Control of dorsal mesoderm specification in Drosophila melanogaster by transcription factors networks.

Nicolas Doll

Master BioSciences, Département de Biologie, École Normale Supérieure de Lyon.
Institut de Génomique Fonctionnelle de Lyon
2014-01-21

Keywords: Drosophila, dorsal mesoderm, transcription factors networks, Tinman, Wingless, cell specification

During embryonic development, the action of a great number of transcription factors (TFs) and pathways enables the specification of a tissue and the cell differentiation. Here we focus on the Drosophila dorsal mesoderm which is well studied. At early stages, two pathways, the Decapentaplegic (Dpp) and Wingless (Wg) pathways as well as the Tinman TF characterize the dorsal mesoderm. They activate the expression of master regulators which control the development of different parts of the dorsal mesoderm. Indeed, combinations of master regulators TFs lead to the activation of a great number of enhancers that regulating other TFs and pathways proteins, enlarging the TFs network in cells. These TFs activate particular combinations of genes coding for proteins directly involved in cell differentiation.

Introduction

How transcription factors (TFs) control embryogenesis is a longstanding and still poorly understood question. Mesoderm development is the best characterized model to answer it. The mesoderm is one of the three primary germ layers and is induced during the gastrulation. Mesoderm needs to be specified during early embryogenesis to subsequently differentiate into different subtypes that will give rise to various derived organs such as fat body, heart or muscles. An interesting model to study embryonic development is Drosophila melanogaster. Indeed, genetics and more recent molecular approaches allow describing quite in detail many actors involved in the different differentiation steps of mesoderm all along embryogenesis.

Drosophila mesoderm development proceeds in different phases. First of all, at stage 5, mesoderm is induced at the ventral pole of the embryo. At the next stage, it invaginates and extends along the anterior-posterior axis (Figure 1A). During these early stages, a few number of TFs expressed all over the mesoderm is induced. At stage 10, the Wingless (Wg) pathway is activated in stripes in the mesoderm, along the anterior-posterior axis. Then, at stage 11, stripes with the same identity migrate and merge together [1]. Wg pathway regulates specific TFs targets. So in the two domains, different TFs networks are established. These networks are controlled by master regulators TFs which activated more specialized genes inducing cell specific differentiation programs. TFs networks in the Wg domain will give rise to the visceral mesoderm (splanchnopleure) whereas the Wg* domain will form the cardiacl and somatic mesoderm (somatopleure) (Figure 1B) [2]. The splanchnopleure differentiates into fat body, gonads and some somatic muscles [2,3], whereas somatopleure leads to the formation of different somatic muscles, of the heart (Figure 1C), which is the first embryonic tissue formed, and of heart associated tissues.

The TFs networks leading to cell differentiation has been first characterized by genetic analysis such as loss of function studies. These approaches have allowed to identified the master regulators of these networks. However redundant genes or genes inducing low phenotype had not been identified. New approaches of transcriptome and Chromatin Immunoprecipitation (ChIP) analysis allow to characterized all genes targeted by a master regulator TF and so belonging to the network that induced cell differentiation. These studies have been performed on the dorsal mesoderm and wherein the focus will. Here we present how the dorsal mesoderm is specified over time, depending on signaling pathways and TFs networks.

Early signals characterizing the dorsal mesoderm

At early stages of development, the patterning of the dorsal ventral axis in the Drosophila embryo is controlled by the Dorsal protein. Dorsal is present in a dorsal-ventral nuclear gradient. Indeed, at the ventral pole, Dorsal enters in the nucleus, whereas at the dorsal pole, Dorsal stays in the cytoplasm [4]. Dorsal acts as a morphogene: different target enhancers are activated depending on the Dorsal concentration along the dorsal-ventral axis. Enhancers with a low affinity for Dorsal are activated at the ventral pole by high concentrations of Dorsal. So, some genes are only activated at the ventral pole. This is the case for twist (twi) which is the inductor of the mesoderm [5]. The mesoderm is so induced in the ventral part of the embryo, at stage 5. Just after gastrulation, at stage 6, Twi activates Tinman (Tin) expression, another key regulator of the mesoderm (Figure 1A) [6]. This activation occurs after the Twi binding on one of its four enhancers and leads to a broad expression of tin.
throughout the mesoderm [6]. Tin is the vertebrate NK-2 homologue and binds target enhancers on the NK-2 motif with its homeodomain. Tin is regarded as a pioneer TF that induces chromatin remodeling which may possible the binding of other TFs [7]. Thereby, Tin gives the ability for the dorsal mesoderm cells to respond to the different specification signals. Then the Twi enhancer of tin is inhibited by the binding of the Even skipped (Eve) repressor and a second enhancer is activated (Figure 1B). This enhancer allows a dorsal mesoderm restricted expression [6]. So at stage 10, tin expression is restricted to the dorsal mesoderm.

Another signal characterizing the dorsal mesoderm is the decapentaplegic pathway (Dpp). Unlike twi, dpp gene is inhibited by high concentrations of Dorsal. As a consequence, dpp is more activated at the dorsal pole.

Figure 1. Tinman expression is controlled by four enhancers that activate its expression in different regions over time. The genomic region with the four tin enhancers is represented on the left. The enhancers’ names were given by Yin et al [6] and are related to the domains in which they activate tin expression: Tin A = Anterior head element, Tin B = Broad mesodermal element, Tin C = Cardioblast element, Tin D = Dorsal mesoderm element. The TFs that inhibit the enhancers are represented in red and those that activate them in black. For three different stages, tin enhancers activation states and tin expression profiles are given. Tissues in which tin is expressed are represented on the embryos on the right. A Stage 6: The whole early mesoderm is delimited by a red line. Tin B is activated by Twist. Tin is expressed throughout the early mesoderm, except in the head mesoderm. Tin A is also activated by an unknown TF and tin is so expressed in the head too. B Stage 10: The dorsal mesoderm is delimited by a solid orange line and the ventral mesoderm by a dotted orange line. Tin B and Tin C are inhibited by Eve and Tin D is activated by Mad, Pnr and Tin itself. Tin D is repressed by Doc and Eve. Tin A is still active. So tin is expressed in the head and with fluctuations along the dorsal mesoderm. C Stage 16: The cardiac system is delimited by a yellow line. Tin B is still repressed by Eve. Tin D is inhibited by many TFs (Eve, Doc, Seven up (Svp), E(spl)-c) and Tin A is repressed by Buttonhead (btd). But Tin C is activated by unknown TFs. Tin expression is restricted to 4 cardiomyocytes and 2 pericardial cells per segment.
and so expressed in a reverse gradient of nuclear Dorsal along the dorsal-ventral axis [5]. Dpp acts as a morphogene too by activating genes depending on its concentration [8]. Dpp is only expressed in the ectoderm but diffuses within the underlying mesoderm and activates its own pathway into mesodermal cells. The activation occurs after the binding of the diffusible ligand Dpp on its receptor. This binding leads to the phosphorylation of the Mad protein, which recruits the Madea protein. The Mad/Madea complex penetrates into the nucleus and activates target genes. Dpp pathway is necessary for the specification of the dorsal mesoderm [9].

At stage 10, another important pathway is present in the dorsal mesoderm: the Wingless (Wg) pathway. The Wg pathway is activated by the binding of the Wg ligand on its receptor. Its effectors are the TF Pangolin (dTCF) and the two Sloppy Paired TFs: slp 1 and 2 [10]. The Wg ligand is secreted by the ectoderm and diffuses through five underlying cell layers within the mesoderm. But unlike Dpp, Wg ligand is expressed in stripes in the ectoderm [11]. As a consequence, the underlying dorsal mesoderm is striped by the Wg pathway in different patches along the anterior-posterior axis (Figure 2B) [10]. At stage 10, these Wg stripes are the only major difference observed within the dorsal mesoderm. But it leads to the specification between the visceral mesoderm, and the cardiac and somatic mesoderm.

**Activation of master regulators involved in the dorsal mesoderm specification**

The Wg pathway and its effectors induce changes in the activation state of many enhancers [12]. Some enhancers, inactive in the absence of Wg pathway are activated in Wg+ domains [10]. As a results the genes they control are expressed in stripes along the dorsal mesoderm. An example of these pair-rule genes is eve. During early stages Dpp effectors and Tin bind eve enhancer, but a repressor inhibits their activating action (Figure 2A). The binding of the Wg effector, dTCF, leads to an ousting of the repressor and an activation of the enhancer (figure 2B) [12]. Another master regulator of the somatopleure, the Dorsocross gene (Doc), is expressed in these Wg+ stripes. However Doc is only activated by Mad and dTCF, not by Tin [13]. These genes which are activated in the Wg+ domains, are master regulators of all somatic and cardiac mesoderm. Their activation is the first step of somatopleural specification.

In contrast, other enhancers are inhibited by the Wg pathway. Among them, some enhancers are already activated by Tin or Dpp effectors. Their activation is so restricted to the Wg- domain and the genes they control are expressed in negative stripes along the dorsal mesoderm. This is the case for bagpipe (bag)
coding for a NK homeobox protein, like tin [14]. On the bap enhancer, the bindings of Tin and the Dpp effector Mad are sufficient for its activation. However the fixation of an effector of the Wg pathway, the repressor Slp, inhibits the bap enhancer (Figure 2B) [10] [12]. The action of Slp may cause a chromatin remodeling or a recruitment of other repressors like Groucho, which represses the bap enhancer [12]. As a result, bap expression is restricted to the Wg stripes. The same mechanism is observed on the biniu (bin) enhancer. Bin is a FoxF TF with a forkhead domain that regulates directly or indirectly 1164 genes, more than half of which negatively [15]. Biniu induces a positive feedback on the dpp and the bap enhancers [14]. The Wg* stripes will form the future visceral mesoderm and Bap and Bin are its master regulators.

Furthermore, the activation state of some enhancers may not be affected by the Wg pathway. Some of them are already activated by Tin and Mad. In this case they control genes which are expressed in the whole mesoderm, at stage 10. For example tin is still expressed at this stage throughout the mesoderm. This is also the case of pannier (pnr), expressed in the dorsal mesoderm since stage 8 (Figure 2). Pnr expression in the dorsal mesoderm had been first considered as a consequence of intermediate Dpp concentrations [16]. But a recent study has proved that the pnr enhancer is not activated by a Dpp threshold [8]. On its two enhancers, a direct Tin binding has been showed while Mad, despite having specific binding motifs on the enhancer, interacts with Tin protein to activate pnr transcription (Figure 2) [17]. Pnr codes for a GATA TF that binds Tin proteins in vivo and is so regarded as a Tin cofactor [8]. Pnr activates Dpp and Tin expression, in a positive feedback loop [18]. Tin and Pnr are master regulators of both splanchnopleure and somatopleure. However, pnr and tin are differentially regulated by the other master regulators and their targets along the anterior-posterior axis. For example, in the Wg* patches, Doc and Eve inhibit directly tin expression [6]. On the other hand, Doc is an activator of pnr, a positive regulator of tin [13]. But ultimately, tin and pnr are more expressed in the Wg* than in the Wg* cells (Figure 1B) [8].

The Wg pathway induces so a particular regulation of gene expression from stage 10. The expression of different TFs in the two domains of the dorsal mesoderm is the first step of specification. In the Wg* stripes, Doc, Eve and low concentrations of Tin and Pnr control the differentiation program in somatic and cardiac muscles whereas in the Wg stripes, Bin, Bap and high concentrations of Tin and Pnr induce the visceral mesoderm specification. Furthermore, pair-rule genes, expressed in stripes, are mutually exclusive, allowing a clear molecular distinction between the different bands [13].

**Downstream networks inducing dorsal mesoderm specification**

At stage 11, the stripes of same identity migrate and merge together [1]. But the morphological cell differentiation and organs formation occur later. The first larval differentiated organ is the heart and it takes shape only at stage 16. So between the stages 11 and 16, splanchnopleural and somatopleural master regulator TFs induce the activation of a large number of target genes leading to cell differentiation and organ formation.

The first identified target genes of early master regulators had been characterized one by one thanks to loss of function studies and only genes with high phenotype were studied. Genes with a redundant function or specialized genes had not been taken into account. This approach had limited the view of downstream networks [19,20]. Recent progress in ChIP and transcriptome analysis techniques allows the identification of a great number of target enhancers [15,21]. The target genes activated by these enhancers are then identified thanks to a computer program. Let us take the case of Tin: Tin is the first inductor of the dorsal mesoderm. In addition, Tin is a pioneer TF which recruits other TFs by fixating on enhancers. A technique of ChIP on chip allows to localize the different regions bound by Tin and the cofactors recruited. So ChIP on chip with Tin provides the different enhancers involved in the dorsal mesoderm specification and differentiation [21]. Tin regulated more than 480 enhancers involved in the regulation of approximately 260 genes [22].

In the Wg* domain different enhancers are activated with particular combinations of TFs: enhancers only activated by Tin; enhancers activated by Tin, Doc and Pnr; enhancers activated by Tin, dTCF and Mad and enhancers activated by these five master regulators (Figure 3) [23]. These enhancers lead to the expression of specific somatopleural genes. Moreover, in the Wg* domain, enhancers of splanchnopleural genes are occupied by these TFs but they are not active [24]. The binding of the Slp repressor inhibits them (Figure 3). On the other hand, in the Wg* domain, the bindings of the Pnr, Tin, Bin and Bap TFs allow the activation of enhancers of splanchnopleural genes (Figure 3). Conversely, in this domain, enhancers of somatopleural genes are inactive because of a lack of Eve, Doc or dTCF.

In general, two different models are proposed to explain a TF binding on an enhancer: the enhanceosome model, in which the binding is highly cooperative and all proteins are fixed to one another; and the billboard model, in which the bindings of TFs are independent and the binding motifs order (grammar) on the enhancer is not very important [24]. Here bindings on enhancers are a mix of these two models. Tin is the first TF that binds on DNA. Then the other TFs are recruited cooperatively on the enhancer but the grammar of their bindings is not important [23]. The binding of a TF does not lead necessarily to the enhancer activation. Only a good combination of TFs on the enhancer can allow its activation [21,23].

Lots of genes are activated in the downstream networks. They code for proteins with manifold functions. These proteins could be classified into three categories: proteins involved in pathways, TFs
controlling the differentiation of a cell type, and proteins which directly induce cell differentiation [14]. Indeed Tin regulates general pathways, the JAK/STAT and the Eya pathways, both involved in somatic muscles formation [22]. Transcription factors, regulating more specific differentiation processes in the dorsal mesoderm, such as Hand, Pointed or Mef2 are also activated [25–27]. These TFs are involved in positive or negative feedbacks, and feedforwards, which contribute to the robustness and the refinement of the processes [22]. Finally genes that code for proteins directly involved in the morphogenesis are activated in these networks too. For example a cytoskeleton protein, the β-3 Tubulin, has its expression directly activated by Pnr and Bin [14] and likewise the enhancer of katp, a gene coding for an ATP-sensitive potassium channel involved in cardiomyocytes development, is directly bound by Tin and Pnr [28]. Among the Tin and Pnr targets, we find genes coding for proteins involved in the actin cytoskeleton organization such as scr68B [23,29]. The timing of the enhancer binding by master regulators and downstream TFs is related to the function of the gene. General genes are activated during all the specification and differentiation steps in many cells whereas specific genes are expressed during late stages of differentiation in a low number of cells [15].

Unexpectedly, some early master regulators are recycled and play a role in the differentiation of specific tissues. For example, at stage 16, tin is only expressed in a few cardiomyocytes and pericardial cells per segment (Figure 1C) [13,30]. Tin expression is restricted by the action of different actors, like the TF Lame Duck, a target of the Dpp pathway [31], and later by Seven up and Doc, in the Tin’s cardiomyocytes [30]. Interestingly a simple auto-restriction of tin expression is observed via the JAK/STAT signalling pathway. Tin activates the JAK/STAT pathway via Stat92E, which activates e(spl)-c genes that inhibites tin expression, inducing its restriction in some cardiomyocytes. Twi expression is restricted by this same pathway at late stages too [32]. The tin restricted expression in four cardiomyocytes and a few number of pericardial cells per segment is permitted by the activation of another enhancer, that controls its late expression (Figure 1C). This late tin enhancer is regulated by Wg [6]. The same mechanism is observed for mef2. As in the vertebrates, mef2 controls the differentiation in all Drosophila larval muscles [33]. At early stages mef2 is expressed in the whole mesoderm, then in the dorsal mesoderm and finally in muscle precursors. Early roles of mef2 are not well known but at later stages, mef2 is the key regulator for the muscles differentiation. The mechanisms of regulation of mef2 expression are
networks at a cellular level, rather than at a tissue level. Hundreds of TFs interact within these networks. Differentiation stages, thousands of genes and their targets are not well characterized over time and space. Today, the biggest genes networks had been established between them. But recent ChIP and transcriptomic studies show that their targets are the result of a change in the techniques used for reading enhancers, enabling the identification of many master regulators of a large number of enhancers. Then it is responsible of an induction of different master regulators in the two domains. This little difference is amplified thanks to the differential activation by these master regulators of a large number of enhancers. These enhancers control different types of genes: genes coding for pathway proteins, for TFs or for proteins directly involved in differentiation. These target genes are more specialized and their expressions are less broad. Furthermore, some master regulators are recycled and control the differentiation of a particular subset of cells. Our new knowledge on networks leading to the cell differentiation show that they are more complex than what was previously established. This is the result of a change in the techniques used for network characterization.

Indeed, during the 1990s and 2000s, studies of loss of function had enabled the identification of many master regulator genes. Then cis-regulatory networks had been established between them. But recent ChiP and transcriptomic studies show that their targets are very numerous. Feedforward and feedback loops make the prediction of their expression very complicated over time and space. Today, the biggest genes networks take into consideration 60 genes and all interactions are not well characterized [20]. In fact, during the differentiation stages, thousands of genes and hundreds of TFs interact within these networks. Current technologies make possible to study these networks at a cellular level, rather than at a tissue level [34,35]. Nucleus purification provides the epigenetic and transcriptomic landscapes for a particular cell [34]. One of the future challenges is to correctly characterize the trans-regulation of the different actors of these networks, and perhaps rebuild them numerically to predict gene expression in different cells.

**Conclusion**

The dorsal mesoderm specification is a well characterized model to understand the complexity of TF networks leading to cell differentiation. During early stages of its development, only a low number of actors are involved and they are broadly expressed. At stage 10, the dorsal mesoderm is striped by alternating Wg+ and Wg- domains. The presence or the absence of the Wg pathway and so of their effectors Slp and dTCF is the only difference within the dorsal mesoderm. This little difference is responsible of an induction of different master regulators in the two domains. Then it is amplified thanks to the differential activation by these master regulators of a large number of enhancers. These enhancers control different types of genes: genes coding for pathway proteins, for TFs or for proteins directly involved in differentiation. These target genes are more specialized and their expressions are less broad. Furthermore, some master regulators are recycled and control the differentiation of a particular subset of cells. Our new knowledge on networks leading to the cell differentiation show that they are more complex than what was previously established. This is the result of a change in the techniques used for network characterization.

Indeed, during the 1990s and 2000s, studies of loss of function had enabled the identification of many master regulator genes. Then cis-regulatory networks had been established between them. But recent ChiP and transcriptomic studies show that their targets are very numerous. Feedforward and feedback loops make the prediction of their expression very complicated over time and space. Today, the biggest genes networks take into consideration 60 genes and all interactions are not well characterized [20]. In fact, during the differentiation stages, thousands of genes and hundreds of TFs interact within these networks. Current technologies make possible to study these networks at a cellular level, rather than at a tissue level [34,35]. Nucleus purification provides the epigenetic and transcriptomic landscapes for a particular cell [34]. One of the future challenges is to correctly characterize the trans-regulation of the different actors of these networks, and perhaps rebuild them numerically to predict gene expression in different cells.
Lee et Frasch: Wingless effects mesoderm patterning and ectoderm segmentation events via induction of its downstream target sloppy paired. [date unknown], [no volume].


