en clusters de 6 ou 7 gènes contigus sur le génome, leur fonction reste encore inconnue ; leur restriction aux arthropodes ainsi que leur expression suggèrent cependant un rôle dans la mise en place de la cuticule. Une diminution d'expression en absence de *pri* des gènes *Osiris* présents dans les trachées a été validée par hybridation *in situ*, montrant bien une influence de Pri sur ces gènes. *Osi 14* est une nouvelle cible de Svb, montrant une forte diminution d'expression épidermique en hybridation *in situ* dans les embryons *svb* (Fig 25). *Osi 14* est également exprimé dans les trachées, il serait donc intéressant de tester l'implication de Pri sur son expression dans ce tissu.

### A la recherche des gènes dont l'expression est augmentée...

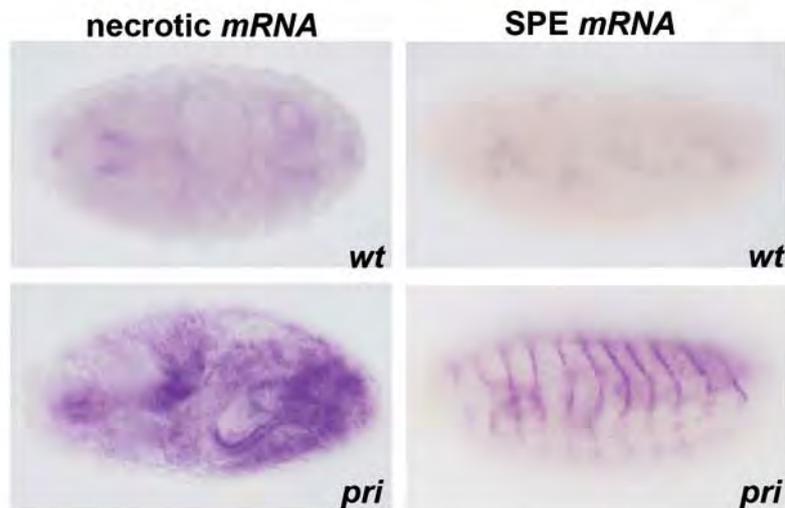
La fonction activatrice de Svb est indispensable pour la formation des trichomes. L'expression artificielle d'une forme exclusivement répresseur est en effet suffisante pour inhiber leur formation (dans un contexte par ailleurs normal). Il semble donc que les effecteurs indispensables correspondent aux gènes down-régulés en absence de Svb (et/ou Pri). Ainsi, il n'existe que très peu de gènes présentant une up-régulation en absence de Svb (Fig 17C). En utilisant des critères volontairement relâchés ( $\log FC > 0.5$ ,  $pvalue > 0.05$ ), on obtient une liste de 76 gènes. Il est à noter que ces gènes ne présentent pas d'expression épidermique connue, confortant l'hypothèse qu'ils ne jouent pas un rôle majeur dans la morphogénèse des trichomes. Il reste cependant possible que le FT Svb ait une fonction importante comme répresseur de transcription, à des étapes plus précoces de l'embryogénèse que nous n'avons pas étudiées par nos microarrays.



**Figure 38. Voie de signalisation spz/Toll chez *Drosophila melanogaster*.**

Cette voie est activée par la présence de champignons, bactéries ou levures dans l'hémolymphe, et permet l'induction de l'expression de molécules antifongiques ou antibactériennes en réponse à la reconnaissance de ces pathogènes. Les acteurs dont l'expression est augmentée en absence de Pri sont représentés par des rectangles violets.

D'après Walshe, Haines 2009.



**Figure 39. Les expressions de *necrotic* et *SPE* sont augmentées en absence de *pri*.**

Hybridation *in situ* pour les gènes *necrotic* et *SPE* dans des embryons *pri*. L'expression de ces gènes est augmentée de façon ubiquitaire pour *necrotic* et dans l'épiderme pour *SPE*.

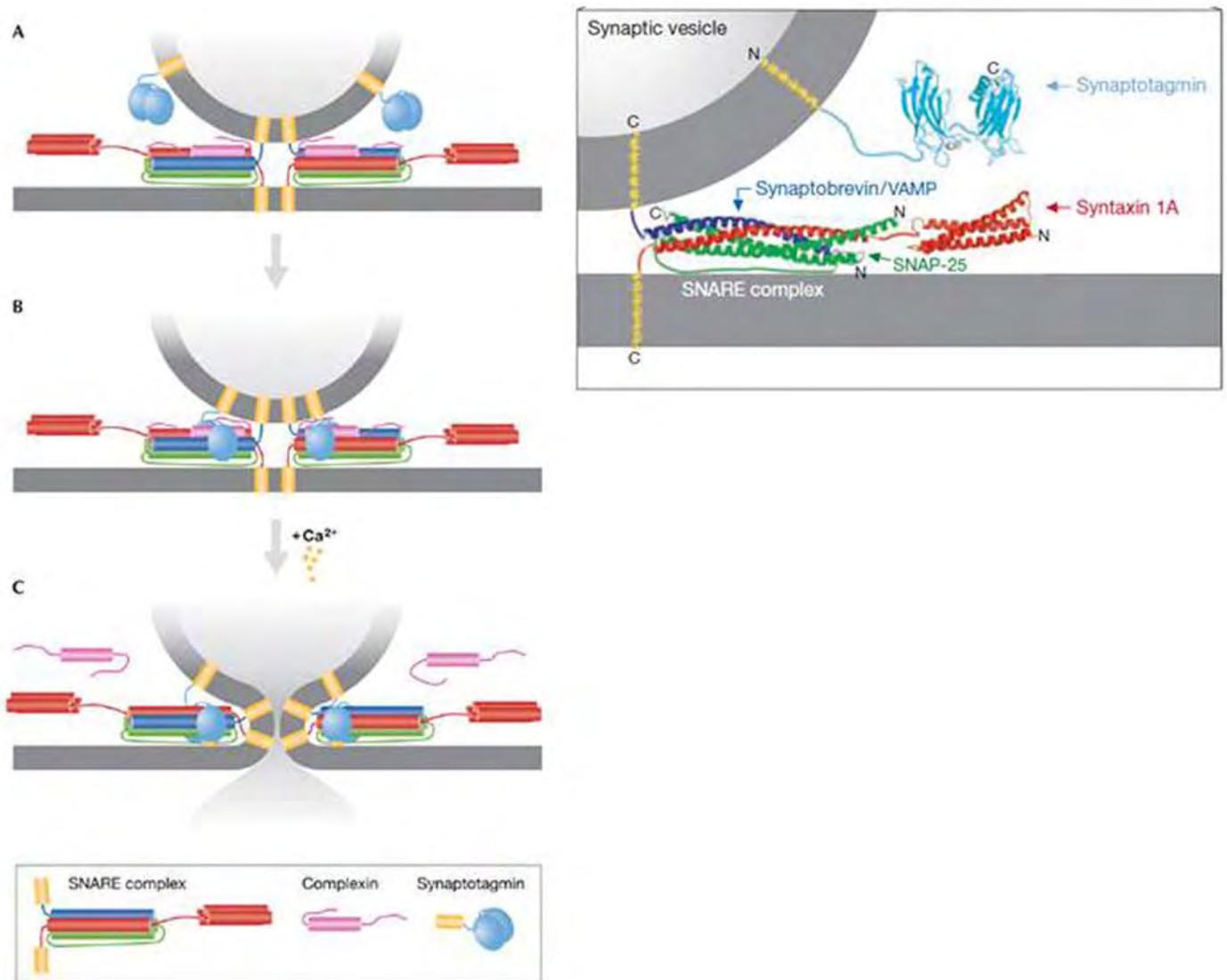
### *De nombreux gènes surexprimés dans des embryons pri.*

En absence de gènes connus permettant de mettre en place un « benchmarking », nous nous sommes basés sur des critères statistiques pour identifier les gènes sur-exprimés dans les embryons *pri*. De façon intéressante et contrairement à ce qui a été observé pour les embryons *svb*, de nombreux gènes sont sur-exprimés en absence de *pri*, et ceci de façon extrêmement robuste. Ainsi, alors que seulement 5 gènes présentent un Fold Change de 4 en absence de *svb* (FC>4, pvalue>0.05), plus de 180 sont sur-exprimés en absence de *pri* dans les même conditions (Fig 37). Pour un Fold Change de 2, la liste en embryons *pri* contient 510 gènes (Fig 37). Ainsi le nombre de gènes affectés par l'absence de Pri est beaucoup plus grand que le nombre de gènes affectés par l'absence de Svb (Fig 17, 37). Cette observation suggère que Pri régule d'autres processus lors de la différenciation terminale des embryons, ces processus étant indépendants de Svb. Ce résultat est d'ailleurs largement compatible avec l'expression de *pri* dans des tissus embryonnaires ou *svb* n'est pas exprimé.

### *Les peptides Pri et la réponse immunitaire chez la drosophile.*

Parmi les gènes dont l'expression est augmentée en absence de Pri, on remarque un enrichissement en acteurs de la voie Toll / spatzle, activée en réponse à une infection par des bactéries Gram+. On retrouve ainsi *necrotic* (*nec*), l'enzyme SPE (Spatzle Processing Enzyme), la protéine PGRP-SD qui reconnaît des peptidoglycans, le ligand de Spatzle, Cactus, qui participe à la synthèse de protéines antifongiques, et enfin Relish, un facteur de transcription qui active l'expression de gènes anti-microbiens (Fig 38).

Une up-régulation des ARNm *SPE* spécifiquement dans l'épiderme, et une plus ubiquitaire pour *necrotic* ont été validées par hybridation *in situ* (Fig 39). Cet enrichissement en gènes participant à la réponse immunitaire présentant une forte hausse de leur expression en absence de Pri laisse à penser que ces peptides participent à la régulation de la réponse immunitaire chez la drosophile.



**Figure 40. L'expression des composants de la transmission synaptique est régulée par Pri.** Interactions entre les différents acteurs impliqués dans l'arrimage et la fusion des membranes entre une vésicule synaptique et la membrane plasmique. L'expression de l'ensemble de ces gènes est augmentée dans le contexte mutant *pri*. D'après Carr, Munson 2007.

### *Un rôle de Pri dans la transmission synaptique ?*

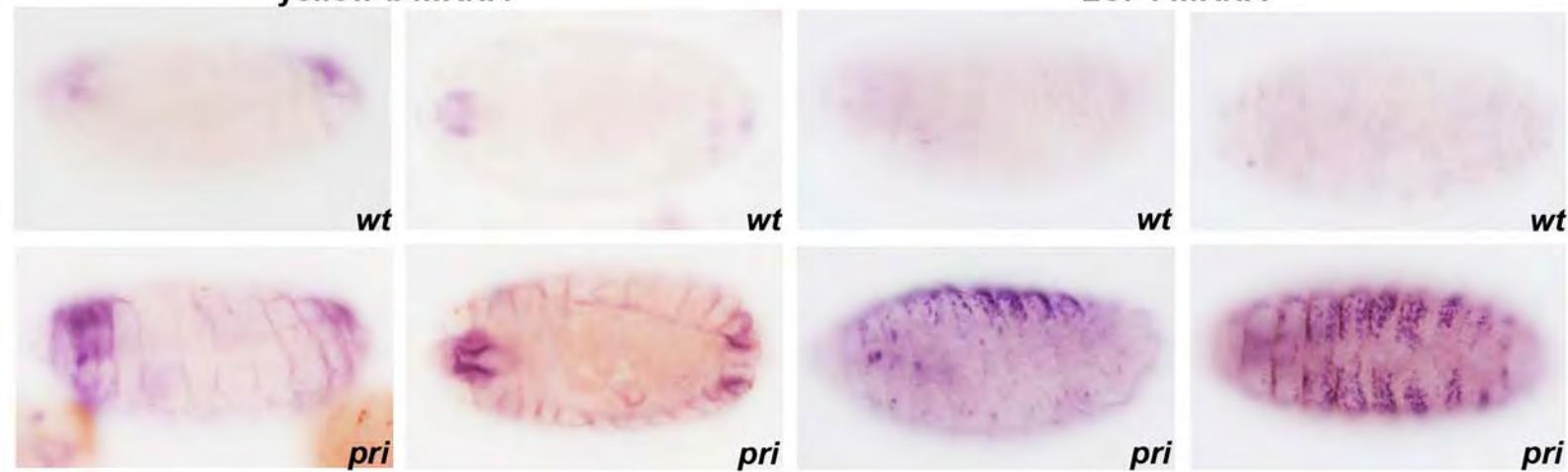
Nous avons aussi identifié des gènes codants des facteurs nécessaires à la transmission synaptique up-régulés dans les embryons *pri*. Au cours de l'embryogénèse, il existe une expression transitoire de *pri* dans le système nerveux central, notamment les ganglions procéphaliques. Les facteurs dérégulés en absence de *pri* sont des composants de la vésicule ou de la membrane pré-synaptique, dont la mutation individuelle entraîne des défauts de transmission synaptique. Une vésicule synaptique est composée de différentes protéines trans-membranaires : les Synaptotagmines-1 et 2 (Syt1 et 2), les Synaptobrévines-1 et 2 (VAMPs-1 et 2), les Synaptogyrines-1 et 3, les Synapsines-1 à 3, ainsi que la petite GTPase Rab3. En réponse à un potentiel d'action, Syt1 est un capteur de Ca<sup>2+</sup> qui va établir un complexe avec VAMP, ainsi qu'avec les protéines liées à la membrane plasmique : la Syntaxine-1 et SNAP-25. On parle du complexe SNARE qui arrime la vésicule à la membrane plasmique. La Complexine (Cpx) se lie ensuite au complexe SNARE et le stabilise, ce qui entraîne la fusion des membranes et la libération de neurotransmetteurs dans la fente synaptique (Carr and Munson 2007) (Fig 40).

Parmi l'ensemble de ces acteurs, les gènes *synaptogyrine-1*, *synapsine-1*, *Rab3*, *synaptotagmine-1*, *syntaxine-1*, *SNAP25* et *complexine* sont dérégulés en absence de Pri. Ainsi ces résultats ouvrent la possibilité d'une implication des peptides Pri dans la transmission synaptique. Des mesures physiologiques pourraient permettre de tester les conséquences de l'absence de Pri sur l'activité synaptique.

Par ailleurs, un rôle de la Syntaxine-1 (Syx1A) a été mis en évidence dans la morphogénèse épidermique chez la drosophile. L'inactivation de *syx1A* n'altère pas la sécrétion de chitine, mais les embryons présentent une cuticule altérée, incluant l'absence d'épicuticule (Moussian, Veerkamp et al. 2007). Il est donc possible que la dérégulation de Syx1A contribue au moins en partie aux altérations de synthèse/assemblage de la cuticule qu'on observe en absence d'activité des peptides Pri.

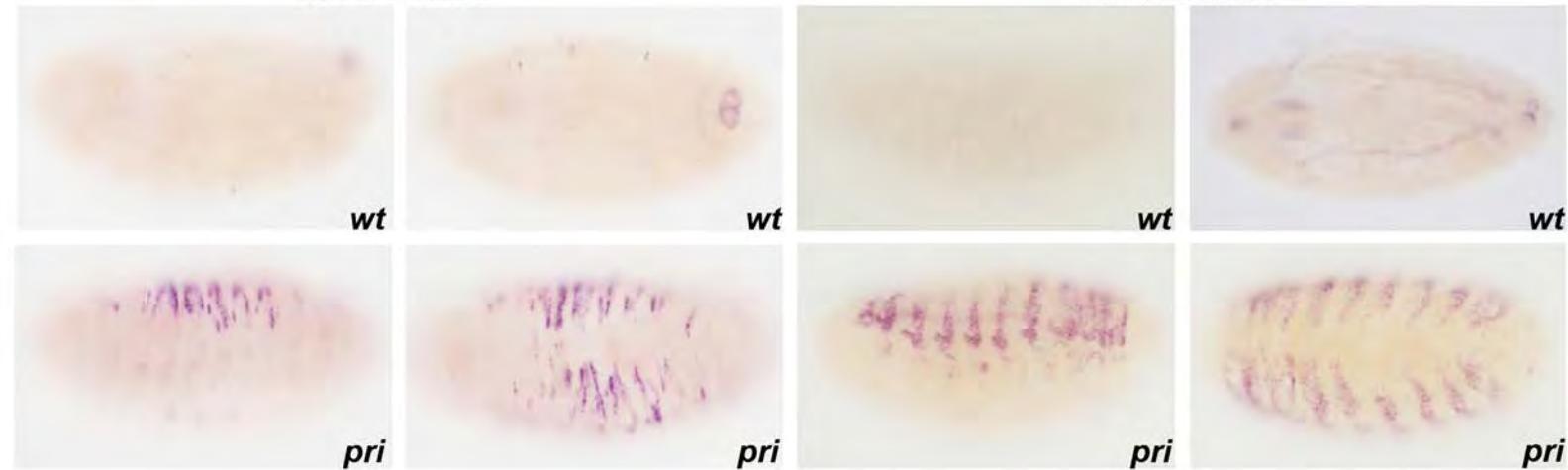
**yellow-b mRNA**

**LCP1 mRNA**



**LCP2 mRNA**

**LCP3 mRNA**

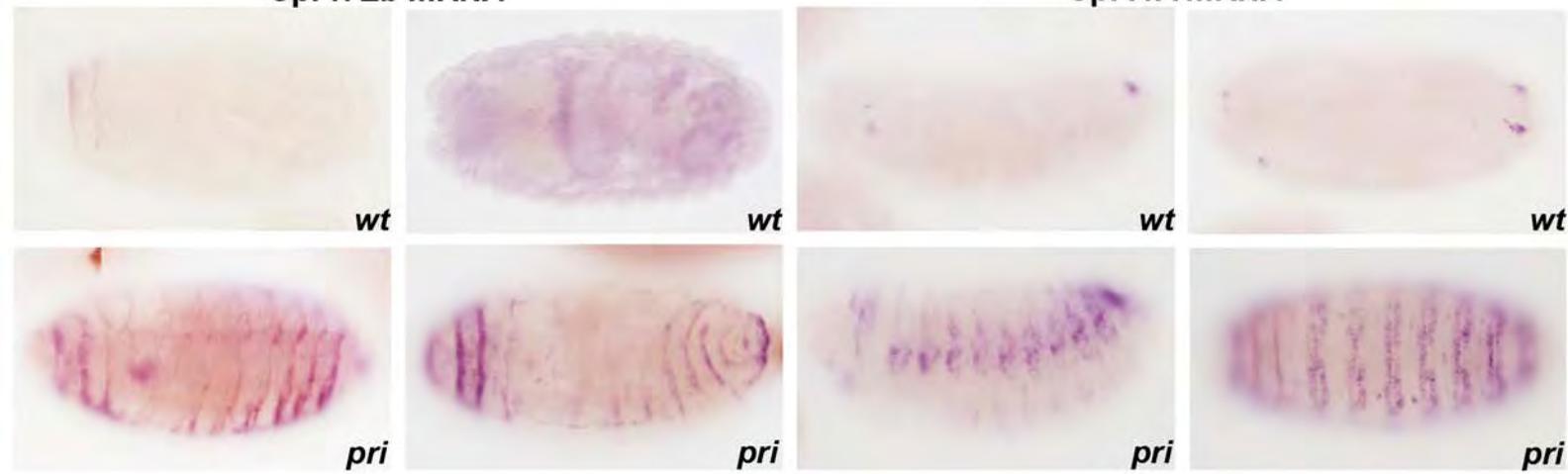
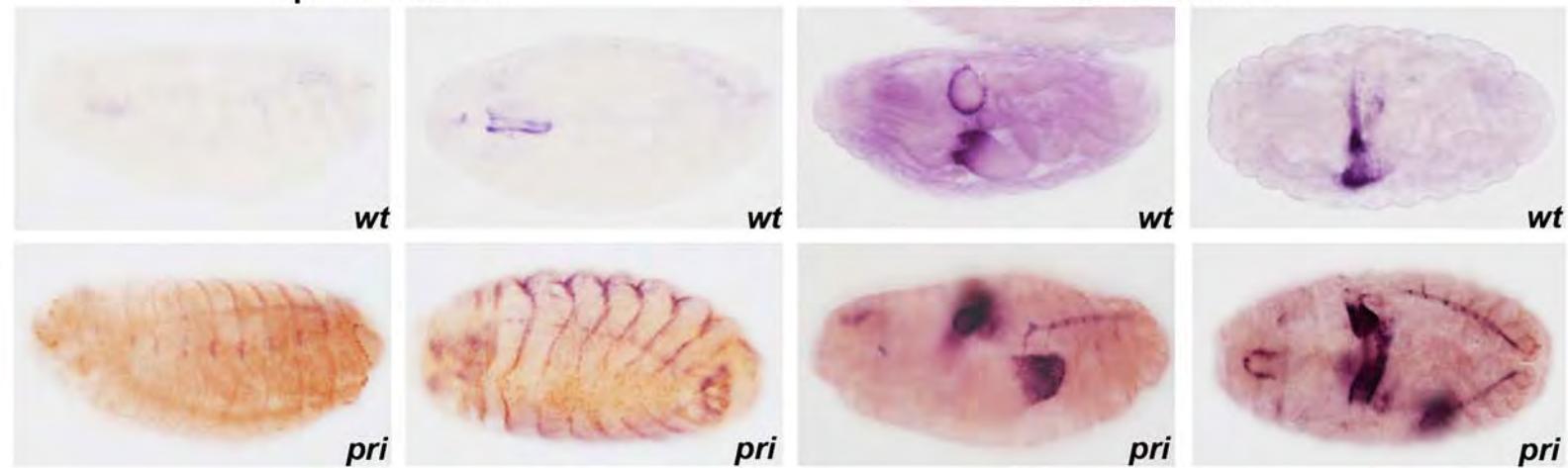


**Figure 41. Les peptides Pri régulent l'expression de composants de la cuticule.**

Hybridation *in situ* pour des gènes dont l'expression est augmentée dans le contexte mutant *pri*. Pour chaque candidat une vue latérale et une vue ventrale ou dorsale sont présentées. L'expression de *yellow-b* est augmentée dans l'épiderme et les trachées en absence de Pri. Les gènes *LCP1*, *LCP2* et *LCP3* présentent une augmentation de leur expression dans l'épiderme en absence de Pri.

## Les peptides Pri régulent l'expression de composants de la cuticule.

De nombreux composants avérés ou présumés de la cuticule sont également présents dans les gènes sur-exprimés en absence de Pri. Alors que le gène *yellow* est une cible avérée de Svb, on retrouve ici 2 nouveaux membres de la famille des protéines Yellow : *Yellow-e2* et *Yellow-b*. Il est cependant à noter une différence de comportement entre ces deux gènes : alors que *yellow-b* est exprimé dans les embryons sauvages, le gène *yellow-e2* lui ne semble pas être normalement exprimé au cours de l'embryogenèse. On peut donc imaginer deux types de dérégulations différentes en absence de Pri : des gènes exprimés à un niveau supérieur comme *yellow-b*, et des gènes normalement absents mais dont l'expression est activée comme *yellow-e2*. Bien que le rôle des protéines Yellow ne soit pas encore précisément élucidé, ces facteurs participent au processus de mélanisation des insectes. En effet l'absence du gène *yellow* entraîne une dépigmentation de différents tissus chez *Drosophila melanogaster* (Brehme 1941; Prud'homme, Gompel et al. 2006). L'analyse des gènes *yellow* chez le lépidoptère *Heliconius* montre que leur expression dans l'aile adulte corrèle avec les étapes de pigmentation et mélanisation du tissu (Ferguson, Green et al. 2011). La validation par hybridation *in situ* d'une augmentation de l'expression de *yellow-b* dans l'épiderme et les trachées aux stades tardifs en absence de Pri suggère une régulation d'une partie des gènes de la famille *yellow* par Pri (Fig 41). On retrouve également d'autres gènes impliqués dans la mise en place de la cuticule et qui présentent une up-régulation en absence de Pri. Ainsi la famille des Larval Cuticle Protein (LCP) présente une augmentation massive de leur expression en absence de Pri. Pour ces gènes, on observe une absence totale d'expression en condition sauvage et une hausse drastique de l'expression en absence de *pri*. Des expériences d'hybridation *in situ* nous ont permis de valider cette hausse d'expression en absence de Pri pour *LCP1*, *LCP2* et *LCP3*, ceci spécifiquement dans l'épiderme et plutôt en région dorsale, ce qui suggère une implication de Pri dans la régulation de la mise en place de la cuticule (Fig 41).

**Cpr47Eb mRNA****Cpr11A mRNA****Cpr78Cc mRNA****Muc18B mRNA****Figure 42. Les peptides Pri régulent l'expression de composants de la cuticule.**

Hybridation *in situ* pour des gènes dont l'expression est augmentée dans le contexte mutant *pri*. Pour chaque candidat une vue latérale et une vue ventrale ou dorsale sont présentées. L'expression de *Cpr47Eb*, *Cpr11A* et *Cpr78Cc* sont augmentées dans l'épiderme et les trachées en absence de Pri. Le gène *Muc18B* présente une augmentation d'expression dans les trachées en absence de Pri.

De plus, 14 gènes de la famille Cpr (Cuticle Protein Related) présentent une augmentation significative de leur expression en absence de Pri. L'apparition d'une expression épidermique des gènes *CPr11A*, *Cpr78Cc*, et *Cpr47Eb* dans cette condition a été validée par hybridation *in situ* (Fig 42) alors qu'ils présentent un niveau d'expression faible ou nul en condition sauvage.

Enfin, trois membres de la famille des mucines (*Muc18B*, *Muc55B* et *Muc26B*) présentent également une augmentation d'expression en absence de Pri, mais de façon moins drastique que les autres facteurs de la cuticule. L'augmentation d'expression de *Muc18B* dans les trachées en absence de Pri a d'ailleurs été validée par hybridation *in situ* (Fig 42). L'expression de 2 membres de la famille *obstructor* (*Obst-E* et *Obst-J*) est également augmentée en absence de Pri. Les Mucines sont des protéines constituant la matrice extracellulaire, et les protéines Obstructor sont des constituants de la membrane pérित्रophique. Ces deux familles de protéines contiennent des domaines de liaison à la chitine, ce qui suggère un rôle dans le maintien de la cuticule.

La « néo-expression » de ces gènes en absence de Pri, à un stade où la cuticule n'est pas encore synthétisée, soutient l'importance du contrôle temporel de la synthèse de la cuticule par Pri. Il serait intéressant d'étudier si le patron d'expression observé dans les embryons *pri* correspond à un profil d'expression spécifique qui devrait se mettre en route normalement à des étapes plus tardives, *i.e.* durant le développement larvaire, où s'il s'agit d'un profil normalement jamais exprimé au cours de la différenciation de l'épiderme. Les mutants *pri* possèdent une cuticule très altérée, qui pourrait provenir d'une expression prématurée ou/et dérégulée de ses constituants structuraux. Ainsi, nos données suggèrent que les peptides Pri servent de régulateur temporel de la synthèse de la cuticule.



Collectivement ces résultats révèlent donc l'existence de nouvelles fonctions des peptides Pri au cours de l'embryogenèse, étendant l'importance de ces petits peptides régulateurs pour le contrôle transcriptionnel du développement.



## DISCUSSION

L'analyse des conséquences transcriptomiques de l'inactivation du FT Shavenbaby et des peptides Pri récemment découverts, a permis d'améliorer significativement la connaissance des mécanismes régulés par ces facteurs. Pour les peptides Pri, ces travaux étendent leur implication potentielle à la transmission synaptique, l'immunité via la voie Toll, ainsi que dans la régulation temporelle de la mise en place de la cuticule. Ces données ouvrent de nouvelles voies pour la compréhension du rôle de longs ARNs codant des petits peptides dans le développement animal.

### De nouvelles fonctions attractives des peptides Pri ...

L'utilisation d'ARN extrait à partir d'embryons entiers a rendu difficile l'interprétation des données du transcriptome *svb*. En effet, plusieurs cibles avérées de Svb ne présentent pas de variation d'expression détectable par microarrays, sans doute à cause de territoires additionnels d'expression. Dans le cas du transcriptome *pri*, l'utilisation d'embryons entiers a par contre permis d'avoir une idée plus large des différentes implications de Pri à cette étape de l'embryogenèse.

#### ... dans la différenciation des trachées

En plus de la formation des trichomes, Pri est nécessaire à la différenciation des trachées, notamment la mise en place des *taenidia*, des structures riches en actine qui soutiennent la lumière des tubes respiratoire de l'embryon. Il est établi que la fonction de Pri dans les trachées est indépendante de Svb. Svb n'est en effet pas exprimé à ce stade dans les trachées, et surtout l'absence de Svb n'affecte pas la différenciation des trachées, en tout cas la mise en place des *taenidia*. Nos travaux ont permis d'identifier plusieurs gènes qui sont spécifiquement exprimés dans les trachées, et dont le niveau et/ou le timing d'expression est affecté dans les mutants *pri*. Ces premières données sont intéressantes et permettent de



découvrir de nouveaux gènes de la différenciation terminale du système respiratoire. Contrairement à la formation des trichomes, le rôle moléculaire de Pri pour la régulation de l'expression de ces effecteurs moléculaires n'est pas connu. Il est tentant d'inférer que, comme pour la formation des trichomes, la fonction transcriptionnelle de Pri dans les trachées passe par la modification protéolytique d'un régulateur de transcription. Il reste bien sûr possible qu'il s'agisse d'un mécanisme totalement différent.

*... dans la transmission synaptique*

En plus de ces données sur les médiateurs moléculaires de la fonction de Pri dans les trachées, nos travaux suggèrent une implication des peptides Pri dans la régulation de processus développementaux additionnels. La dérégulation des acteurs de la transmission synaptique suggère par exemple une nouvelle fonction aujourd'hui inconnue des peptides Pri dans le système nerveux central. Cette hypothèse est aussi compatible avec le profil d'expression de l'ARN *pri*, et une analyse plus détaillée des expressions de *pri* d'une part et des composants du complexe de transmission d'autre part sera donc intéressante. En fonction des résultats de ces analyses, une analyse phénotypique détaillée, incluant le test de la transmission nerveuse, pourrait être attractive.

*... dans la mise en place de la cuticule*

De nombreux composants structuraux de la cuticule (familles LCP, Cpr, Twdl, Mucines, Obstructor) sont dérégulés en absence de Pri, leur transcription étant fortement augmentée. Les enzymes de synthèse de la chitine, telles que la chitine syntase, ne sont pas cependant pas dérégulés en absence de Pri. Ainsi, les peptides Pri pourraient servir à réprimer l'expression de nombreuses protéines de la cuticule, tant que d'autres composants comme la chitine ne sont pas synthétisés et/ou correctement déposés. A la fin de l'embryogénèse l'expression de *pri* diminue, ce qui lève la répression sur les composants



structuraux dont l'expression est alors activée. Ils peuvent ensuite être sécrétés pour compléter la formation correcte de la cuticule. Via son rôle d'activateur de Svb, Pri pourrait ainsi permettre de coordonner temporellement la mise en place des extensions épidermiques ainsi que la formation générale de la cuticule.

**En conclusion, nos travaux ont permis d'identifier de nombreux effecteurs cellulaires de la différenciation épidermique. Nous ouvrons aussi de nouvelles pistes concernant les rôles des peptides Pri dans des processus indépendants de Svb et de la morphogénèse trichomes. Le contrôle temporel de Pri sur la mise en place de la cuticule ainsi que dans la transmission synaptique sont, pour moi, deux pistes particulièrement attrayantes.**



# CONCLUSION et PERSPECTIVES



Mes travaux de thèse ont porté principalement sur la nature et la logique des mécanismes régulant la transcription des gènes au cours du développement. Il s'agissait tout d'abord d'identifier les acteurs directs d'un processus de différenciation terminale, la morphogénèse du tissu épidermique de l'embryon de drosophile. D'autre part, nous voulions découvrir les éléments cis-régulateurs (enhancers) coordonnant le contrôle transcriptionnel de ce programme cellulaire. Enfin, un enjeu important était de découvrir le code fonctionnel de ces éléments contrôlant la transcription lors de l'embryogenèse tardive, notamment en comparaison des modèles généralement répandus bâtis sur l'étude des étapes précoces du développement. L'ambition de nos travaux a été de mettre en place, et d'analyser grâce une démarche multidisciplinaire, des approches à l'échelle du génome entier. Une question générale est quelle est en pratique la puissance heuristique de ces approches, potentiellement à même de réaliser une rupture majeure dans la collecte des données biologiques.

En effet il devient possible de concilier profondeur et étendue des analyses, avec une description détaillée de l'ensemble des éléments observés ou découverts *ab initio* par l'analyse. On substitue ainsi à une analyse qualitative, objet par objet (par exemple l'étude d'un gène), une mesure quantitative de comportement à un temps donné de l'ensemble de constituants du système (l'ensemble des gènes d'un organisme). Pour fixer les idées, rappelons par exemple pour l'analyse de l'expression génique, qu'il s'agirait d'un gain de plus de  $4 \log(10)!$  Cependant les limitations, biais, voire distorsions de la description des systèmes biologiques par ces approches génome entier ont été largement discutées et décriées.

Par-delà une certaine fascination de la manipulation numérique et d'une approche exhaustive, quels sont les apports et les pièges du genome-wide pour comprendre les mécanismes *in vivo* du développement ?



Différents éléments ont été déjà discutés dans chacun des chapitres de résultats et cette discussion générale s'articule autour de trois axes principaux. Le premier est une réflexion sur les évolutions techniques et conceptuelles permettant d'analyser les systèmes biologiques dans leur ensemble. Le second concerne la nature des mécanismes du contrôle de l'expression génique dans un cas de différenciation terminale. Le dernier axe ouvre des perspectives sur le rôle de peptides régulateurs pour le contrôle de différentes étapes du développement.

type de NGS	applications
reséquençage des génomes	découvertes de mutations dans le génome humain
séquençage "paired-end"	détection de variants de structure
séquençage méta-génomique	découverte flore commensale / infectieuse mise à jour de la diversité des virus
transcriptome	expression génique épissage alternatif annotations de transcrits SNPs
séquençage des petits ARNs	profil des microRNAs
séquençage de l'ADN traité au bisulfite	patron de méthylation des cytosines
ChIP-seq	interactions protéine-ADN
séquençage après fragmentation par nucléases	position des nucléosomes

**Figure 43. Diversité des NGS et applications.**

D'après Shendure, Ji 2008.

## A. vers un changement d'échelle de l'analyse des systèmes biologiques

### A.1 Opportunité et enjeu

L'introduction des techniques de séquençage de l'ADN (Sanger, Coulson 1977) a valu à F. Sanger l'attribution de son deuxième prix Nobel de chimie (en 1980, qu'il partage avec W. Gilbert). La publication en 1995 de la première séquence génomique complète d'un organisme vivant, celui de la bactérie *Haemophilus influenzae* (1830kb), allait ouvrir une nouvelle ère en biologie. Grâce à l'établissement de larges consortiums, la robotisation, et les progrès de la bioinformatique, l'année 2000 verra la détermination des premières séquences du génome des animaux, notamment la drosophile (Adams, Celniker et al. 2000), puis de celle de l'Homme. Ces résultats pionniers démontrent l'unicité du mode vivant, le relatif faible nombre de gènes, et finalement l'étendue de notre ignorance sur la façon dont notre génome code l'organisation et la diversité du vivant.

Une étape importante nécessite la comparaison de ces génomes pour mieux comprendre leur logique et dynamique évolutive. Le développement des méthodes de séquençage de nouvelle génération (NGS) permet aujourd'hui d'obtenir des séquences génomiques pour des sommes modiques (<10000\$), quand on les compare à l'effort considérable du premier génome humain (>3 milliards \$). On dispose du génome de centaines d'espèces vivantes et de larges programmes sont en cours, *e.g.* i5K, qui se propose de séquencer 5000 nouvelles espèces d'arthropodes d'ici 4 ans ([i5k-project](#)). Ces approches de génomique comparative sont aussi appliquées à l'analyse intraspécifique, ouvrant de nouvelles possibilités pour le génotypage massivement parallèle (identification de SNPs et de mutations). Grâce à cette mine d'information on peut aussi accéder à de nouvelles façons d'identifier l'ensemble de différents éléments fonctionnels du génome (pour revue (Shendure and Ji 2008) (Figure 43)). Par exemple le séquençage après fragmentation par des nucléases permet de localiser les nucléosomes (Schones, Cui et al. 2008), et surtout on peut identifier



les produits de l'expression du génome, les ARN messagers mais aussi des molécules régulatrices importantes comme les petits ARNs non-codants (Morin, O'Connor et al. 2008).

L'analyse générale et des niveaux d'ARN par puces à ADN, ou plus récemment RNA-seq, permet d'accéder au répertoire transcriptionnel d'un échantillon. Ces données transcriptomiques permettent d'identifier, outre le profil d'accumulation des ARNs, leur dynamique d'expression, maturation, traduction, etc...

Le séquençage de fragments génomiques fixés *in vivo* par une protéine (ChIP-seq) a rapidement remplacé le ChIP-on-Chip (hybridation des fragments précipités sur puces à ADN), et offre une vue des interactions régulatrices ADN/protéines à l'échelle du génome.

De nombreux outils bio-informatiques ont été développés afin d'analyser les résultats des approches de NGS. Par exemple, pour les résultats de transcriptome, CisTargetX permet d'analyser les régions non-codantes associées à une liste de gènes dérégulés, et d'identifier au sein de ces régions des clusters de motifs évolutivement conservés à partir d'une banque de PWMs (Aerts, Quan et al. 2010). Il est possible d'implémenter cette banque avec des matrices personnelles, ce qui nous a permis de comparer l'efficacité de prédiction entre la PWM OvoQ6 définie *in vitro* et SvbF7 définie sur un set de séquences validées *in vivo*. L'outil a ensuite évolué pour pouvoir également analyser des données de ChIP-seq.

Les logiciels MACS, QuEST, PICS, CisGenome et Findpeaks identifient les pics de ChIP-seq via un calcul du profil de densité des séquences. Une fois ces régions déterminées, on peut rechercher les motifs ou sites de fixation enrichis au sein de ces régions régulatrices potentielles. Ainsi, en fonction des données (gènes, séquences, motifs, ...) la plateforme Regulatory Sequence Analysis Tools (RSAT) propose une palette d'outils d'analyse des régions non-codantes (Thomas-Chollier, Sand et al. 2008). Peak-motifs combine différents algorithmes pour identifier des motifs de régulation au sein de séquences issues de NGS (ChIP-seq ou ChIP-on-chip), de comparer ces motifs à des banques de données pour



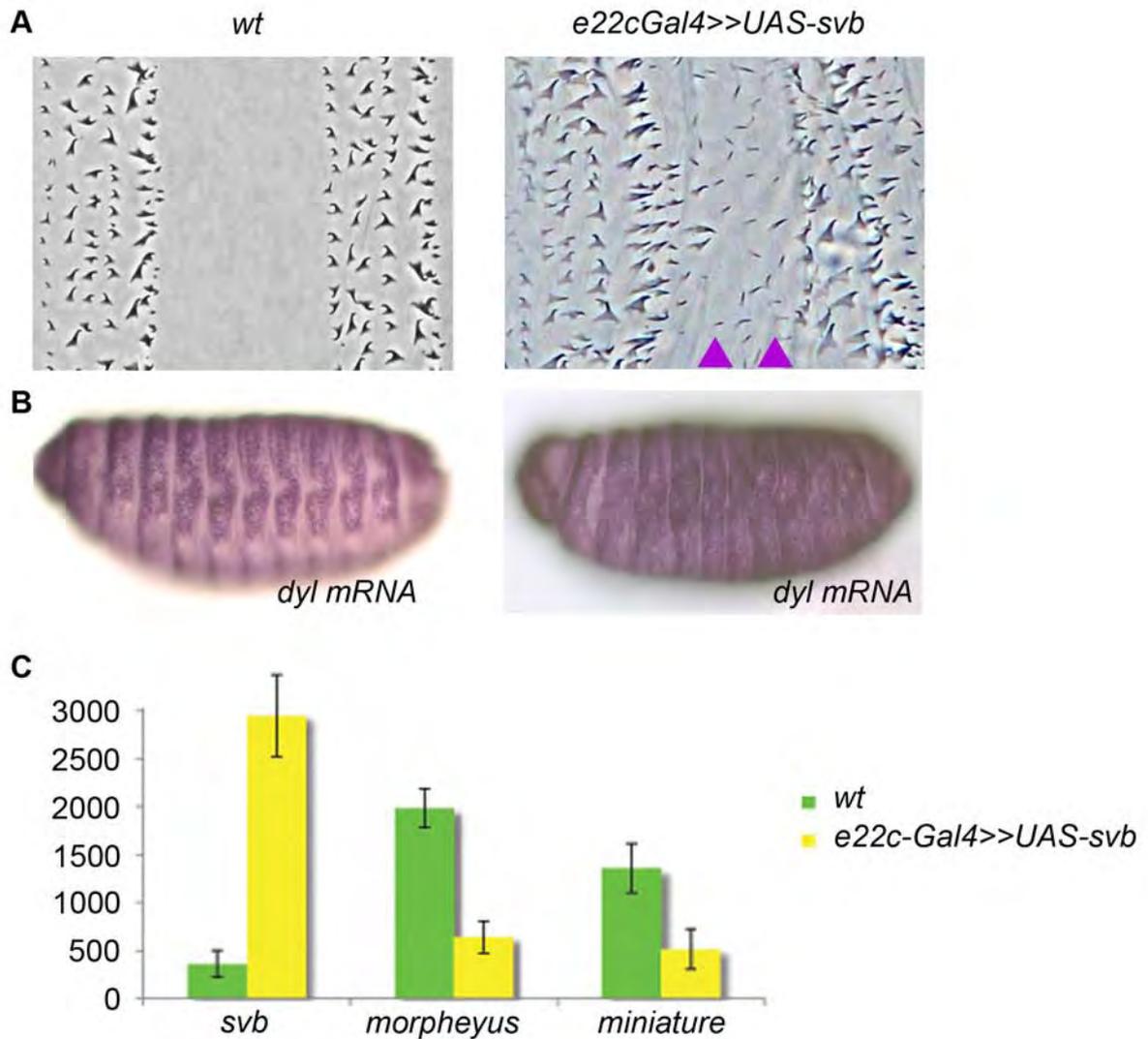
identifier les FTs fixés et de prédire des sites de fixation potentiels pour ces FTs. Pour nos analyses des ChIP-seq du FT Shavenbaby, l'outil de découverte de motif a confirmé que la matrice SvbF7 est enrichie dans les séquences ChIP-ées, alors que les motifs Blue et Yellow semblent plus dispersés sur l'ensemble du génome.

L'essor des approches de NGS entraînent cependant la génération d'une importante masse de données difficiles à relier entre elles. Des outils ont donc été développés pour aider à identifier les interactions et réseaux géniques, tels que GINsim qui permet de modéliser qualitativement un réseau, puis d'analyser et simuler des modifications engendrées au sein de ce réseau (Gonzalez, Naldi et al. 2006).

#### *A.2 apports et limites des approches transcriptome*

L'analyse du transcriptome génère des données quantitatives d'expression dans différentes conditions. Cette quantification du niveau d'expression de chacun des gènes requière évidemment un test de reproductibilité engageant la réalisation de réplicats. L'ACP effectuée sur nos données des variations du transcriptome engendrées par l'inactivation du facteur Svb, ou des peptides Pri, montre une bonne corrélation entre 5 échantillons indépendants d'un même génotype, validant la qualité « technique » de nos expériences. Une première difficulté importante réside cependant dans la réification d'un lien causal unique entre les dimensions macroscopique et microscopique de l'échantillon.

On spéculé en effet que le composant majeur des différences observées entre les échantillons sauvages et, par exemple mutants pour *svb*, est justement l'absence de ce seul gène. Il est cependant clair que la nature, le nombre et l'effet de multiples différences entre les génomes comparés, toutes choses égales par ailleurs, restent en l'état complètement inconnu. Chaque souche de laboratoire présente une origine, une histoire microévolutive différentes et est le réservoir d'une diversité génétique certainement importante (notamment dans le cas des souches mutantes qui sont difficiles à isogéniser). Autrement dit, il ne faut pas



**Figure 44. Une incohérence entre les données qualitatives et quantitatives.**

**A. L'expression ectopique de *svb* induit la formation de denticules dans les cellules lisses.**

L'expression ubiquitaire dans l'épiderme de *svb* (driver *e22c-Gal4*) est suffisante pour induire la formation de denticules ectopiques (triangles violets).

**B. *Svb* active ses effecteurs lorsqu'il est exprimé de façon ubiquitaire dans l'épiderme.**

Hybridation *in situ* de *dyl* après sur-expression de *svb* dans l'épiderme (driver *e22c-Gal4*). *dyl* est exprimé dans les cellules lisses.

**C. L'expression ubiquitaire de *svb* induit une baisse d'expression de ses effecteurs.**

Niveau d'expression moyen de *svb*, *morpheus* et *miniature* dans le transcriptome *wt* ou après expression ectopique de *svb*.

sous-estimer l'influence du contexte génétique sur l'expression et la fonction du génome. Cette difficulté est par exemple bien illustrée par les études génétiques du comportement, qui nécessitent un souci constant de réduire autant que faire se puisse les différences génomiques (par isogénisation) entre les souches de drosophile à comparer. Il est aussi bien connu que le phénotype d'une inactivation génétique chez la souris peut varier considérablement entre deux contextes génétiques différents (Didion, De vilena 2013).

Pour nos propres études, nous avons vu précédemment l'influence non négligeable du contexte génétique utilisé pour le transcriptome ou la validation *in vivo* (Cf partie 2). Alors que les gènes connus ne présentaient qu'une faible variation en absence de *svb*, masquée par des territoires d'expressions secondaires, la condition *pri* (équivalente à une condition Svb répresseur) nous a permis d'améliorer la liste de candidats malgré des extractions d'ARN sur embryons entiers. Dans le but d'obtenir une condition supplémentaire, nous avons également établi le transcriptome d'embryons sur-exprimant *svb* dans tout le tissu épidermique (via le driver *e22c-Gal4*). Alors qu'une hausse d'expression de *svb* a été validée qualitativement par hybridation *in situ*, et quantitativement par les données de puces (Fig 44), la majorité des gènes cibles connus de Svb ne présentent pas de modulation décelable, voire une baisse de leur niveau d'expression (Fig 44). Outre la différence des fonds génétiques, ces résultats surprenants pourraient être dus à un effet de « quenching » d'un éventuel cofacteur par la protéine Svb trop fortement exprimée. Il est aussi possible que cette manipulation génétique ait produit des défauts développementaux importants, altérant par exemple la morphologie et/ou la chronologie de l'embryogenèse polluant la collecte des données à un stade unique de développement.

L'analyse d'un transcriptome nous informe du niveau d'expression des gènes à un instant « T » du développement. Pour améliorer notre compréhension d'un processus biologique, il serait intéressant d'effectuer plutôt une analyse cinétique de la variation du transcriptome. Dans notre cas, les extractions d'ARN des embryons sauvages, *pri* et *svb* ont



été effectués à 13-15h (à 25°C) de l'embryogenèse, période à laquelle Svb active normalement ses effecteurs. La répétition des analyses avec cette fois une dizaine de points expérimentaux encadrant cette fenêtre temporelle serait certainement informative. Par-delà l'amélioration de l'identification des cibles, la comparaison des cinétiques de dérégulation génique produirait sans doute des informations utiles pour comprendre le comportement spécifique de sous classes fonctionnelles des effecteurs de la différenciation épidermique. Nos résultats sur la découverte des cibles moléculaires additionnelles des peptides Pri (*i.e.*, indépendant de Svb) soutiennent très fortement l'intérêt potentiel de cette approche de mesure cinétique des variations du transcriptome pour élargir l'étendue du répertoire des fonctions développementales de cette nouvelle classe de régulateurs.

On a vu aussi la difficulté d'extraire les variations d'expression génique pertinentes par la mesure du, ou en réalité des, transcriptomes de l'embryon entier. Il serait donc intéressant pour s'affranchir de ces limitations de trier spécifiquement les cellules de l'épiderme embryonnaire, idéalement des cellules à trichomes. On se heurte ici à un problème logique. Comment identifier les cellules à trichomes dans les mutants *svb*, ou *pri*, alors que leur détermination même dépend de ces facteurs. Une solution partiellement satisfaisante serait par exemple d'utiliser un BAC contenant l'ensemble des régions cis-régulatrice de *svb*, pour diriger l'expression d'une GFP membranaire (ou nucléaire, voir ci-dessous). Alternativement, les approches d'édition du génome comme la technique CRIPR-cas9 (Mali, Church 2013) pourraient permettre de remplacer *in situ* la séquence codante de *svb* par une GFP, un knocked-IN.

Une autre limitation de nos approches est qu'elles mesurent des niveaux d'accumulation des ARNs messagers, ce qui cumule différentes variables comme le taux de transcription bien sur, mais aussi la stabilité des ARN et donc de l'héritage des cellules à un stade précédent à l'analyse. Une amélioration possible serait donc de ne mesurer que la



formation des transcrits naissants, par une approche de type GRO-Seq (Danko, Kraus 2013), c'est à dire l'extension de l'analyse RUN-ON à l'échelle du génome entier. Pour l'embryon de drosophile, l'analyse des ARN nucléaires serait déjà sans doute un apport intéressant, surtout couplé au tri par des noyaux programmés pour exprimer *svb*. La faisabilité de cette approche, dans un contexte sauvage vient d'être démontrée par l'équipe de David STERN. Son implémentation pour comparer différents contextes génétiques reste à faire, mais le gain constant de sensibilité des RNA-Seq est un élément prometteur.

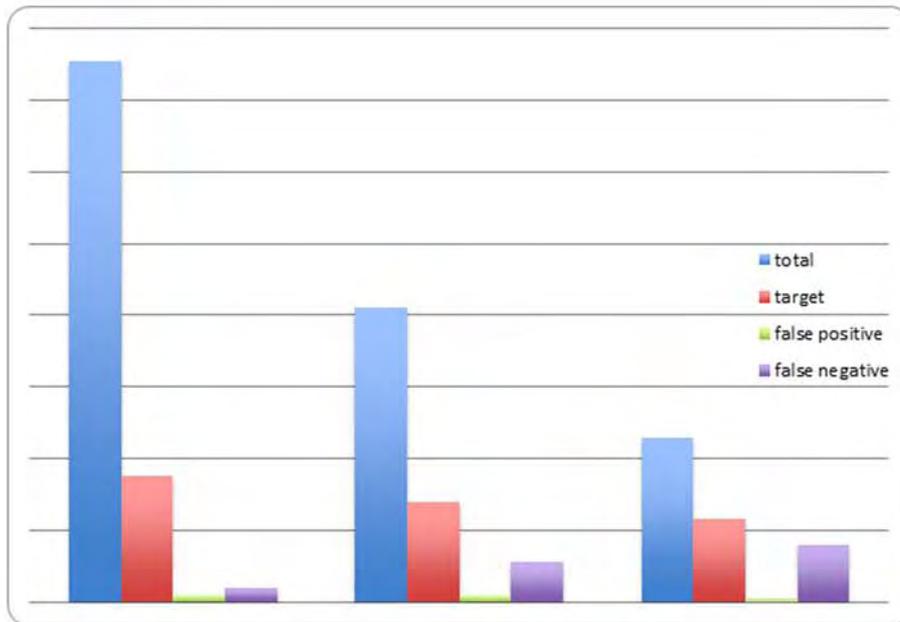
Cependant, la nature même des données génomiques engage une difficulté intrinsèque pour leur analyse. En effet, les outils statistiques habituels sont bien adaptés à l'analyse d'un grand nombre d'échantillons, décrits par un faible nombre de variables. Dans le cas du transcriptome on obtient une matrice pratiquement à l'opposé, c'est à dire quelques échantillons (ici 5 réplicats) ayant chacun plus de 18000 variables (le nombre de gènes, en réalité des sondes présents sur la puce). Un seuil d'erreur généralement acceptable de 1% implique ici un bruit obligatoire qui est du même ordre de grandeur que celui du nombre de gènes qu'on identifie généralement par ces approches. On doit donc garder ces chiffres en tête pour l'interprétation des données, notamment suites aux étapes de normalisation, filtrage (voir ci-dessous) et analyses *ad hoc*. Il faut se rappeler qu'un test statistique ramène en général à l'estimation d'une probabilité (celle de l'hypothèse nulle, *i.e.* pas de variation significative) et constitue une aide à la décision et non une preuve en tant que telle.

Une question importante pour le biologiste, et d'expérience difficile à assumer face au statisticien, est la nature et l'intérêt du filtrage initial des données à analyser. Ce point est bien illustré par les pipe-lines qui se sont avéré être les plus efficaces pour la découverte de nouveaux effecteurs de la différenciation épidermique. Tout d'abord, combien de gènes sont effectivement exprimés dans une condition donnée ? Dans les échantillons sauvages, 11905 sondes (sur les 18951) présentent un signal normalisé médian  $<50$  unités arbitraires (u.a) (en comparaison les valeurs pour *pri* et *svb* sont respectivement 2220 et 366 u.a.). Il semble donc



qu'une grande partie de ces signaux correspondent à du bruit biologique ou technique (8910 sondes  $< 10$  u.a.). De toute façon, il serait impossible de valider par les approches qualitatives que nous avons employées (hybridation *in situ*) une quelconque diminution de leur expression en condition mutante, leur détection en condition sauvage étant déjà largement en deçà de la sensibilité observée (approx.  $>100$  u.a.). Ainsi, nous avons filtré nos données pour ne rechercher les gènes down-régulés dans les conditions mutantes que parmi ceux dont l'expression sauvage est  $>100$ . Réciproquement pour la recherche des gènes up-régulés, ce filtre n'est pas adapté et nous avons choisi au contraire un seuil limite cette fois-ci pour les conditions mutantes (notamment *pri*), exactement pour les mêmes raisons de possibilité de validation par hybridation *in situ*. On voit donc qu'il ne s'agit en aucun cas d'une description exhaustive, mais que chaque liste générée représente un échantillon de gènes dont on pourra vérifier la dérégulation par une méthode informative et indépendante. On comprend bien sûr le côté hétérodoxe de comparer des listes établies avec des filtres différents !

Pire encore, je me rappelle les discussions animées concernant le choix des niveaux de fold-change et autres seuillages de valeurs p. Une fois écarté les gènes peu ou pas exprimés, il est vite apparu que la focalisation sur l'estimation de la « significativité » des variations ne permettrait pas d'identifier les acteurs de la morphogénèse des trichomes en aval du FT Svb. Par exemple, l'ensemble des gènes down-régulés ( $FC > -4$ ,  $p < 0,01$ ) dans les mutants *svb* se réduit à 25 gènes (dont 5 y compris *svb* sont simplement délétés par la déficience *svbR9*). Un seul d'entre eux correspond à un effecteur de la formation des trichomes, *shavenoid*, un autre *CG9460* n'est pas une cible de Svb, et 4 autres sont des acteurs du métabolisme ! Dans les mêmes conditions, on identifie 77 gènes down-régulés ( $FC > -4$ ,  $p < 0,01$ ) dans les embryons *pri* mutants, dont 17 cibles connues de Svb et seulement trois faux positifs (*CG14960*, *CG9196* et *Osi17*). Nous avons finalement choisi, et assumé, une démarche pragmatique pour optimiser la découverte des effecteurs de la formation des



**Figure 45. optimisation des paramètres de sélection pour les gènes candidats en aval de Svb par apprentissage sur le set de gènes cibles connus.**

Les candidats sont sélectionnés selon les critères suivants : i) expression sauvage > 100 u.a., ii)  $(\text{fold-change svb} / \text{fold-change pri}) > 2$ . On voit l'influence du seuil minimal de down-régulation mesuré dans les conditions mutant svb/sauvage sur le nombre : total de candidats (bleu), cibles connues (rouge), faux positifs (vert) et faux négatifs (violet).

Pour la suite de nos expériences, nous avons retenu la première combinaison, qui optimise le potentiel de découverte de nouvelles cibles.

trichomes en sacrifiant l'orthodoxie statistique. Comme présenté dans la partie 1, l'idée générale a été de se laisser guider par les connaissances déjà acquises sur la formation des trichomes, en cherchant les paramètres qui capturent le maximum d'effecteurs connus et les séparent le mieux des faux positifs. Après une première description des valeurs moyennes de chacun des paramètres pour l'ensemble des effecteurs connus, nous avons ajusté par approximation successive des seuils pour garder une liste de gènes aussi étendue que possible tout en maximisant la présence des cibles connues (*i.e.*, minimisant les faux négatifs) et minimisant les faux positifs (*i.e.*, les gènes non régulés par Svb, comme observé en *in situ*) (Fig 45).

L'analyse des résultats de transcriptome fut donc un challenge éprouvant pour obtenir des listes de candidats optimisées en fonction des questions biologiques. La connaissance de gènes cibles de Svb fut donc un atout majeur dans la définition des critères appliqués de gènes candidats pour être activé par Svb dans les cellules à trichomes. Nos résultats confirment l'intérêt de ce choix subjectif des seuils, paramètres et autres bidouillages de biologiste appliqués à nos données de puces pour augmenter l'efficacité de la recherche des effecteurs de la morphogenèse des trichomes.

Au contraire, la recherche des gènes up-régulés en absence de *pri* ne bénéficiait d'aucune donnée préalable susceptible de nous aider au benchmark des conditions. Nous avons alors choisi une approche conservative, se réfugiant derrière l'analyse orthodoxe des données de puces à ADN : à savoir la maximalisation du fold-change et la minimisation des valeurs p. Probablement grâce à l'ampleur des variations observées (FC moyen >280X pour les quatre gènes LCP1-4!), ces critères ont généré de nouvelles connaissances très informatives, notamment pour des gènes peu ou pas exprimé en conditions sauvages et fortement up-régulés en absence de *pri*. Il reste bien sur que des up-régulations plus



modérées, par exemple quantitatives et/ou dans un territoire additionnel pour des gènes déjà fortement exprimés à ce stade de l'embryogenèse, restent complètement à caractériser.

### *A.3 apports et limites des approches ChIP*

La réalisation de ChIP-seq fut une avancée importante pour identifier *in vivo* les régions cis-régulatrices des effecteurs cibles de Svb. L'interprétation de ces données représente elle aussi un véritable travail expérimental, avec ses inévitables erreurs de types I ou de type II. Les résultats de ChIP-seq génèrent un nombre d'occurrences de lecture pour chaque région du génome et il faut ensuite extraire les pics (par exemple en utilisant MACS) pour obtenir une liste de régions régulatrices potentielles. On retrouve des écueils similaires à ceux rencontrés pour l'analyse du transcriptome : d'abord de la « déformation » de la matrice échantillons/variables, les questions de reproductibilité des conditions expérimentales (nombre de réplicats, contrôles), la question de l'embryon entier *vs* une population cellulaire triée, ensuite des traitements pré- et post-analyse avec notamment le choix des méthodes de soustraction du bruit biologique, et bien sûr des seuils pour la reconnaissance des pics. On note cependant des avantages par rapport à nos puces, soit intrinsèques, soit liés au design expérimental. Par exemple, le NGS permet une extraordinaire finesse de l'analyse avec dans notre cas une valeur de reads tous les 40bp (soit >4 millions de variables), dont l'alignement sur le génome permet de séparer facilement le bruit aléatoire du signal, en fonction de la variation des régions voisines. D'autre part, la question de la variation du fond génétique devient presque caduque, le témoin réalisé dans chaque expérience correspondant aux mêmes extraits ChIP-és uniquement avec l'anticorps secondaire. Enfin, nous avons réalisé dans ce cas une cinétique de la variation du signal et analysant trois points temporels successifs (8-10h, 10-12h, 12-14h).

Il existe cependant des difficultés additionnelles au ChIP-Seq. La réalisation expérimentale reste plus complexe (même dans l'équipe de K. White !) et nous n'avons pas



réussi à obtenir un signal adéquat pour le point temporel intermédiaire. D'autre part, la qualité de l'anticorps est évidemment déterminante. Dans le cas de Svb, les nombreux anticorps qui ont été générés (à l'exception de l'anti-Svb1S, mais qui ne reconnaît que la forme répresseur de Svb) restent de qualité moyenne excluant leur utilisation en ChIP. Pour contourner ces difficultés, nous avons opté pour une construction transgénique chimère, exprimant une fusion Svb::GFP sous le contrôle de deux enhancers épidermiques principaux (McGregor et al, Nature 2007).

Nous identifions quand même près de 6000 pics pour le seul point temporel 12-14h que nous avons analysé plus en détail. Avant même toute validation expérimentale, il est probable que de nombreux pics ne correspondent pas à des régions régulatrices fonctionnelles. Plusieurs paramètres sont susceptibles de contribuer à cette « inflation » du nombre de pic ChIP. Différentes études notamment chez la drosophile ont rapporté l'existence d'un grand nombre ( $>10^3$ ) de pics ChIP, *e.g.* (Li, MacArthur et al. 2008), ce qui peut suggérer l'existence de sites de fixation réels mais non fonctionnels (voir la suite de la discussion). On a vu aussi le rôle présumé de régions HOTs permettant une fixation indirecte d'un TF. On ne peut cependant exclure dans nos analyses, une influence délétère du système utilisé, combinant surexpression (faible, probablement de l'ordre de 2X) et protéine Svb fusionnée à la GFP. Là encore, les développements méthodologiques comme l'édition du génome (pour insérer un tag immunologique au sein de la protéine Svb endogène sans altérer son activité, ni son patron spatial et temporel d'expression) et/ou le tri cellulaire, et un design expérimental avec plus de points temporels, offrent de nouvelles possibilités d'amélioration. Enfin, le développement de méthodes de ChIP-Seq beaucoup plus résolutive, comme le ChIP-exo (Rhee, Pugh 2012), semble ouvrir la voie à l'identification des régions liées virtuellement au nucléotide près.



#### A.4 apport et limites de la bioinformatique

De nombreux outils bio-informatiques ont été développés dans le but d'analyser les jeux de données des NGS et de prédire les motifs enrichis. Ces outils se basent évidemment sur la séquence du génome de l'organisme étudié, laissant peu de place à l'influence des éléments extrinsèques au génome, par exemple la conformation tridimensionnelle d'un locus. De plus, il arrive que des implémentations/mises à jour des séquences compliquent les analyses. Dans notre cas, la ré-annotation de certaines régions du génome de la drosophile a mené à la perte de la détection du site B du CRM *tyn2* par cis-targetX, alors qu'il s'agit du site fonctionnel *in vivo*.

Si de nombreuses améliorations ont été apportées aux outils de prédictions des CRMs, il existe un problème théorique qui limite leur potentiel de découverte. En effet les algorithmes ont été développés à partir des connaissances préexistantes sur l'architecture fonctionnelle des enhanceurs. Nous avons vu que pour le moment, les banques de données restent (c'est frappant dans le cas de la drosophile) extrêmement biaisées par les résultats d'analyses de l'embryogenèse précoce. Il est donc possible qu'un « street-light effect » affecte notre compréhension des CRMs, où finalement on ne favorise l'identification que des éléments qui correspondent aux modèles déjà établis, et en retour ces nouvelles données renforcent les hypothèses de départ quant à leur organisation fonctionnelle...

Comment les notions répandues de clustering homo- hétéro-typiques de sites de fixation, de conservation évolutive de ces motifs, d'existence de grammaire pour l'organisation respective de ces motifs etc...(bâties pour la plupart sur l'analyse des toutes premières heures de l'embryogenèse) pourront s'étendre à l'ensemble du développement (>10jours), voire à la vie adulte (>3 mois), reste à établir. Nos données, au moins dans le cas de la différenciation des trichomes, augurent que nous ne disposons aujourd'hui que d'une vision, disons partielle, de l'architecture fonctionnelle du génome régulateur. Des travaux



récents, sur d'autre processus de l'embryogenèse « tardive » (Junion, Spivakov et al. 2012; Jin, Stojnic et al. 2013) et des approches massivement parallèles (White & al, PNAS 2013) soutiennent déjà la portée de nos résultats.

De nouvelles analyses fonctionnelles si possible non biaisées, sont donc indispensables pour améliorer les outils de prédiction en enrichissant la diversité des modèles de conception et/ou paramétrage. On a vu aussi l'intérêt d'analyses statistiques pour découvrir *ab initio* de nouveaux éléments fonctionnels des enhanceurs, notamment quand le set d'apprentissage correspond à des éléments cis-régulateurs validés par analyse *in vivo*.

#### *A.5 Coupler des approches globales pour répondre à une question biologique*

Pour contourner les différents obstacles et limitations inhérents à chaque type d'approche, nous croyons à l'intérêt de coupler intimement ces démarches. Ainsi, pour mieux comprendre les mécanismes de la morphogénèse des cellules épidermiques, nous avons associé plusieurs approches globales : transcriptome, ChIP-seq, prédictions bio-informatiques et criblage génétique. La quantité d'informations issues de ces approches représente un réservoir impressionnant de données, mais il semble sage en l'état de les considérer comme un ensemble de prédictions plutôt que des résultats définitifs. L'ambition de l'exhaustivité peut aussi engendrer des problèmes, et une approche conservatrice est plutôt de définir des listes de candidats (gènes, CRM, motifs régulateurs) plausibles, à tout le moins testables. Les approches genome-wide sont intéressantes par leurs informations quantitatives. Peut-être par habitude, il nous a semblé essentiel de confronter/confirmer certaines de ces données par des approches indépendantes, qualitatives et *in vivo*. Nous avons ainsi choisi de tester les candidats gène cible par hybridation *in situ*, et les CRM présumés par transgénèse stable de constructions rapporteur d'activité transcriptionnelle, nonobstant la lourdeur de ce travail expérimental. De même, nous avons choisi (et j'ai regretté



parfois...) de tester là encore par transgénèse l'influence fonctionnelle des différents motifs découverts par analyse *in silico*, et celle de leur de leur profil de combinaison.

Il est essentiel de souligner qu'en plus de ce travail expérimental, mes interactions directes et multidisciplinaires avec des spécialistes de chaque approche ont été essentielles pour tirer le meilleur des jeux de données pour répondre aux différentes questions posées.

**Cette stratégie a permis : i) une avancée importante dans le décryptage des codes de régulation transcriptionnels de la différenciation terminale, ii) de mieux comprendre les mécanismes mis en jeu lors du remodelage de la forme des cellules, iii) de découvrir de nouvelles fonctions régulatrices des micro-protéines Pri pour le contrôle temporel du développement.**

## **B. les enhanceurs de la morphogenèse de l'épiderme : progrès et futurs challenges**

Les résultats de nos travaux basés sur une combinaison d'approches génome entier ont permis de réfuter l'existence d'une grammaire unique et contrainte pour l'architecture fonctionnelle des CRMs associés aux gènes cibles de Svb. Nous avons ainsi identifié trois nouveaux motifs dont diverses combinaisons sont nécessaires à l'activité des CRMs. Nous montrons également que l'association « au hasard » de sites multiples SvF7 et Blue Motif n'est pas suffisante pour diriger une expression épidermique, suggérant que d'autres éléments, par exemple notre Yellow Motif, sont nécessaires pour leur activité.

### *B.1 Présence et rôle potentiel des sites « fantômes » de fixation des TFs*

Notre étude révèle l'existence de nombreux motifs semblables au site de fixation de Svb dans l'environnement de ses gènes cibles, mais incapables de diriger une expression épidermique dans nos tests d'activité de rapporteurs transgéniques. Ces sites dont l'origine évolutive peut être discutée (Lusk and Eisen 2010) et que j'appellerais « sites fantômes »



sont à rapprocher du grand nombre de pics de ChIP qui ne sont pas fonctionnels (Li, McArthur & al, PLoS Biol 2008 ; Fisher & al, PNAS 2012). Ces sites fantômes pourraient servir de réservoir de FT dans l'environnement du CRM, afin de favoriser le recrutement du FT (Biggin 2010). Des sites fantômes de faible affinité pourraient augmenter la concentration locale des FTs notamment pour la boucle d'ADN entre CRM et promoteur.

### *B.2 Des modèles plus souples pour l'architecture fonctionnelle des CRMs*

On continue souvent à opposer les modèles de CRMs contraints (enhanceosome) et flexibles (billboard) (Yanez-Cuna, Kvon et al. 2012). Nos travaux contribuent à réfuter l'existence d'une grammaire unique entre différents enhanceurs qui dirigent un même programme de différenciation, dans les mêmes cellules, à une même étape de développement, sous le contrôle d'un même « master gene ». Des résultats similaires ont été récemment rapportés pour la différenciation cardiaque de drosophile (Junion, Spivakov et al. 2012; Jin, Stojnic et al. 2013), soutenant une organisation plus souple des CRMs au moins pour l'embryogenèse « tardive ». Cette vision pourrait cependant être élargie à d'autres CRMs. Smith et al (Smith, Taher et al. 2013) ont testé l'effet de modifications de l'organisation de CRMs synthétiques de vertébrés et montrent que l'arrangement des sites de fixation supporte une grande flexibilité pour diriger une expression spécifique.

### *B.3 Conservation évolutive des CRMs : contrainte ou mirage?*

Un paradoxe de notre étude est, malgré l'apparente flexibilité du code régulateur des différents CRMs validés chez *D. melanogaster*, la conservation évolutive de leur organisation respective à travers les différentes espèces de drosophile aujourd'hui séquencées. On a vu en introduction la confusion qui règne concernant l'interprétation des données de conservation fonctionnelle du CRM *eve-2* chez les insectes (Hare, Peterson et al. 2008) (Crocker & al, 2008). L'existence de cluster de sites de fixation ne prouve pas leur importance fonctionnelle



et pourrait résulter de conséquences indirectes des mécanismes de l'évolution des espèces (Lusk and Eisen 2010). En tout état de cause, nous croyons à l'existence de contraintes locales qui influent pour déterminer quel motif est effectivement utilisé *in vivo* pour la régulation de l'expression génique. Nos données sur les conséquences du contexte local des sites reconnus par Svb vont en effet dans ce sens. De même, une étude récente comparant 1300 régions liées par le FT Crx à 3000 régions contrôles soutient que le potentiel régulateur des interactions TF/ADN est déterminé par des caractéristiques « hautement locales » et non le contexte génomique (White, Myers & al 2013).

Collectivement ces données soutiennent la flexibilité d'organisation des CRMs pour diriger une activité tissu-spécifique, associée à une contrainte locale des sites de fixation.

## **C la découverte des effecteurs de la morphogenèse**

### *C.1 les limites de la prédiction fonctionnelle*

L'identification de nouveaux gènes cibles de Svb rendait attractive la découverte de leur fonction présumée pour mieux comprendre la formation des trichomes. Les analyses maintenant classiques de termes d'ontologie génique (GO) réalisées sur différentes plateformes (Bioconductor , DAVID, etc...) se sont toutefois révélées décevantes. Si les protéines de la transcription, des voies de signalisation, de certains aspects du métabolisme cellulaire général sont certainement faciles à identifier, il n'en va pas de même pour les nombreuses protéines cytosoliques ou membranaires qui contiennent des combinaisons variées de domaines fonctionnels. Comment prédire leur activité, la présence d'un domaine par exemple ATPase n'est pas très informatif pour deviner la fonction de la protéine correspondante, qui peut agir dans une foultitude de mécanismes cellulaires différents, de la régulation de la chromatine au fonctionnement du protéasome. Peut-on attribuer une fonction à chaque protéine ? On découvre chaque jour des implications inattendues de tel ou



tel facteur suite à une nouvelle analyse fonctionnelle. On réalise aussi les limites de l'annotation automatique quand on évalue les GO terms associés à une de nos protéines favorites. Enfin, on est frappé par l'ampleur du nombre de gènes dont on ne connaît ni fonction biologique, ni fonction moléculaire présumée, ni même la présence de domaine protéique associés à d'autres facteurs mieux connus. Là encore, par delà les « intelligent guesses », on se rend compte de l'étendue de notre ignorance sur les fonctions géniques et on n'échappera pas à la généralisation de l'analyse fonctionnelle, grâce notamment aux organismes modèles.

### *C.2 qu'avons-nous appris de l'analyse génome-wide*

La recherche des effecteurs de la morphogénèse épidermique nous a permis d'identifier une centaine gènes cibles de Svb pour ce processus. Nous avons montré que Svb régule directement un grand nombre de gènes appartenant à différentes familles : liaison à l'actine, matrice extracellulaire, cuticule, sécrétion, PCP, etc.... Nous détaillons dans la suite la portée de l'identification des acteurs de la voie PCP, en particulier pour la morphogénèse des cellules à trichomes, et pour notre compréhension des mécanismes de polarisation cellulaire en général. Nous avons également identifié de nombreux effecteurs aux fonctions encore inconnues à ce jour, soulignant la surprenante diversité des mécanismes nécessaire au remodelage de l'organisation tridimensionnelle des cellules.

Le transcriptome *pri*, utilisé au départ comme condition complémentaire pour l'identification des cibles de Svb dans la morphogénèse épidermique, représente finalement un réservoir impressionnant de données pour découvrir le rôle de micro-protéines régulatrices lors du développement.

### *C.3 PCP : vive la différence !*

#### C.3.1 enjeu de mieux comprendre la PCP



La polarisation apico-basale des cellules est une caractéristique majeure des tissus épithéliaux qui est indispensable pour leurs fonctions biologiques. De nombreux progrès ont été réalisés pour comprendre les mécanismes et acteurs moléculaires participant à l'acquisition de la polarité apico-basale des cellules. En comparaison, les mécanismes d'acquisition d'une polarité tissulaire, orthogonale à cet axe, dite de polarité planaire ou PCP restent encore mal compris. La découverte de la PCP doit beaucoup à la drosophile et la plupart des molécules impliquées ont été identifiées notamment de par leur influence sur l'organisation planaire des trichomes adultes. On réalise cependant aujourd'hui l'importance de la PCP pour le développement animal, et son influence sur le fonctionnement adulte du système rénal, de l'audition ou du cerveau. Nous allons voir comment nos résultats sur le modeste épiderme embryonnaire de la drosophile peuvent apporter de nouvelles directions pour mieux comprendre l'établissement et l'exécution de la PCP.

### C.3.2 avancées principales

La PCP est supposée mettre en place des coordonnées de polarité à travers l'ensemble d'un tissu, ceci via les protéines du module GLOBAL (Adler 2012; Peng and Axelrod 2012). Les phénotypes observés suite à l'inactivation de *ds* dans l'embryon montrent un rôle local, (i.e., rangée-spécifique) de cet acteur pour la PCP des trichomes. Ces résultats étendent des données obtenues dans l'analyse de larves de troisième stade (Casal, Lawrence et al. 2006; Repiso, Saavedra et al. 2010; Donoughe and DiNardo 2011). Nous montrons par ailleurs un patron d'expression de ce gène spécifiquement dans les cellules à trichomes. Collectivement ces résultats montrent qu'au moins dans l'épiderme, il existe un processus de PCP qui se met en place à travers quelques rangées de cellules, et non sur l'ensemble du tissu.

Des études ont montré l'importance de la voie PCP pour le processus de différenciation des kératinocytes ou des cellules pancréatiques chez les vertébrés (Cortijo, Gouzi et al. 2012; Dai, Li et al. 2013). L'identification d'effecteurs tissu-spécifique de la PCP



comme gènes cibles de Svb dans l'embryon met en lumière un rôle nouveau du processus de différenciation important pour participer à la PCP du tissu. Un dialogue croisé entre PCP et différenciation semble donc être important pour que les deux événements puissent avoir lieu.

Rab23 est un nouvel effecteur de Svb, et fait partie de la famille des petites GTPases impliquées dans le transport vésiculaire. Son rôle connu actuellement est de participer à la PCP dans l'aile (Pataki, Matusek et al. 2010), mais son mécanisme d'action reste inconnu. Des données récentes montrent le rôle d'autres petites GTPases dans la PCP. Rab10 participe à la polarisation de la sécrétion des composants de la membrane basale lors de la morphogénèse de la chambre ovarienne (Lerner, McCoy et al. 2013), et Rab8 est nécessaire à la localisation subcellulaire particulière de Dishevelled lors de la ciliogénèse (Zilber, Babayeva et al. 2013). Collectivement ces résultats montrent une participation des GTPases Rab dans la polarisation d'une cellule, ceci via un transport/séquestration de protéines. Des liens directs ont été montrés entre certaines Rab et les composants de la vésicule COPII comme sec23 et sec31 (Slavin, Garcia et al. 2011), suggérant que PCP et sécrétion COPII sont intimement liées, et que Rab23 pourrait être l'un des acteurs permettant l'interaction entre les 2 mécanismes dans l'épiderme embryonnaire.

### C.3.3 futures directions

Alors que les acteurs principaux de la PCP (module GLOBAL et CORE) sont bien connus dans différents modèles, il apparaît que le rôle respectif de chaque module varie d'un modèle à l'autre. Nous montrons des nouveautés concernant l'implication d'acteurs de ces modules dans la PCP embryonnaire, preuve que des efforts sont à faire pour mieux comprendre le rôle de ces modules selon les tissus. De nombreux autres acteurs spécifiques de chaque type cellulaire participent également à l'acquisition de cette polarité. Le répertoire de ces acteurs spécifiques s'élargit de jour en jour, complexifiant davantage notre compréhension des mécanismes mis en jeu. Les enjeux actuels consistent donc à établir



précisément leur rôle dans l'épiderme embryonnaire. De nouvelles voies semblent également influencer la PCP telle que la sécrétion, il est donc important de définir les « nœuds d'interactions » entre la PCP et ces voies pour mieux comprendre comment se met en place une extension de façon asymétrique sur la face apicale d'une cellule.

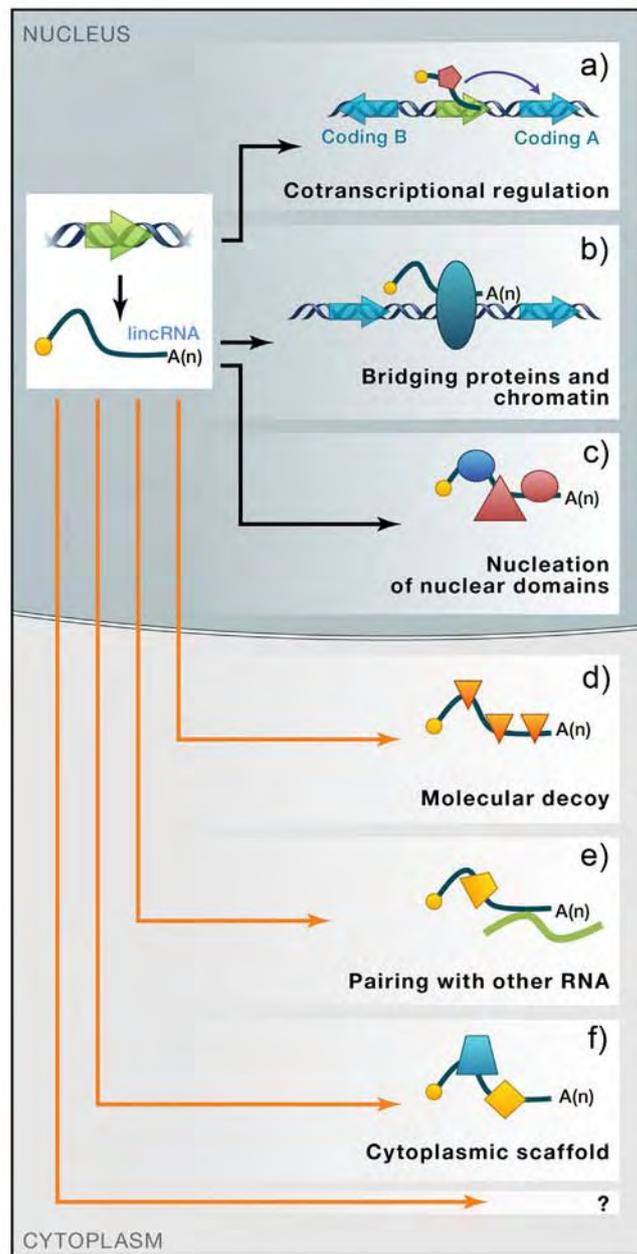
## **D. une nouvelle classe de régulateurs du développement : les petits peptides**

### *D.1 l'importance quantitative et fonctionnelle des lncRNAs*

Les longs ARNs non codants (lncRNAs) font partie de la grande famille des ARNs dont la fonction semble ne pas dépendre de la production d'un intermédiaire protéique. Différentes études, notamment utilisant le NGS, suggèrent qu'une fraction considérable du génome pourrait produire des lncRNAs. Certains auteurs proposent même que 80% du génome humain transcrit de longs ARNs non-codants (Kapranov, Cheng et al. 2007). On comprend cependant encore très mal le registre fonctionnel des lncRNAs. Par exemple, la banque de données LncRNADB, qui ambitionne une couverture de l'ensemble des lncRNAs eucaryotes, ne répertorie qu'une centaine de lncRNAs pour lesquels on dispose de données de leur activité. Malgré la rareté de ces données, l'importance des lncRNAs est maintenant bien établie pour la régulation de l'expression du génome à différents niveaux (pour revue (Guttman and Rinn 2012), (Ulitsky and Bartel 2013)) (Fig 46) : i) transcriptionnels (Feng, Bi et al. 2006), ii) post-transcriptionnels (Beltran, Puig et al. 2008), iii) épigénétiques (Sanchez-Elsner, Gou et al. 2006). Des lncRNAs semblent également impliqués dans des pathologies neurologiques (Lukiw, Handley et al. 1992) et de nombreux cancers. On voit donc l'enjeu de contribuer à élucider le spectre d'action de cette classe de molécules récemment découvertes.

### *D.2 you said long NON-CODING RNAs ?*

Les algorithmes d'annotation des ARNs sont généralement calibrés pour définir un ARN comme non codant s'il ne contient pas d'ORF supérieure à une centaine de codons. La



**Figure 46. Différentes fonctions des lincRNAs.**

Un lincRNA peut participer à la régulation transcriptionnelle via un recrutement de facteurs au niveau du transcrit naissant du lincRNA (a), ou un recrutement de cofacteurs transcriptionnels au niveau d'un locus (b). Il peut également servir de leurre pour piéger des facteurs régulateurs d'ARNs (d), s'associer avec un ARN pour une régulation post-transcriptionnelle, ou participer au maintien de la structure de complexes multiprotéiques (c, f).

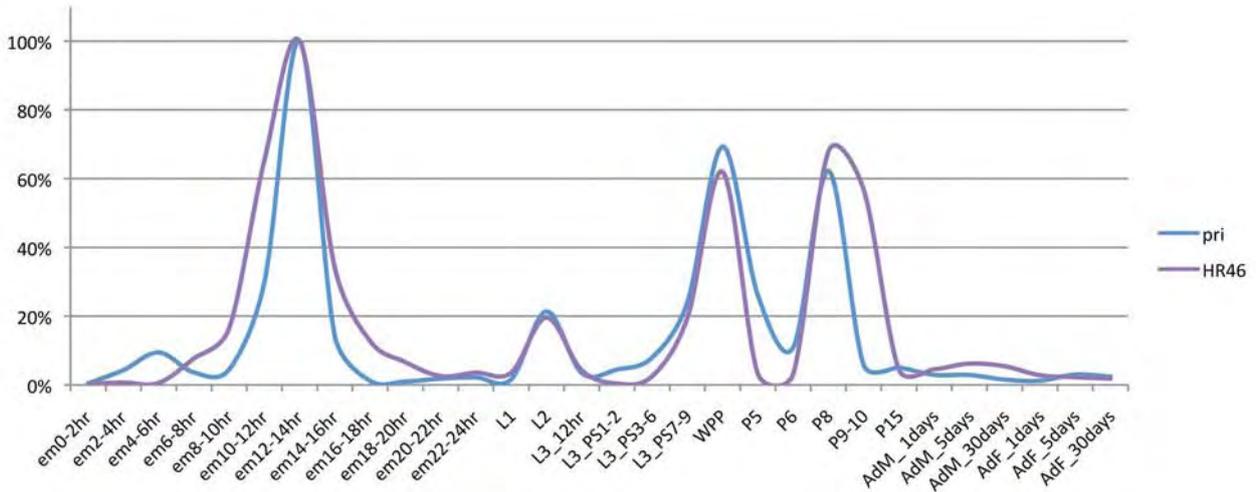
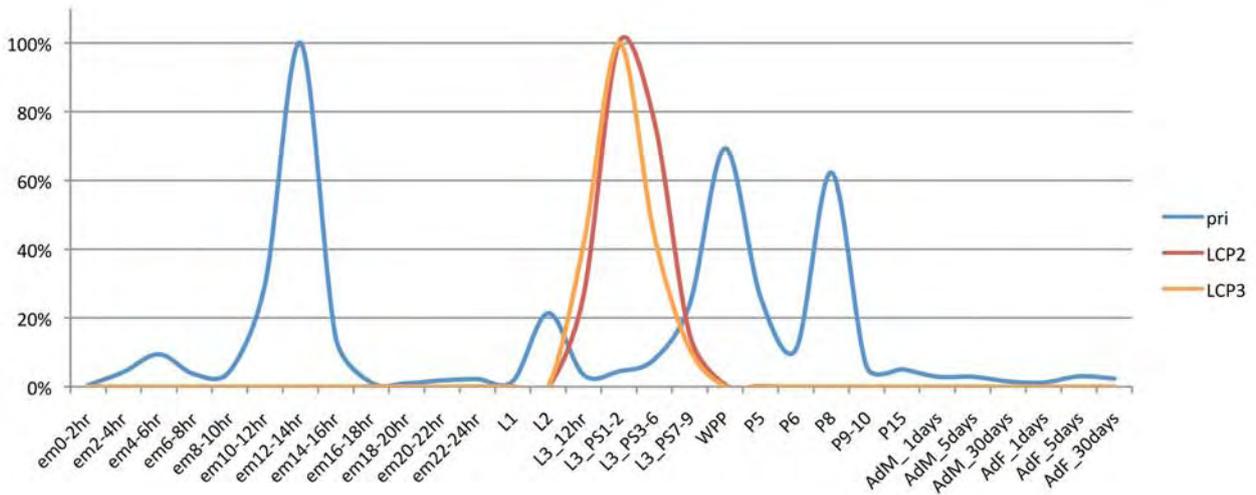
D'après (Ulitsky and Bartel, 2013)

découverte de *pri*, initialement défini comme lncRNA, mais dont la fonction passe par la production de 4 peptides ou micro-protéines (11-32aa), remet en question l'intérêt de ce seuillage, suggérant que de nombreux ARNs réputés non codants pourraient, au moins en partie, exercer aussi leur activité par la production de tels peptides.

Magny et al (Magny, Pueyo et al. 2013) ont identifié un autre ARN apparemment non codant, mais qui produit en réalité deux peptides de 30aa participant à la régulation du transport du calcium lors de la contraction cardiaque chez la drosophile. Une petite ORF codant pour un peptide de 135 acides aminés a également été identifiée dans le génome de *l'herpesvirus bovin-1 (BHV-1)* et participerait au cycle de latence du virus (Inman, Zhou et al. 2004). Une analyse exhaustive des petites ORFs chez *Arabidopsis thaliana* leur attribue un rôle sans doute très important pour la morphogénèse (Hanada, Higuchi-Takeuchi et al. 2013). Une extrapolation suggère que plusieurs centaines de petites ORFs seraient présentes dans le génome de drosophile, certaines présentant une conservation évolutive forte jusqu'à l'Homme (Magny, Pueyo et al. 2013). Il apparaît donc des peptides codés par de petites ORFs seraient porteurs de fonctions régulatrices importantes pour le développement et la différenciation des organismes (Cheng, Chan et al. 2011) (Cabili, Trapnell et al. 2011).

### *D.3 De nouvelles fonctions régulatrices des peptides Pri*

L'ARN *pri* contient donc 4 ORFs, codant des micro-protéines de 11 à 32 acides aminés douées de fonctions régulatrices importantes pour le développement des insectes, par exemple pour la segmentation embryonnaire (Savard, Marques-Souza et al. 2006), la morphogénèse des pattes adultes (Galindo, Pueyo et al. 2007), et la différenciation de l'épiderme (Galindo, Pueyo et al. 2007; Kondo, Hashimoto et al. 2007; Kondo, Plaza et al. 2010). L'analyse du transcriptome des embryons mutants nous a permis de découvrir des fonctions additionnelles des micro-protéines *pri* dans le développement de la drosophile, et d'identifier des centaines de nouvelles cibles de leur action. En absence de Pri, de nombreux

**A****B**

**Fig 47. Les cinétiques d'expression de *pri* et des LCPs corrént avec l'activité de la voie ecdysone.**

**A. Expression de *pri* et *HR46*.**

Données de RNA-seq (Flybase) montrant les niveaux d'expression de *pri* et *HR46* (readout de la voie ecdysone), à différentes étapes du développement. L'expression de *pri* corréle parfaitement avec les pics d'ecdysone observés en milieu d'embryogenèse, et aux transitions larvaires et pupales.

**B. Expression de *pri* et des LCPs.**

Données de RNA-seq (Flybase) montrant les niveaux d'expression de *pri* et des *LCPs* à différentes étapes du développement. L'expression des *LCPs* est restreinte aux stades larvaires tardifs (L3).

gènes sont en effet transcriptionnellement dérégulés, indépendamment de l'activité du FT Svb. Une hypothèse simple serait que les peptides Pri régulent aussi l'activité d'un ou plusieurs FTs différents de Svb dans d'autres tissus.

De nombreux composants de la cuticule sont exprimés prématurément en absence de *pri*. Nous montrons par exemple, une forte expression épidermique de ces gènes dès le stade 16 de l'embryogenèse alors que leur expression est normalement restreinte à des étapes plus tardives du développement embryonnaire (18h pour Edg91 et 20h pour les Cprs), post-embryonnaire (*i.e.*, stade L3 pour les LCPs) et finalement au cours de la métamorphose ou même le développement des jeunes adultes, suggérant un rôle des micro-protéines dans le contrôle temporel des différentes étapes développementales. Cette hypothèse avait déjà été proposée pour la morphogénèse épidermique gouvernée par Svb (Kondo, Plaza et al. 2010). L'expression de chacun des gènes codant pour un composant de la cuticule corrèle avec un pic d'ecdysone. Une hypothèse attrayante serait un lien entre les peptides Pri et cette hormone connue comme régulateur temporel des mues de la larve de drosophile. Des données récentes de l'équipe montrent que l'expression de *pri* dépend en effet de la voie de l'ecdysone (Jérôme Bohere). On peut ainsi imaginer les peptides Pri comme régulateurs temporels du développement, contribuant à synchroniser les différents programmes de différenciation les uns par rapport aux autres, en lien étroit avec le développement global de l'organisme et probablement aussi en interaction avec le milieu extérieur.

## E. Conclusion

La mise en place et l'analyse des résultats des approches génome entier ne sont pas exemptes de difficultés. Elles font appel à des démarches multidisciplinaires, des mathématiques à la physique, en passant par l'informatique et la biologie expérimentale. Il est difficile aujourd'hui de cumuler ces compétences pour une seule personne. Le biologiste est souvent noyé dans cette avalanche de données quantitatives et n'a pas forcément les



réflexes pour analyser ces données dans de bonnes conditions. Le statisticien se méfie de nos tripatouillages des données et préfère des approches plus systématiques basées sur l'analyse des variations et de la probabilité de leur signification. Le temps de la modélisation semble aussi plus court que celui de l'analyse *in vivo*. Le spécialiste des sciences dures est souvent frustré de la relative lenteur du test biologique de ses nombreuses hypothèses.

Il faut aussi faire le deuil du mythe d'une analyse moléculaire exhaustive, comme celui d'un criblage génétique à saturation. Même si nous ne sommes pas encore à une analyse véritablement globale, nous démontrons qu'on peut découvrir et étudier la fonction *in vivo* de centaines d'interactions moléculaires, avec un niveau de détail sans précédent.

Nous avons ainsi découvert qu'un processus morphologique apparemment simple dépend de l'interconnexion d'une variété de mécanismes moléculaires et fonctions cellulaires dont l'enjeu de la compréhension dépasse la drosophile et son épiderme embryonnaire.

La découverte des micro-protéines montre que malgré le décryptage de milliers de génomes, des régulateurs clés du développement et de la physiologie sont restés ignorés à cause de paramètres arbitraires, cachés dans une valeur de variable des algorithmes d'annotation. A lui seul, cet exemple illustre la nécessité critique d'un va et vient constant entre les recherches « wet » et « dry ».

Sans oublier les passages difficiles, cette approche interdisciplinaire s'est finalement imposée comme une évidence pour identifier les éléments et reconstruire les interactions d'un système biologique complexe via l'association des compétences scientifiques de chacun.



## REFERENCES

- Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of *Drosophila melanogaster*." *Science* **287**(5461): 2185-2195.
- Adler, P. N. (2012). "The frizzled/stan pathway and planar cell polarity in the *Drosophila* wing." *Curr Top Dev Biol* **101**: 1-31.
- Aerts, S., J. van Helden, et al. (2007). "Fine-tuning enhancer models to predict transcriptional targets across multiple genomes." *PLoS One* **2**(11): e1115.
- Aerts, S., P. Van Loo, et al. (2003). "Computational detection of cis-regulatory modules." *Bioinformatics* **19 Suppl 2**: ii5-14.
- Agawa, Y., M. Sarhan, et al. (2007). "*Drosophila* Blimp-1 is a transient transcriptional repressor that controls timing of the ecdysone-induced developmental pathway." *Mol Cell Biol* **27**(24): 8739-8747.
- Andrew, D. J. and B. S. Baker (2008). "Expression of the *Drosophila* secreted cuticle protein 73 (*dsc73*) requires Shavenbaby." *Dev Dyn* **237**(4): 1198-1206.
- Andrews, J., D. Garcia-Estefania, et al. (2000). "OVO transcription factors function antagonistically in the *Drosophila* female germline." *Development* **127**(4): 881-892.
- Arbouzova, N. I., E. A. Bach, et al. (2006). "Ken & barbie selectively regulates the expression of a subset of Jak/STAT pathway target genes." *Curr Biol* **16**(1): 80-88.
- Arnosti, D. N., S. Barolo, et al. (1996). "The eve stripe 2 enhancer employs multiple modes of transcriptional synergy." *Development* **122**(1): 205-214.
- Ayer, S. and C. Benyajati (1990). "Conserved enhancer and silencer elements responsible for differential *Adh* transcription in *Drosophila* cell lines." *Mol Cell Biol* **10**(7): 3512-3523.
- Beltran, M., I. Puig, et al. (2008). "A natural antisense transcript regulates *Zeb2/Sip1* gene expression during Snail1-induced epithelial-mesenchymal transition." *Genes Dev* **22**(6): 756-769.
- Benoist, C. and P. Chambon (1981). "In vivo sequence requirements of the SV40 early promoter region." *Nature* **290**(5804): 304-310.
- Berger, M. F. and M. L. Bulyk (2006). "Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins." *Methods Mol Biol* **338**: 245-260.
- Bergman, C. M., B. D. Pfeiffer, et al. (2002). "Assessing the impact of comparative genomic sequence data on the functional annotation of the *Drosophila* genome." *Genome Biol* **3**(12): RESEARCH0086.
- Bergson, C. and W. McGinnis (1990). "An autoregulatory enhancer element of the *Drosophila* homeotic gene *Deformed*." *Embo J* **9**(13): 4287-4297.
- Berman, B. P., Y. Nibu, et al. (2002). "Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome." *Proc Natl Acad Sci U S A* **99**(2): 757-762.
- Berman, B. P., B. D. Pfeiffer, et al. (2004). "Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*." *Genome Biol* **5**(9): R61.
- Bielinska, B., J. Lu, et al. (2005). "Core promoter sequences contribute to ovo-B regulation in the *Drosophila melanogaster* germline." *Genetics* **169**(1): 161-172.
- Bischof, J., R. K. Maeda, et al. (2007). "An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases." *Proc Natl Acad Sci U S A* **104**(9): 3312-3317.
- Blair, A., A. Tomlinson, et al. (2006). "Twinstar, the *Drosophila* homolog of cofilin/ADF, is required for planar cell polarity patterning." *Development* **133**(9): 1789-1797.

- Borggrefe, T. and X. Yue (2011). "Interactions between subunits of the Mediator complex with gene-specific transcription factors." *Semin Cell Dev Biol* **22**(7): 759-768.
- Brehme, K. S. (1941). "The Effect of Adult Body Color Mutations upon the Larva of *Drosophila Melanogaster*." *Proc Natl Acad Sci U S A* **27**(6): 254-261.
- Britten, R. J. and E. H. Davidson (1969). "Gene regulation for higher cells: a theory." *Science* **165**(3891): 349-357.
- Brittle, A., C. Thomas, et al. (2012). "Planar polarity specification through asymmetric subcellular localization of Fat and Dachshous." *Curr Biol* **22**(10): 907-914.
- Brown, C. D., D. S. Johnson, et al. (2007). "Functional architecture and evolution of transcriptional elements that drive gene coexpression." *Science* **317**(5844): 1557-1560.
- Cabili, M. N., C. Trapnell, et al. (2011). "Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses." *Genes Dev* **25**(18): 1915-1927.
- Cande, J., Y. Goltsev, et al. (2009). "Conservation of enhancer location in divergent insects." *Proc Natl Acad Sci U S A* **106**(34): 14414-14419.
- Carr, C. M. and M. Munson (2007). "Tag team action at the synapse." *EMBO Rep* **8**(9): 834-838.
- Casal, J., P. A. Lawrence, et al. (2006). "Two separate molecular systems, Dachshous/Fat and Starry night/Frizzled, act independently to confer planar cell polarity." *Development* **133**(22): 4561-4572.
- Chanut-Delalande, H., I. Fernandes, et al. (2006). "Shavenbaby couples patterning to epidermal cell shape control." *PLoS Biol* **4**(9): e290.
- Chavoshi, T. M., B. Moussian, et al. (2010). "Tissue-autonomous EcR functions are required for concurrent organ morphogenesis in the *Drosophila* embryo." *Mech Dev* **127**(5-6): 308-319.
- Cheng, H., W. S. Chan, et al. (2011). "Small open reading frames: current prediction techniques and future prospect." *Curr Protein Pept Sci* **12**(6): 503-507.
- Cortijo, C., M. Gouzi, et al. (2012). "Planar cell polarity controls pancreatic beta cell differentiation and glucose homeostasis." *Cell Rep* **2**(6): 1593-1606.
- Crocker, J. and A. Erives (2008). "A closer look at the eve stripe 2 enhancers of *Drosophila* and *Themira*." *PLoS Genet* **4**(11): e1000276.
- Cunha, P. M., T. Sandmann, et al. (2010). "Combinatorial binding leads to diverse regulatory responses: Lmd is a tissue-specific modulator of Mef2 activity." *PLoS Genet* **6**(7): e1001014.
- Dai, D., L. Li, et al. (2013). "Planar cell polarity effector gene Intu regulates cell fate-specific differentiation of keratinocytes through the primary cilia." *Cell Death Differ* **20**(1): 130-138.
- Davidson, B. P., S. C. Dogra, et al. (2001). "A duplicated HNF-3 binding site in the CYP2H2 promoter underlies the weak phenobarbital induction response." *Int J Biochem Cell Biol* **33**(11): 1080-1093.
- Delon, I., H. Chanut-Delalande, et al. (2003). "The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in *Drosophila*." *Mech Dev* **120**(7): 747-758.
- Dickinson, W. J. and J. W. Thatcher (1997). "Morphogenesis of denticles and hairs in *Drosophila* embryos: involvement of actin-associated proteins that also affect adult structures." *Cell Motil Cytoskeleton* **38**(1): 9-21.
- Donoughe, S. and S. DiNardo (2011). "dachshous and frizzled contribute separately to planar polarity in the *Drosophila* ventral epidermis." *Development* **138**(13): 2751-2759.
- Doyle, H. J., R. Kraut, et al. (1989). "Spatial regulation of zerknüllt: a dorsal-ventral patterning gene in *Drosophila*." *Genes Dev* **3**(10): 1518-1533.

- Eaton, S. (1997). "Planar polarization of *Drosophila* and vertebrate epithelia." *Curr Opin Cell Biol* **9**(6): 860-866.
- Elemento, O. and S. Tavazoie (2005). "Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach." *Genome Biol* **6**(2): R18.
- Erives, A. and M. Levine (2004). "Coordinate enhancers share common organizational features in the *Drosophila* genome." *Proc Natl Acad Sci U S A* **101**(11): 3851-3856.
- Ezan, J. and M. Montcouquiol (2013). "Revisiting planar cell polarity in the inner ear." *Semin Cell Dev Biol* **24**(5): 499-506.
- Farley, E. and M. Levine (2012). "HOT DNAs: a novel class of developmental enhancers." *Genes Dev* **26**(9): 873-876.
- Feng, J., C. Bi, et al. (2006). "The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator." *Genes Dev* **20**(11): 1470-1484.
- Ferguson, L. C., J. Green, et al. (2011). "Evolution of the insect yellow gene family." *Mol Biol Evol* **28**(1): 257-272.
- Fernandes, I., H. Chanut-Delalande, et al. (2010). "Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes." *Dev Cell* **18**(1): 64-76.
- Fisher, W. W., J. J. Li, et al. (2012). "DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*." *Proc Natl Acad Sci U S A* **109**(52): 21330-21335.
- Frankel, N., G. K. Davis, et al. (2010). "Phenotypic robustness conferred by apparently redundant transcriptional enhancers." *Nature* **466**(7305): 490-493.
- Frankel, N., D. F. Erezylmaz, et al. (2011). "Morphological evolution caused by many subtle-effect substitutions in regulatory DNA." *Nature* **474**(7353): 598-603.
- Frasch, M. and M. Levine (1987). "Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in *Drosophila*." *Genes Dev* **1**(9): 981-995.
- Galindo, M. I., J. I. Pueyo, et al. (2007). "Peptides encoded by short ORFs control development and define a new eukaryotic gene family." *PLoS Biol* **5**(5): e106.
- Garfinkel, M. D., J. Wang, et al. (1994). "Multiple products from the shavenbaby-ovo gene region of *Drosophila melanogaster*: relationship to genetic complexity." *Mol Cell Biol* **14**(10): 6809-6818.
- Glaser, R. L., M. F. Wolfner, et al. (1986). "Spatial and temporal pattern of hsp26 expression during normal development." *Embo J* **5**(4): 747-754.
- Gonzalez, A. G., A. Naldi, et al. (2006). "GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks." *Biosystems* **84**(2): 91-100.
- Goto, T., P. Macdonald, et al. (1989). "Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues." *Cell* **57**(3): 413-422.
- Guan, X., B. W. Middlebrooks, et al. (2006). "Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in *Drosophila*." *Proc Natl Acad Sci U S A* **103**(45): 16794-16799.
- Guruharsha, K. G., J. F. Rual, et al. (2011). "A protein complex network of *Drosophila melanogaster*." *Cell* **147**(3): 690-703.
- Guttman, M. and J. L. Rinn (2012). "Modular regulatory principles of large non-coding RNAs." *Nature* **482**(7385): 339-346.
- Halder, G., P. Callaerts, et al. (1995). "Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*." *Science* **267**(5205): 1788-1792.

- Hanada, K., M. Higuchi-Takeuchi, et al. (2013). "Small open reading frames associated with morphogenesis are hidden in plant genomes." *Proc Natl Acad Sci U S A* **110**(6): 2395-2400.
- Hare, E. E., B. K. Peterson, et al. (2008). "A careful look at binding site reorganization in the even-skipped enhancers of *Drosophila* and sepsids." *PLoS Genet* **4**(11): e1000268.
- Hertz, G. Z. and G. D. Stormo (1999). "Identifying DNA and protein patterns with statistically significant alignments of multiple sequences." *Bioinformatics* **15**(7-8): 563-577.
- Hiromi, Y., A. Kuroiwa, et al. (1985). "Control elements of the *Drosophila* segmentation gene *fushi tarazu*." *Cell* **43**(3 Pt 2): 603-613.
- Hong, J. W., D. A. Hendrix, et al. (2008). "Shadow enhancers as a source of evolutionary novelty." *Science* **321**(5894): 1314.
- Inagaki, S., K. Numata, et al. (2005). "Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*." *Genes Cells* **10**(12): 1163-1173.
- Inman, M., J. Zhou, et al. (2004). "Identification of a novel bovine herpesvirus 1 transcript containing a small open reading frame that is expressed in trigeminal ganglia of latently infected cattle." *J Virol* **78**(10): 5438-5447.
- Jakobsen, J. S., M. Braun, et al. (2007). "Temporal ChIP-on-chip reveals *Biniou* as a universal regulator of the visceral muscle transcriptional network." *Genes Dev* **21**(19): 2448-2460.
- Jin, H., R. Stojnic, et al. (2013). "Genome-wide screens for in vivo Tinman binding sites identify cardiac enhancers with diverse functional architectures." *PLoS Genet* **9**(1): e1003195.
- Johnson, D. S., A. Mortazavi, et al. (2007). "Genome-wide mapping of in vivo protein-DNA interactions." *Science* **316**(5830): 1497-1502.
- Junion, G., M. Spivakov, et al. (2012). "A transcription factor collective defines cardiac cell fate and reflects lineage history." *Cell* **148**(3): 473-486.
- Kapranov, P., J. Cheng, et al. (2007). "RNA maps reveal new RNA classes and a possible function for pervasive transcription." *Science* **316**(5830): 1484-1488.
- Kel, A. E., E. Gossling, et al. (2003). "MATCH: A tool for searching transcription factor binding sites in DNA sequences." *Nucleic Acids Res* **31**(13): 3576-3579.
- Kheradpour, P., A. Stark, et al. (2007). "Reliable prediction of regulator targets using 12 *Drosophila* genomes." *Genome Res* **17**(12): 1919-1931.
- Khoeiry, P., U. Rothbacher, et al. (2010). "A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites." *Curr Biol* **20**(9): 792-802.
- Kondo, T., Y. Hashimoto, et al. (2007). "Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA." *Nat Cell Biol* **9**(6): 660-665.
- Kondo, T., S. Plaza, et al. (2010). "Small peptides switch the transcriptional activity of *Shavenbaby* during *Drosophila* embryogenesis." *Science* **329**(5989): 336-339.
- Kulkarni, M. M. and D. N. Arnosti (2003). "Information display by transcriptional enhancers." *Development* **130**(26): 6569-6575.
- Lee, S. and M. D. Garfinkel (2000). "Characterization of *Drosophila* OVO protein DNA binding specificity using random DNA oligomer selection suggests zinc finger degeneration." *Nucleic Acids Res* **28**(3): 826-834.
- Lerner, D. W., D. McCoy, et al. (2013). "A Rab10-dependent mechanism for polarized basement membrane secretion during organ morphogenesis." *Dev Cell* **24**(2): 159-168.
- Li, X. Y., S. MacArthur, et al. (2008). "Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm." *PLoS Biol* **6**(2): e27.
- Liberman, L. M. and A. Stathopoulos (2009). "Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence." *Dev Biol* **327**(2): 578-589.

- Locke, M. (2001). "The Wigglesworth Lecture: Insects for studying fundamental problems in biology." *J Insect Physiol* **47**(4-5): 495-507.
- Locke, M., A. Kiss, et al. (1994). "The cuticular localization of integument peptides from particular routing categories." *Tissue Cell* **26**(5): 707-734.
- Locke, M. and N. Krishnan (1971). "The distribution of phenoloxidases and polyphenols during cuticle formation." *Tissue Cell* **3**(1): 103-126.
- Logan, S. K., M. J. Garabedian, et al. (1989). "DNA regions that regulate the ovarian transcriptional specificity of *Drosophila* yolk protein genes." *Genes Dev* **3**(9): 1453-1461.
- Loots, G. G. and I. Ovcharenko (2004). "rVISTA 2.0: evolutionary analysis of transcription factor binding sites." *Nucleic Acids Res* **32**(Web Server issue): W217-221.
- Lu, J., J. Andrews, et al. (1998). "*Drosophila* OVO zinc-finger protein regulates ovo and ovarian tumor target promoters." *Dev Genes Evol* **208**(4): 213-222.
- Ludwig, M. Z. and M. Kreitman (1995). "Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila*." *Mol Biol Evol* **12**(6): 1002-1011.
- Ludwig, M. Z., N. H. Patel, et al. (1998). "Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change." *Development* **125**(5): 949-958.
- Lukacsovich, T., K. Yuge, et al. (2003). "The ken and barbie gene encoding a putative transcription factor with a BTB domain and three zinc finger motifs functions in terminalia development of *Drosophila*." *Arch Insect Biochem Physiol* **54**(2): 77-94.
- Lukiw, W. J., P. Handley, et al. (1992). "BC200 RNA in normal human neocortex, non-Alzheimer dementia (NAD), and senile dementia of the Alzheimer type (AD)." *Neurochem Res* **17**(6): 591-597.
- Lusk, R. W. and M. B. Eisen (2010). "Evolutionary mirages: selection on binding site composition creates the illusion of conserved grammars in *Drosophila* enhancers." *PLoS Genet* **6**(1): e1000829.
- MacArthur, S., X. Y. Li, et al. (2009). "Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions." *Genome Biol* **10**(7): R80.
- Maerkl, S. J. and S. R. Quake (2007). "A systems approach to measuring the binding energy landscapes of transcription factors." *Science* **315**(5809): 233-237.
- Magny, E. G., J. I. Pueyo, et al. (2013). "Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames." *Science* **341**(6150): 1116-1120.
- Malik, S. and R. G. Roeder (2011). "The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation." *Nat Rev Genet* **11**(11): 761-772.
- Marcinkevicius, E. and J. A. Zallen (2013). "Regulation of cytoskeletal organization and junctional remodeling by the atypical cadherin Fat." *Development* **140**(2): 433-443.
- McGregor, A. P., V. Orgogozo, et al. (2007). "Morphological evolution through multiple cis-regulatory mutations at a single gene." *Nature* **448**(7153): 587-590.
- Meireles-Filho, A. C. and A. Stark (2009). "Comparative genomics of gene regulation-conservation and divergence of cis-regulatory information." *Curr Opin Genet Dev* **19**(6): 565-570.
- Mevel-Ninio, M., R. Terracol, et al. (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." *Mech Dev* **49**(1-2): 83-95.
- Michaut, L., S. Flister, et al. (2003). "Analysis of the eye developmental pathway in *Drosophila* using DNA microarrays." *Proc Natl Acad Sci U S A* **100**(7): 4024-4029.
- Morin, R. D., M. D. O'Connor, et al. (2008). "Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells." *Genome Res* **18**(4): 610-621.

- Mottola, G., A. K. Classen, et al. (2010). "A novel function for the Rab5 effector Rabenosyn-5 in planar cell polarity." *Development* **137**(14): 2353-2364.
- Moussian, B., J. Veerkamp, et al. (2007). "Assembly of the Drosophila larval exoskeleton requires controlled secretion and shaping of the apical plasma membrane." *Matrix Biol* **26**(5): 337-347.
- Murugasu-Oei, B., R. Balakrishnan, et al. (1996). "Mutations in masquerade, a novel serine-protease-like molecule, affect axonal guidance and taste behavior in Drosophila." *Mech Dev* **57**(1): 91-101.
- Negre, N., C. D. Brown, et al. (2011). "A cis-regulatory map of the Drosophila genome." *Nature* **471**(7339): 527-531.
- Ng, T., F. Yu, et al. (2006). "A homologue of the vertebrate SET domain and zinc finger protein Blimp-1 regulates terminal differentiation of the tracheal system in the Drosophila embryo." *Dev Genes Evol* **216**(5): 243-252.
- Noyes, M. B., X. Meng, et al. (2008). "A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system." *Nucleic Acids Res* **36**(8): 2547-2560.
- Nusslein-Volhard, C. and E. Wieschaus (1980). "Mutations affecting segment number and polarity in Drosophila." *Nature* **287**(5785): 795-801.
- Odenwald, W. F., W. Rasband, et al. (2005). "EVOPRINTER, a multigenomic comparative tool for rapid identification of functionally important DNA." *Proc Natl Acad Sci U S A* **102**(41): 14700-14705.
- Orlando, K. and W. Guo (2009). "Membrane organization and dynamics in cell polarity." *Cold Spring Harb Perspect Biol* **1**(5): a001321.
- Papatsenko, D., Y. Goltsev, et al. (2009). "Organization of developmental enhancers in the Drosophila embryo." *Nucleic Acids Res* **37**(17): 5665-5677.
- Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." *Nat Rev Genet* **10**(10): 669-680.
- Pataki, C., T. Matussek, et al. (2010). "Drosophila Rab23 is involved in the regulation of the number and planar polarization of the adult cuticular hairs." *Genetics* **184**(4): 1051-1065.
- Payre, F., A. Vincent, et al. (1999). "ovo/svb integrates Wingless and DER pathways to control epidermis differentiation." *Nature* **400**(6741): 271-275.
- Peng, Y. and J. D. Axelrod (2012). "Asymmetric protein localization in planar cell polarity: mechanisms, puzzles, and challenges." *Curr Top Dev Biol* **101**: 33-53.
- Plaza, S., H. Chanut-Delalande, et al. (2010). "From A to Z: apical structures and zona pellucida-domain proteins." *Trends Cell Biol* **20**(9): 524-532.
- Portales-Casamar, E., S. Thongjuea, et al. (2010). "JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles." *Nucleic Acids Res* **38**(Database issue): D105-110.
- Price, M. H., D. M. Roberts, et al. (2006). "Cytoskeletal dynamics and cell signaling during planar polarity establishment in the Drosophila embryonic denticle." *J Cell Sci* **119**(Pt 3): 403-415.
- Pritsker, M., Y. C. Liu, et al. (2004). "Whole-genome discovery of transcription factor binding sites by network-level conservation." *Genome Res* **14**(1): 99-108.
- Prud'homme, B., N. Gompel, et al. (2006). "Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene." *Nature* **440**(7087): 1050-1053.
- Purvanov, V., A. Koval, et al. (2010). "A direct and functional interaction between Go and Rab5 during G protein-coupled receptor signaling." *Sci Signal* **3**(136): ra65.
- Rana, N. A. and R. S. Haltiwanger (2011). "Fringe benefits: functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors." *Curr Opin Struct Biol* **21**(5): 583-589.

- Read, D., T. Nishigaki, et al. (1990). "The *Drosophila* even-skipped promoter is transcribed in a stage-specific manner in vitro and contains multiple, overlapping factor-binding sites." *Mol Cell Biol* **10**(8): 4334-4344.
- Rebers, J. E. and J. H. Willis (2001). "A conserved domain in arthropod cuticular proteins binds chitin." *Insect Biochem Mol Biol* **31**(11): 1083-1093.
- Ren, N., J. Charlton, et al. (2007). "The flare gene, which encodes the AIP1 protein of *Drosophila*, functions to regulate F-actin disassembly in pupal epidermal cells." *Genetics* **176**(4): 2223-2234.
- Repiso, A., P. Saavedra, et al. (2010). "Planar cell polarity: the orientation of larval denticles in *Drosophila* appears to depend on gradients of Dachshous and Fat." *Development* **137**(20): 3411-3415.
- Rister, J. and C. Desplan (2010). "Deciphering the genome's regulatory code: the many languages of DNA." *Bioessays* **32**(5): 381-384.
- Robasky, K. and M. L. Bulyk (2010). "UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions." *Nucleic Acids Res* **39**(Database issue): D124-128.
- Roy, S., J. Ernst, et al. (2010). "Identification of functional elements and regulatory circuits by *Drosophila* modENCODE." *Science* **330**(6012): 1787-1797.
- Sanchez-Elsner, T., D. Gou, et al. (2006). "Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax." *Science* **311**(5764): 1118-1123.
- Sandelin, A., W. W. Wasserman, et al. (2004). "ConSite: web-based prediction of regulatory elements using cross-species comparison." *Nucleic Acids Res* **32**(Web Server issue): W249-252.
- Sandmann, T., C. Girardot, et al. (2007). "A core transcriptional network for early mesoderm development in *Drosophila melanogaster*." *Genes Dev* **21**(4): 436-449.
- Sandmann, T., L. J. Jensen, et al. (2006). "A temporal map of transcription factor activity: *mef2* directly regulates target genes at all stages of muscle development." *Dev Cell* **10**(6): 797-807.
- Satija, R. and R. K. Bradley (2012). "The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the *Drosophila* embryo." *Genome Res* **22**(4): 656-665.
- Savard, J., H. Marques-Souza, et al. (2006). "A segmentation gene in *tribolium* produces a polycistronic mRNA that codes for multiple conserved peptides." *Cell* **126**(3): 559-569.
- Schena, M., D. Shalon, et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." *Science* **270**(5235): 467-470.
- Schier, A. F. and W. J. Gehring (1993). "Analysis of a fushi tarazu autoregulatory element: multiple sequence elements contribute to enhancer activity." *Embo J* **12**(3): 1111-1119.
- Schones, D. E., K. Cui, et al. (2008). "Dynamic regulation of nucleosome positioning in the human genome." *Cell* **132**(5): 887-898.
- Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." *Nat Biotechnol* **26**(10): 1135-1145.
- Slavin, I., I. A. Garcia, et al. (2011). "Role of Rab1b in COPII dynamics and function." *Eur J Cell Biol* **90**(4): 301-311.
- Small, S., A. Blair, et al. (1992). "Regulation of even-skipped stripe 2 in the *Drosophila* embryo." *Embo J* **11**(11): 4047-4057.
- Small, S., R. Kraut, et al. (1991). "Transcriptional regulation of a pair-rule stripe in *Drosophila*." *Genes Dev* **5**(5): 827-839.
- Smith, R. P., L. Taher, et al. (2013). "Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model." *Nat Genet* **45**(9): 1021-1028.
- Snyder, M., J. Hirsh, et al. (1981). "The cuticle genes of *drosophila*: a developmentally regulated gene cluster." *Cell* **25**(1): 165-177.

- Stanojevic, D., S. Small, et al. (1991). "Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo." *Science* **254**(5036): 1385-1387.
- Stark, A. (2013). "Regulatory Genomics - Decoding *Drosophila* Regulatory Sequences." *Biomed Tech (Berl)*.
- Stark, A., M. F. Lin, et al. (2007). "Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures." *Nature* **450**(7167): 219-232.
- Stathopoulos, A., M. Van Drenth, et al. (2002). "Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo." *Cell* **111**(5): 687-701.
- Stedman, E. (1950). "Cell specificity of histones." *Nature* **166**(4227): 780-781.
- Strutt, D. and S. J. Warrington (2008). "Planar polarity genes in the *Drosophila* wing regulate the localisation of the FH3-domain protein Multiple Wing Hairs to control the site of hair production." *Development* **135**(18): 3103-3111.
- Strutt, H. and D. Strutt (2009). "Asymmetric localisation of planar polarity proteins: Mechanisms and consequences." *Semin Cell Dev Biol* **20**(8): 957-963.
- Tagle, D. A., B. F. Koop, et al. (1988). "Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints." *J Mol Biol* **203**(2): 439-455.
- ten Bosch, J. R., J. A. Benavides, et al. (2006). "The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription." *Development* **133**(10): 1967-1977.
- Thanos, D. and T. Maniatis (1995). "Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome." *Cell* **83**(7): 1091-1100.
- Thomas-Chollier, M., O. Sand, et al. (2008). "RSAT: regulatory sequence analysis tools." *Nucleic Acids Res* **36**(Web Server issue): W119-127.
- Togawa, T., H. Nakato, et al. (2004). "Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, *Bombyx mori*." *Insect Biochem Mol Biol* **34**(10): 1059-1067.
- Ulitsky, I. and D. P. Bartel (2013). "lincRNAs: genomics, evolution, and mechanisms." *Cell* **154**(1): 26-46
- Walters, J. W., S. A. Dilks, et al. (2006). "Planar polarization of the denticle field in the *Drosophila* embryo: roles for Myosin II (zipper) and fringe." *Dev Biol* **297**(2): 323-339.
- Wang, Z., M. Gerstein, et al. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." *Nat Rev Genet* **10**(1): 57-63.
- Wasserman, W. W., M. Palumbo, et al. (2000). "Human-mouse genome comparisons to locate regulatory sites." *Nat Genet* **26**(2): 225-228.
- Wu, J. Y. and Y. Rao (1999). "Fringe: defining borders by regulating the notch pathway." *Curr Opin Neurobiol* **9**(5): 537-543.
- Yan, J., D. Huen, et al. (2008). "The multiple-wing-hairs gene encodes a novel GBD-FH3 domain-containing protein that functions both prior to and after wing hair initiation." *Genetics* **180**(1): 219-228.
- Zaret, K. S. and J. S. Carroll (2011). "Pioneer transcription factors: establishing competence for gene expression." *Genes Dev* **25**(21): 2227-2241.
- Zeitlinger, J., R. P. Zinzen, et al. (2007). "Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo." *Genes Dev* **21**(4): 385-390.
- Zhu, L. J., R. G. Christensen, et al. (2010). "FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system." *Nucleic Acids Res* **39**(Database issue): D111-117.

Zilber, Y., S. Babayeva, et al. (2013). "The PCP effector Fuzzy controls cilia assembly and signaling by recruiting Rab8 and Dishevelled to the primary cilium." Mol Biol Cell **24**(5): 555-565.

## Résumé

La fonction de chacune de nos cellules dépend d'un contrôle précis de son organisation tridimensionnelle. Des surdités génétiques sont par exemple dues à des défauts des extensions cellulaires de l'oreille interne, et la morphologie altérée des cellules cancéreuses favorise l'invasion tumorale. Il est bien établi que les cellules acquièrent progressivement leur forme au cours du développement de l'embryon, mais les mécanismes intimes de la différenciation morphologique restent mal compris. Comment des cellules qui contiennent toutes le même génome peuvent produire des formes aussi différentes ; certains de nos neurones sont 200 000 fois plus longs que nos globules rouges !

Pour aborder ces questions, mes travaux se sont concentrés sur le décryptage des mécanismes de différenciation morphologique de l'épiderme embryonnaire de drosophile. Nous avons entrepris un large effort multidisciplinaire international, impliquant génomique fonctionnelle, bioinformatique, physique et mathématique pour identifier, à l'échelle du génome entier, l'ensemble des effecteurs de la morphogénèse épidermique et leurs éléments de régulation transcriptionnels associés.

Nos résultats établissent, pour la première fois à notre connaissance, le répertoire « exhaustif » des gènes directement responsables du remodelage de la forme d'un type cellulaire. Ces données montrent la variété insoupçonnée des différentes fonctions impliquées, incluant régulateurs du cytosquelette, de la matrice extracellulaire, du trafic vésiculaire, de la polarité cellulaire, et de nombreuses protéines au rôle encore inconnu.

D'autre part, nous avons établi la cartographie génomique des éléments cis régulateurs, dits CRMs, qui dirigent la transcription des gènes effecteurs de la morphogénèse dans les cellules épidermiques. Par approche transgénique, nous avons démontré la fonctionnalité de plus d'une vingtaine de CRMs. En couplant dissection fonctionnelle *in vivo* et modélisation computationnelle, nous avons pu commencer à décrypter l'architecture fonctionnelle de ces CRMs, éclairant d'un jour nouveau la logique du contrôle de l'expression du génome.

Enfin, nous avons découvert un deuxième programme transcriptionnel qui contrôle la temporalité de la différenciation de l'épiderme. Nous identifions plusieurs centaines de gènes dont l'expression doit être normalement restreinte aux étapes post-embryonnaires du développement. Nous montrons que ces deux programmes indépendants, morphogénèse et différenciation, sont cependant coordonnés par des microprotéines régulatrices, récemment découvertes au laboratoire.

Ce travail met en évidence l'intérêt et les limites des nouvelles approches globales d'analyse de l'expression des gènes et démontre la puissance d'une démarche multidisciplinaire en biologie du développement. Il révèle l'étendue des acteurs du contrôle de la morphologie cellulaire et élucide les mécanismes permettant la coordination de leur expression.

## Summary

Cell function depends on a fine control of its morphogenesis. Morphological defects of cancer cell promote cell invasion, and genetic deafness depend on defects of inner ear cells extension. It is well known that cells acquire their shape during development, but mechanisms underlying morphological differentiation are not well described. How cells containing the same genome can finally present such morphological diversity?

To answer to this question, we undertook the deciphering of mechanisms implicated in morphological differentiation of the epidermis during drosophila embryonic development. Based on multidisciplinary approaches (functional genomic, bio-informatic, physic and mathematic), we identify at the genome scale the set of epidermal morphogenesis effectors and their respective regulatory elements.

For the first time, the “exhaustive” set of genes implicated in shape remodeling of a specific cell type is described, bringing on light an unexpected wide variety of functions, including cytoskeleton, extracellular matrix, vesicular trafficking, planar cell polarity regulators, and unknown proteins.

We also describe the genomic mapping of cis-regulatory elements (CRMs) driving transcription of effector genes implicated in epidermal morphogenesis. By transgenic assays we demonstrate *in vivo* the functionality of about twenty CRMs. Functional dissections associated to *in silico* modeling permit the deciphering of functional organization of these CRMs, putting forward the logic of the gene expression.

Finally we also identify a new transcriptional program controlling the timing of epidermis differentiation. We discover thousands of genes of which expression is restricted to post-embryonic steps of development, and show that the differentiation and morphogenesis programs are finally coordinated by recently identified regulatory micro-proteins.

This work brings to light the interest and the limits of new genome wide technics for gene expression analysis, and demonstrates the power of a multidisciplinary approach for developmental biology. It also reveals the extent of cell morphology regulators, and mechanisms implicated in the coordination of their expression.