AUS DEM ADOLF-BUTENANDT-INSTITUT, LEHRSTUHL: MOLEKULARBIOLOGIE IM BIOMEDIZINISCHEN CENTRUM DER LMU
VORSTAND: PROF. DR. PETER BECKER

HISTONE H4K20 METHYLATION PHASES CYTOSKELETAL DYNAMICS WITH THE CELL CYCLE DURING XENOPUS MUCOCILIARY EPITHELIUM FORMATION

DISSERTATION ZUM ERWERB DES DOKTORGRADES DER NATURWISSENSCHAFTEN

AN DER MEDIZINISCHEN FAKULTÄT DER LUDWIG-MAXIMILIANS-UNIVERSITÄT ZU MÜNCHEN

vorgelegt von Alessandro Angerilli
aus Rom, Italien
2018

Mit Genehmigung der Medizinischen Fakultät der Universität München

Betreuer: Prof. Dr. Ralph A.W. Rupp
Zweitgutachter: Prof. Dr. Magdalena Götz
Dekan: Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung: 12.11.2018

EIDESSTATTLICHE VERSICHERUNG

Angerilli, Alessandro

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

Histone H4K20 methylation phases cytoskeletal dynamics with the cell cycle during Xenopus mucociliary epithelium formation.

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 23-11-2018
Alessandro Angerilli

Ort, Datum Unterschrift
Doktorandin/Doktorand

Tag der mündlichen Prüfung: 12.11.2018
Eidesstattliche Versicherung Stand: 31.01.2013

Table of Contents:

Zusammenfassung

Die Embryonalentwicklung ist durch ein konstantes Komplexitätswachstum gekennzeichnet, das von der totipotenten Zygote zu den mehr als 200 verschiedenen Zelltypen beim Erwachsenen führt. Die differentielle Regulation der Genexpression und des daraus resultierenden Transkriptionsprofils garantiert die Etablierung einer solchen zellulären Diversität. In diesen transkriptionellen Unterschieden zwischen verschiedenen Zelltypen spielen Transkriptionsfaktoren und epigenetische Mechanismen eine zentrale Rolle. In dieser Dissertation untersuche ich die Differenzierung der embryonalen Epidermis in einem Xenopus-Modellsystem.

In dem ersten Teil meiner Dissertation untersuchte ich das Transkriptom eines autonom differenzierenden embryonalen Gewebes, das epidermale Differenzierung rekapituliert (*animal cap organoids*). Ich habe die Transkriptionsleistung dieser Organoide im Entwicklungszeitverlauf charakterisiert. Hierbei habe ich besonders Augenmerk auf bekannte Regulatoren der Zellidentität, wie Transkriptionsfaktoren, aber auch neue potentielle Regulatoren wie repetitive DNA-Elemente und zirkuläre RNAs, gerichtet.

Im zweiten Teil meiner Dissertation habe ich den Zusammenhang zwischen dem Verlust der H4K20 Histon Methyltransferase Suv4-20h1 / 2 und der darauf folgenden Reduktion der Multizilogenese in der embryonale Epidermis untersucht. Meine Ergebnisse deuten darauf hin, dass die H4K20-Methylierung sowohl die Dynamik des Zytoskeletts als auch den Zellzyklus beeinflussen. Somit kann ich in dieser Arbeit auf einen Einfluss der H4K20 Methylierung hinweisen, der weit über die Bedeutung für die epidermale Differentierung im Xenopusmodell hinausgeht.

Summary

Embryonic development is characterized by a constant growth in complexity, which leads, from the totipotent zygote, to the more than 200 different cell types in the adult. The differential regulation of gene expression, and thus of the transcriptional output, is what guarantees the establishment of such cellular diversity. Transcription factors and epigenetics mechanisms can be accounted mainly for the establishment of transcriptional differences between different cell types. In this dissertation I focused my attention on the differentiation of the embryonic epidermis in the Xenopus model system.

In the first part of my thesis I investigated the transcriptome of an autonomously differentiating embryonic tissue, the animal cap organoids, which recapitulates epidermal differentiation. I characterized the transcriptional output of these organoids in developmental time course. I focused my attention on the well-established fate regulators, such as transcription factors, as well as on novel potential regulators such as repetitive DNA elements and circular RNAs.

The second part of my dissertation aimed to explain why upon depletion of the H4K20 histone methyltransferases Suv4-20h1/2 multiciliogenesis on the embryonic epidermis is strongly impaired. My results strongly suggest that H4K20 methylation phases cytoskeletal dynamics in concert with the cell cycle. Here I will propose a model of the function of H4K20 methylation that extends far beyond the realm of Xenopus epidermal differentiation.

1.1 A Modern view of development

How is the genesis of a new complex individual achieved? How is a new human being manufactured? How does individuality arise? These questions have always captivated human curiosity. In the antiquity, Pythagoras and Aristotle defined a theory that we now call Epigenesis, which claims that complexity is built during embryonic development as a result of interactions that are established between the female and male fluids and is thus susceptible to environmental fluctuations. On the other side a completely different vision is represented by the Preformationism theory, that supports the idea that the complexity of a new-born organism is preformed prior to conception thus during gestation it can only grow and expand in size.

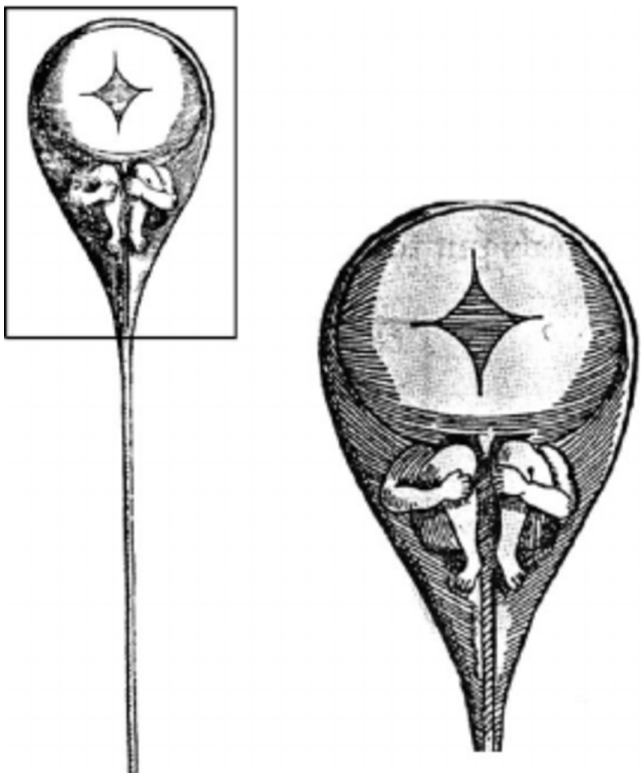


Figure 1. Source: http://blogs.plos.org/dnascience/2017/01/26/12-alternative-facts-of-human-genetics/figure-1-drawing-of-a-homunculus-by-nicholas-hartsoeker-a-famous-misperception-of/

One of the best representations of this theory, which represents the preformationist vision of the 17th century, comes from a famous illustration of Nicolaas Hartsoeker (Figure 1), that support the idea that preformed miniscule beings, called homunculi or animalculi, were already contained, with their fully-developed complexity, inside the male

sperm (or egg depending whether you were a “spermist” or an “ovist”).

In different historical periods one or the other theory has prevailed on the other, fed by new discoveries and observations. For example, the theory of evolution boosted the fame of *preformationism*: the observation that the ontogeny recapitulates the phylogeny and Darwin’s interpretation of this phenomenon implicitly suggested that preformed and defined modules not only exist prior to the embryo but existed long before the species. This left little room for Aristotelian interactions happening during the mating to play a role in the determination of the body plan complexity.

Both views of development are still very popular nowadays and none of the two has ever prevailed completely over the other. The dichotomy Preformationism vs Epigenesis is identifiable in what we now call Nature vs Nurture or Genetics vs Epigenetics.

In fact, the preformationist or genetic view of development recognizes the DNA as the source of all the information required to develop an individual. Little or no environmental perturbations are able to affect the embryogenesis. The individual is deterministically preformed in its DNA and the only possible outcome of the development is already encoded in the genetic information. On the other side, the epigenesis or epigenetic view of development suggests that environmental or stochastic factors take part in the formation of the body. Our individuality is not deterministically predicted and coded in our DNA but arises via the integration of environmental factors with our genetically encoded developmental program (Maienschein 2017).

Where do we stand? In an era were genetic and epigenetics manipulations have become ordinary we can weight how much we know so far of the two processes.

Let us consider the case of identical twins. Identical twins are identical because they share the same identic DNA. During development epigenetic mechanisms act on their genes independently though their action is not stochastic enough to render the twins different at the moment of birth and during their early lives. At age 3 to 4 years their transcriptome is virtually identical. Interestingly though is the fact that the more identical twins age the more differences accumulate in their transcriptomes and in their epigenomes (Fazzari and Greally 2004, Fraga, Ballestar et al. 2005, Robertson 2005). Finally, some epigenetic differences can be transmitted through the germ-line and affect the embryonic development in the following generation. This phenomenon is called transgenerational epigenetics and it has been described as playing a role for many different inheritable characters (Morgan, Sutherland et al. 1999, Heijmans, Tobi et al. 2009, Zenk, Loeser et al. 2017). In this scenario the epigenetic mechanisms integrate environmental information into the DNA-driven developmental process. As a result, Preformationism is not completely deterministic given the fact that body plans are susceptible to environmental stimuli. As well as epigenesis is never dominant; as we have seen for identical twins the outcome of their development can be clumsily predicted on the base of their DNA. Thus, embryonic developmental is the result of a

combined interaction between the predetermined genetic information and the environment it lives in.
It appears clear that we cannot stand on either side. Nature and nurture, genetics and epigenetics are two facets of embryonic development and their integration ultimately shapes who we are.

1.2 Xenopus as a model system

The African clawed frog Xenopus *laevis* is a very convenient animal model to study developmental biology. First of all, it is a tetrapod, meaning it is closer to the mammalian lineage than the fish model. A single female can lay up to several thousand mesolecithal eggs per day that can be all simultaneously fertilized. The fertilization is achieved using stored testis preparations in which sperms remain vital up to three weeks post-mortem. Most importantly, the embryonic development is simultaneous for all the embryos and takes place entirely outside the body of the adults; the eggs and the resulting embryos are big and can be easily observed and manipulated by hand without the use of sophisticated machinery. Doing molecular biology in Xenopus embryos is easy as well. DNA constructs, synthetic RNAs, recombinant proteins, fluorescent dextrans, antibodies and whatever small molecule can be easily injected inside the egg/blastomeres´ cytoplasm. Moreover, the eggs and the embryos can be used to prepare egg/embryonic extracts that is molecularly active and recapitulates ex-vivo many complex cellular mechanisms (Loughlin, Wilbur et al. 2011, Maller 2012, Hoogenboom, Klein Douwel et al. 2017). The Xenopus model comes with two major drawbacks. The first one is its life cycle: it takes up to

1 year for an animal to reach sexual maturity and can live up to 30 years. The second is its genome. In fact, the species is an allotetraploid that originated from an interbreeding event between two different frog species that happened around 17-18 million years ago. Its 3.1 x 10^9 bp genome is divided into 36 chromosomes where 56% of the genes retain a homoeologous allele pair. These characteristics made it hard to work with and just in 2016 X laevis genome has been published (Session, Uno et al. 2016).

For these reasons since many years the developmental community has identified in the Western clawed frog Xenopus *tropicalis* a good surrogate to bypass *laevis´*s drawbacks. X. *tropicalis* has a much shorter generation time (less than 5 months) and a much simpler genome. Its 1.7 x 10^9 diploid genome is divided in 20 chromosomes and its sequence is publicly available since 2010 (Hellsten, Harland et al. 2010). In general X. *laevis* is preferred to perform embryonic manipulations, gain of function experiments or RNA *in situ* hybridization, while X. *tropicalis* results more convenient when it comes to genetics or high throughput experiments. The two species are closely related and their embryonic development is almost undistinguishable (Grainger 2012).

Upon fertilization the resulting zygotes undergo a first division after 1.5 hours followed by 12 holoblastic cell divisions at 30 minute-intervals from one another, relying entirely on pre-deposited maternal transcripts. After the first 13 divisions the embryos reach the midblastula transition, when the zygotic genome gets first transcribed (Collart, Owens et al. 2014). At this stage the cells already

present a primary coelomic cavity, the blastocoel, which separates the cells on the animal hemisphere from the ones in the vegetal hemisphere. This primary coelomic cavity is thought to keep the animal pole cells away from the inductive signals coming from the vegetal cells and to favor the gastrulation movements; it disappears later on, being displaced by the archenteron. In fact, the vegetal cells have the capacity to induce neighboring cells to become mesoderm including transplanted animal pole cells that would by default give rise to ectoderm (De Robertis 2006). Most remarkably they do so in a position dependent manner. Animal pole cells transplanted next to the dorsal vegetal hemisphere will give rise to dorsal mesoderm derivates (notochord and muscle), while the ventral vegetal cells will induce ventral mesoderm derivates (mesenchyme, blood). This is possible because of the presence of dorsalizing signals produced by the Nieuwkoop center in the dorsal vegetal cells, whose position is pre-established inside the egg cytoplasm by the maternal Wnt signaling. The Nieuwkoop center determines as well the formation of the Spemann-Mangold organizer, the organizer of the gastrulation (and thus of the body plan), which corresponds to the dorsal lip of the blastopore. (Vonica and Gumbiner 2007).

The dorsal lip of the blastopore secretes a series of soluble signaling factors that have mainly 2 functions. First, factors like *chordin*, *follistatin* and *noggin* antagonize *bmp4* signaling that would otherwise induce all the ectoderm to become epidermis. Second, they give anterior-posterior positional information. In fact, when the gastrulation movements begin and the mesoderm starts invaginating and involuting into the blastopore the first mesodermal cells to

migrate enter the embryo from the dorsal lip. These cells leading the migration pass next to the dorsal lip experiencing a higher concentration of lip secreted factors such as Cerberus. Cerberus inhibits Xwnt8 signal and induce the formation of anterior structures. Later on, while the blastopore enlarges, cells start entering the embryo from its lateral portions and from the ventral lip. These cells receive less signals from the dorsal lip and will give rise to posterior mesoderm derivates (De Robertis and Kuroda 2004). The involuting axial mesoderm continues to secrete bmp inhibitors shielding the ectoderm above from its influence. The portion of the ectoderm that is now devoid of bmp4 will give rise to the neural lineage (Wilson and Hemmati-Brivanlou 1995). In fact, the ectoderm has a natural neural tendency and the epidermal fate has to be actively induced by *bmp4* (Hemmati-Brivanlou and Melton 1994).

As complexity increases inductive signals shape different cell fates in different tissues. Many of these inductive signals are known and can be easily manipulated experimentally as we will discuss in the next chapter.

1.3 The animal caps model

As we said before the blastula presents a blastocoelic cavity that separates the animal pole blastomeres from the vegetal pole ones. The uppermost part of the animal pole is called animal cap (AC). The animal cap constitutes the prospective ectoderm. When ACs are excised from the embryos at the blastula stage (NF6-9) they immediately round up to form spheres of prospective ectodermal

cells. The cells of these explants are pluripotent and they can be induced to form derivates from the 3 germ layers. For example by modulating the exogenous administration of *activin a,* the explants can be induced to develop into several mesoendodermal derivates (Asashima, Michiue et al. 2008). Similar results can be achieved via gain of function experiments that modulate the inductive signals required for lineage commitment such as BMP or FGF (Sasai, Lu et al. 1996). By combining the two strategies authors have been able to induce animal caps explants to differentiate into sophisticated tissues and organs (or organoids) such as cartilage, pancreas, liver, kidney, eye and even beating heart tissue (Okabayashi and Asashima 2003, Borchers and Pieler 2010). Competence of ACs to respond to inductive signals disappears by NF11 when cells become committed (Jones and Woodland 1987).

What happens to the animal caps when they are not induced? Normally in the context of the embryo they develop into epidermis and neural derivates following the natural gradient of *bmp4* and its inhibitors.

However, when they are isolated from the embryo they can no longer receive the *bmp4* inhibiting factors emanating from the dorsal lip of the blastopore. Their fate can still be switched towards neural, if *bmp4* gets inhibited experimentally by over expression of natural *bmp4* antagonists, such as *noggin*. In this case the entire AC gets neuralized (Zimmerman, De Jesus-Escobar et al. 1996). Otherwise when no manipulation is applied they develop into epidermal tissue (Sive, Grainger et al. 2007). ACs´ default differentiation recapitulates the development of the embryonic epidermis and leads to the formation of a skin organoid that is composed of the same exact 4 cell

types that constitutes the embryonic skin: multiciliated cells, serotonin producing cells, ionocytes and goblet cells (Werner and Mitchell 2013, Walentek, Hagenlocher et al. 2015). The embryonic epidermis and the resulting ACs organoids are what is called a mucociliary epithelium, characterized by the minimal presence of mucus producing cells (goblet cells) and multiciliated cells. The latter harbor hundreds of ciliary axonemes in apical tufts that beat synchronously to generate a fluid flow required for the clearance of bacteria and pollutant. As a result, the ACs explants spin around propelled by the cilia beating.

The amphibian´s embryonic epidermis is of particular interest to us because the human respiratory tract is covered by a mucociliary epithelium that closely resembles it (Werner and Mitchell 2012, Brooks and Wallingford 2014). It can be used to deduce important principles regarding the development, function and diseases affecting the human airflow tract such as cystic fibrosis (Cibois, Luxardi et al. 2015). Moreover, they are much more easily accessible than the human or the mouse tracheal epithelia. In recent years protocols have been developed to produce primary cultures of mammalian respiratory epithelia, though their generation and maintenance is extremely laborious (Davidson, Kilanowski et al. 2000).

The original aim of this dissertation was to make use of the AC model to investigate the embryonic differentiation in a simplified system such as the embryonic epidermis. This choice allowed us to study development in a context that is isolated from the complexity of the whole embryo. First, we explored the potential of the ACs system by

performing an in-depth transcriptomic study. Second, we took again advantage of this system to investigate the epigenetic control of multiciliated cells differentiation.

1.4 Multiciliated cells

Cilia are specialized cytoskeletal structures encapsulated within the plasma membrane and supported by a centriole. There are 2 types of cilia: non motile or primary cilia that function as sensory organelles, and motile cilia (or flagella) that are used for locomotion or for producing liquid flows. Cilia have an ancient evolutionary origin: except in few rare cases they present the same structure in all eukaryotes: 9 microtubules around a central axis (9+0) for primary cilia and 9 microtubules surrounding another pair of microtubules (9+2) in motile cilia or flagella. It is thought that the last common ancestor of all eukaryotes possessed a 9+2 flagellum (Mitchell 2007, Mitchell 2017). Primary cilia are found in almost all cells of the human body (Satir, Pedersen et al. 2010) and they are important cellular organelles. Defects in the formation of cilia are at the root of several human pathologies, commonly defined “ciliopathies”, such as polycystic kidney disease, situ inversus viscerum and primary ciliary dyskinesia (Mahjoub 2013, Pennekamp, Menchen et al. 2015, Milla 2016).

In humans, cells with multiple motile cilia are also found in the respiratory tract where they are required for the clearance of the mucus; in the epithelium covering the brain ventricles, where they generate a flow in the cerebrospinal fluid; and in the fallopian tubes/oviduct, where they are needed to move the egg cells away

from the ovaries (Brooks and Wallingford 2014). Cells harboring multiple cilia are defined as multiciliated cells or MCCs. Multiciliated cells usually present hundreds of ciliary axonemes (150 on average) each one being supported by a centriole (Brooks and Wallingford 2014). In the epidermis of the tadpole MCCs differentiate from a basal cell precursor found in the internal layer of the ectoderm. This progenitor cells express p63 and can differentiate

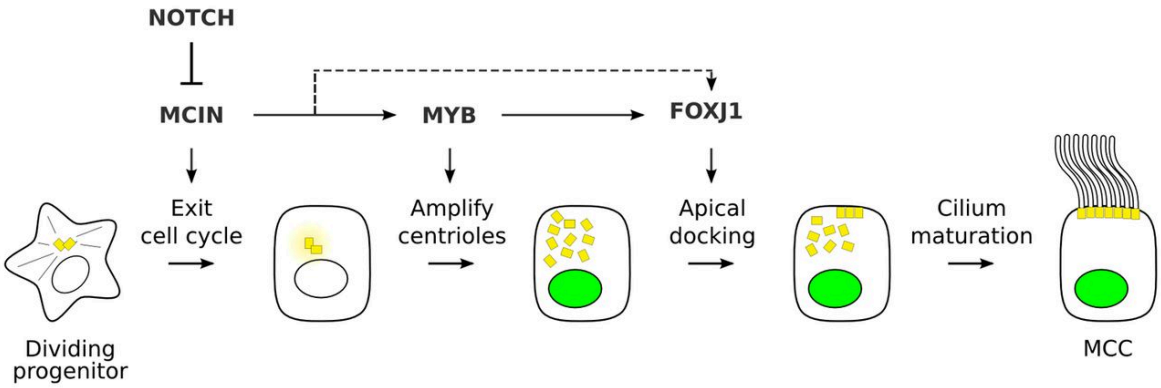


Figure 2: Multistep differentiation process of MCCs and molecular pathways involved as described in the text. (source: Tan, Vladar et al. 2013)

into either small secretory cells or multiciliated cells (Cole, Smith et al. 2010) (Warner, Hackett et al. 2013) that will then intercalate to the upper layer of the epidermis. On the embryonic epidermis MCCs are scattered in a salt a pepper manner such that 2 MCCs are rarely found adjacent to each other. This observation already suggests a mechanism of specification of multiciliated cell involving Notch lateral inhibition. Indeed, when Notch is inhibited there is an overproduction of MCCs, while the opposite is true when it gets up-regulated (Stubbs, Davidson et al. 2006)

In fact, Notch controls the expression of *mcin* (*multicilin,* also called *mcidas* or *mci*), the master transcription factor controlling multiciliogenesis (Stubbs, Vladar et al. 2012). Mcin is so potent that

its ectopic expression induces the formation of cilia tufts in other cell types and ectopic locations, while its down-regulation abolishes ciliogenesis in MCCs (Stubbs, Vladar et al. 2012). Mcin is a coiled coiled transcription regulator that activates the transcription of genes involved in multiciliogenesis via interactions with E2F4 and E2F5. Among its targets is the proto-oncogene Myb that is essential for inducing a S-phase like cell cycle state, during which centrioles are multiplied. Myb can activate the expression of *foxj1* (Tan, Vladar et al. 2013), a transcription factor that positively regulates the expression of structural proteins required for motile ciliogenesis (Yu, Ng et al. 2008). Among the genes controlled by Foxj1 we find intraflagellar transport proteins (IFT) required for the assembly of the cilium, as well as tubulins and tubulin-modifying enzymes (Didon, Zwick et al. 2013, Choksi, Lauter et al. 2014). The last steps of multiciliogenic differentiation require the apical docking of the multiplied centrioles at the cell membrane that can now support the formation of axonemal structures. Last, the ciliary axonemes can be visualized by staining with specific markers such as acetylated alpha-tubulin (Figure 2).

Historically the first regulator of ciliogenesis to be identified has been DAF-19 in C. *elegans*. In the worm DAF-19 controls the formation of non-motile cilia in sensory neurons. It binds and activates the transcription of ciliary specific genes and its loss provokes loss of cilia (Swoboda, Adler et al. 2000). DAF-19 belongs to the RFX family of transcription factors. In all vertebrates (excluded the actinopterygii clade) eight RFX factors have been identified so far (Chu, Baillie et al. 2010). Data show an involvement of many of these factors (demonstrated for RFX1,2,3 and 4) in ciliogenesis (Choksi,

Lauter et al. 2014) in a cell-type specific manner. For example, in human airways Rfx3 directly binds to Foxj1 and amplifies Foxj1-induced activation of ciliary genes (Didon, Zwick et al. 2013). My interest for multiciliogenesis arises from an observation of our lab (Hsam 2015) that upon depletion of two histone-modifying enzymes, namely Suv4-20h1 and Suv4-20h2, ciliogenesis in the embryo gets impaired. This is particularly true for multiciliated cells on the embryonic skin.

1.5 Definition of Epigenetics

As we said before the word epigenetics comes from *epigenesis.* The first modern definition of Epigenetics applied to embryonic development came from Conrad H. Waddington in the 1950´s. He theorized that a totipotent cell would have full differentiation potential like a ball resting on top of a hill has potential energy. When the ball rolls down one of many furrows, representing for the cell its possible fates, it loses its potential energy. The ridges separating the different furrows would be epigenetic barriers that restrict the fate of the cell by preventing its passage to another "furrow". What is the nature of this epigenetic barriers or signals? The first one to be discovered was the acetylation of histones in 1964 (Allfrey, Faulkner et al. 1964). Allfrey and colleagues noted that the histones associated with the DNA are often acetylated. Histones are basic proteins that form the nucleosome, an octameric complex composed of 2 copies of the 4 core histone proteins (H2A, H2B, H3 and H4) that form the basic unit of the DNA packaging system. Given the fact that DNA has a negative charge and the acetic acid group added to the histones

removes a positive charge from them. Allfrey proposed that this modification could diminish the affinity of the nucleosome for the DNA making it more accessible. Soon other epigenetic modifications were identified, the most important of which being histone methylation and DNA methylation. In 2001 it led Thomas Jenuwein and Dave Allis to speculate about the existence of a *histone code* (named after the genetic code), a code of chemical modifications that dictate (thus can be predictive) of the transcriptional status of a locus. In 2008 common agreement has been reached during a meeting in Cold Spring Harbor over a definition of epigenetics. The definition is the following*: "Stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence"* (Berger, Kouzarides et al. 2009). Many factors that contribute to the formation of an epigenetic memory have been identified. Among many the most relevant and best described include: small RNAs that can be transmitted from one generation to the other and influence chromatin formation; Histone modifications like the one described by Allfrey; Histone variants taking part in the formation of non-canonical nucleosomes; ATP-dependent nucleosome remodelers that move nucleosomes in order to induce chromatin compaction or decondensation; and last, DNA methylation that falls on CpG dinucleotides and represses transcription. In the next chapter we will explore more in detail some of these processes.

1.6 Chromatin

The human haploid genome is made of 3.3 billion base pairs. If one would stretch the DNA present in each single human cell its length

would be of approximately 2 meters. The cell nucleus though is usually smaller than 10 microns. How can such a long molecule fit into such a small volume? The answer is compaction. The actual human (haploid) genome in each cell is divided into 23 DNA molecules condensed around proteins to form the so-called chromatin fiber (i.e. 23 chromosomes).

As we said before the packaging unit of the DNA is the nucleosome. DNA has a width of 2 nm while the nucleosome is approximately 11nm in radius. Roughly 146 base pairs are wrapped around a single nucleosome.

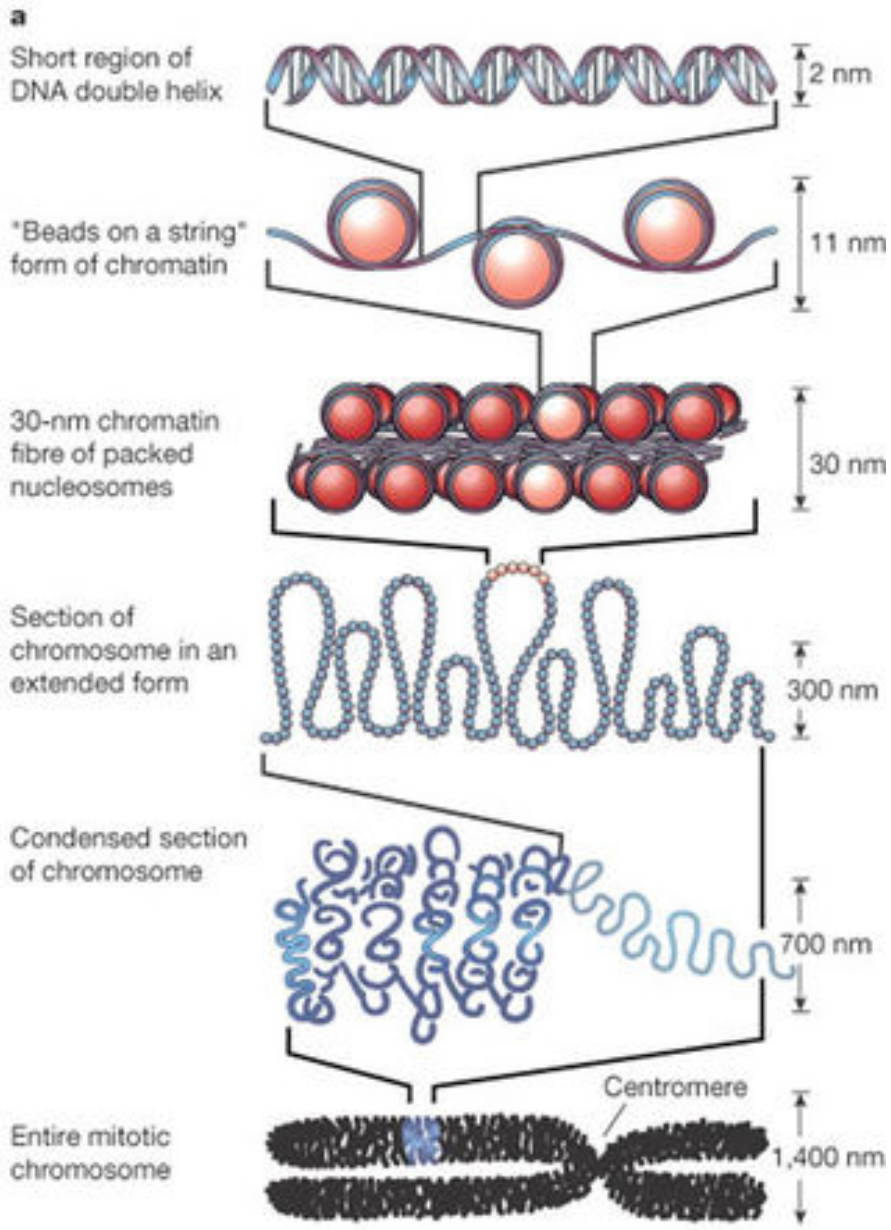


Figure 3: Organization of the eukaryotic genome. Source: (Felsenfeld and Groudine 2003).

The association of the nucleosomes with the DNA leads to the formation of a fiber that is usually defined as “beads on a string”. This already results in a compaction estimated around 5 to 10 folds (Figure 3)(Felsenfeld and Groudine 2003). The position of single nucleosomes is in general not fixed. Single nucleosomes or “beads” can be moved and repositioned for example to expose a DNA motif

that would otherwise be tightly bound or to arrange the spacing between adjacent nucleosomal particles.

Together with the 4 core histones another protein called histone H1 takes part in the compaction of the chromatin. H1 binds the so-called linker region, spanning between 10 and 80 bp, and is required to induce the formation of higher order structures. In particular H1 is thought to mediate contacts between nucleosomal particles and induce the formation of a 30nm chromatin fiber.

The 30nm chromatin fiber generates a compaction of approximately 50 folds. Starting from the 30nm fiber higher order chromatin structures are formed. The mechanism of assembly of such structures has not been fully elucidated yet. Other proteins such as cohesins, condensins and topoisomerases take part in the process. In the end the maximum level of compaction is reached by mitotic/meiotic metaphase chromosomes that reach a 10,000-20,000 fold compaction (Woodcock and Ghosh 2010).

Despite its unitary building block, the nucleosome, chromatin represents a heterogeneous matrix, meaning it is not identical in every portion of the chromosome. This is because histone proteins can be post-translationally modified altering their association with DNA. On top of that, histone variants can substitute canonical histones in the formation of the nucleosome. A typical example of histone variant is CENP-A. CENP-A takes part in the formation of nucleosomes replacing one or both H3 histones. Remarkably, CENP-A is found only at the centromere of the chromosomes, serving as scaffolding and anchoring point for many proteins involved in the regulation of the mitotic spindle (Regnier, Vagnarelli et al. 2005). CENP-A has mainly a *structural* function defining a particular domain

of the chromosome. In the same way other histone variants have a *functional* asymmetry. That is for example the case of the H2A variant macroH2A. MacroH2A is found associated with transcriptional repressed loci in the genome such as the inactivated X chromosome or polycomb-repressed genes (Zink and Hake 2016). It has been shown that macroH2A helps keeping genetic loci silenced by interfering with the binding of transcription factors and DNA remodeling machines (Angelov, Molla et al. 2003).

In this chapter of the introduction we have discussed the structural implications involved in the organization of the genetic material into chromatin. In the next chapter we will explore how chromatin defines and directs the functional status of the genetic material.

1.7 Chromatin colors

Histones can carry information by post-translational modifications (PTMs). The majority of these PTMs happen on the amino-terminal tails of the histones that protrude outside the nucleosome (Luger, Mader et al. 1997). This is particularly true for H3 and H4 that bear the highest number of PTMs. Most modifications are attached to the sidechains of lysines and arginines and their functions are the best understood so far. Arginine residues in the histone tails can be methylated (in a symmetric or asymmetric manner), de-methylated and deiminated. For the sake of this dissertation we will explore only lysine modifications in more detail. The most common modifications associated with lysine residues are methylation and acetylation. Histone acetylation is written by a class of enzymes called HAT (histone acetyl-transferases). As we said before from Allfrey on,

histone acetylation was predicted as a modification facilitating transcription, since it reduces the affinity of the histone proteins for the negatively charged DNA. This prediction was confirmed later, as for instance by the observation that H4K16 acetylation decorates transcriptionally active chromatin. It is known that K16 in histone H4 contacts the acidic patch of neighboring nucleosomes inducing chromatin compaction. The acetylation of this residue releases the binding of adjacent nucleosome particles making the underlying DNA more accessible to transcription factors and the transcription machinery (Zhang, Erler et al. 2017). H4K16 acetylation is for example enriched on the only male X chromosome in Drosophila melanogaster, that requires to be over-activated in order to produce a transcriptional output that is comparable to the one of the two female X-chromosomes. Once histones are acetylated, bromodomain-containing proteins can recognize the modification. These proteins are typically other HATs, which may increase locally the acetylated status on to other lysine residues, or proteins involved in the activation of transcription. For example the chromatin remodeler Brahma/brm possesses a bromodomain and thus associates with acetylated histone residues. Its function is that of remodeling the chromatin in transcriptionally active loci to facilitate transcription (Zhang, Erler et al. 2017).

In contrast, methylation of lysines can have different meanings. Lysines present a primary amino group that can be mono, di and tri-methylated. Lysine methylation is written by the so-called histone methyltransferases (HMTs). These enzymes can be recognized because they possess a common catalytic domain required for writing the modification: the SET domain, that takes its name from 3

of the best studied HMTs **S**u(var)3-9, **E**nhancer of Zeste and **T**rithorax. The SET-domain containing proteins use S-adenosyl methionine as a donor of methyl groups. Among the better-understood residues that undergo methylation there are the substrates of polycomb and trithorax group proteins H3K27 and H3K4 respectively, and the residues that get methylated in heterochromatic foci such as H3K9 and H4K20. K27 is of primary importance for the developmental processes. First of all, this residue can be both acetylated and methylated. Acetylation of K27 defines active enhancers (Creyghton, Cheng et al. 2010). The Polycomb group proteins operate by methylating K27 (Schuettengruber, Bourbon et al. 2017). K27 monomethylation (K27me1) is considered to be an activating mark. On the other hand K27me2 and K27me3 are the typical signatures of a Polycomb-repressed loci (Wang, Joshi et al. 2018). Polycomb was initially discovered for repressing the Hox genes cluster in D. melanogaster. Thus K27 methylation is considered a repressive mark associated with developmentally regulated loci. Opposite to Polycomb, the Hox gene cluster is activated by another HMT complex called Trithorax. Trithorax group proteins induce gene activation by tri-methylating the residue H3K4 along the gene body. This modification recruits chromatin remodelers that activate transcription and avoids the binding of repressing complexes (Schuettengruber, Bourbon et al. 2017).

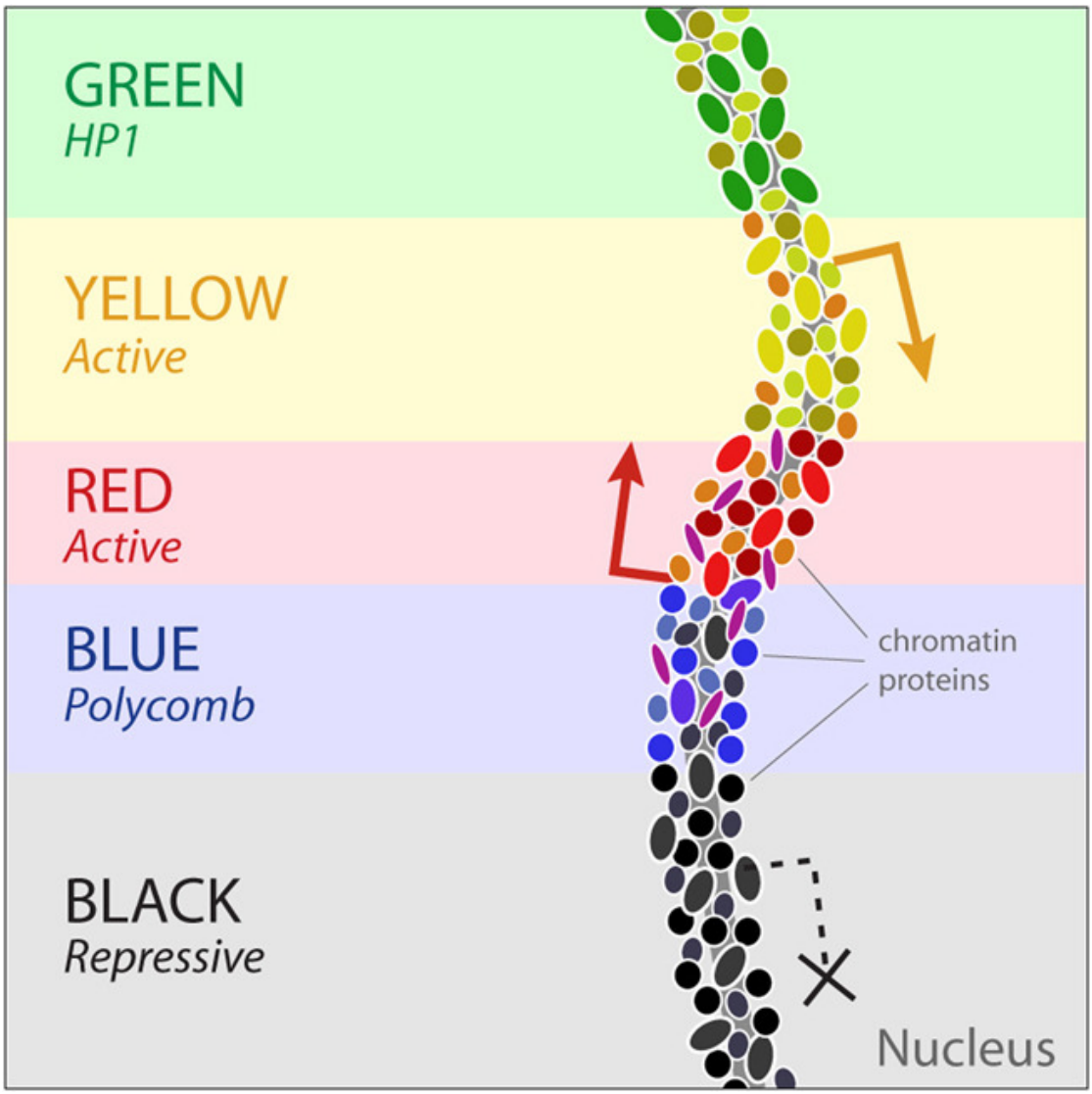
Apart from genes involved in embryonic development, large portions of the eukaryotic genomes have to be constantly silenced. This is particularly true for transposons derived sequences that, when expressed could become mobile and provoke mutagenic insertional events. These regions of the genome are repressed by other means.

In fact, they are enriched for H3K9me3, H4K20me3, HP1 (heterochromatin-protein1) and DNA methylation (Saksouk, Simboeck et al. 2015).

As we have seen up to now, different kinds of chromatin can be distinguished based on the modifications that are associated with it. Historically the chromatin that is transcriptionally active is defined euchromatin. It is less compacted and easy to access for transcription factors and the replication machinery (in fact it replicates in early S-phase). On the other hand, chromatin that is repressed is characterized by other histone modification, it is compacted and its contact points with TFs and the replication machinery are generally believed to be hidden within the fiber structure. Moreover, in order to distinguish cell type specific heterochromatin from the portion of the genome that is repressed in every cell type, the latter is called constitutive heterochromatin. It is particularly electron-dense and characterizes important areas of the chromosome such as the centromere and the telomeric regions (Saksouk, Simboeck et al. 2015).

The sub-functionalization of specific chromatin districts has become more complex in recent years. In 2010, Bas van Steensel´s lab published an article were they tracked the genomic location of 53 known chromatin proteins in a embryonic Drosophila cell line (Filion, van Bemmel et al. 2010). The occupancy overlap of different combinations of proteins allowed the definition of 5 different chromatin types named after colors: Blue, Green, Black, Red, and Yellow (Figure 4). Blue chromatin is one of the three types of repressive heterochromatin; it is enriched for Polycomb group proteins and H3K27me3. It affects developmentally regulated genes

and cell-type specific genes. Green chromatin is also heterochromatic, but it is enriched with Su(var)3-9 and HP1 and colocalizes with H3K9me2. Functionally it overlaps with transposable elements, centromeres, peri-centromeres, telomeres and peri-telomeric regions. Also some genes are found in Green heterochromatin, as for example those subjected to position effect variegation (Elgin and Reuter 2013). The most diffused kind of heterochromatin though is the black heterochromatin. This type of fiber covers more than 48% of the fly genome and is characterized by the co-occurrence of 4 proteins: (linker histone) H1, D1, IAL and SUUR. This heterochromatin has almost no transcriptional output; in fact 66% of the genes present in it have less than 1 read per 10 million mapped reads in RNA-seq studies. Some developmentally regulated genes fall into the black heterochromatin. These genes are expressed in a cell-type specific fashion, which let the authors speculate that black heterochromatin can be remodeled into different kind of chromatin different cell types. Finally, Red and Yellow chromatin represent euchromatin. Both of them have a large transcriptional output and produce the majority of the cellular mRNAs. Red chromatin is defined by the co-occurrence of proteins such as: Brahma, Su(var)2-10, Med31 ECR, GAF and JRA. Yellow chromatin is slightly different from the Red one. The features being particularly enriched in it are MRG15 and H3K36 trimethylation. The difference between these two kinds of euchromatin is evident, also when comparing replication timing of the underlying DNA and the expression patterns in the embryo of the contained genes.



(Filion, van Bemmel et al. 2010)

Very interestingly this model of the variety of forms that chromatin can assume, reflects end-point states and is relative to specific cell lineages. In fact, we know that the sub-functionalization of chromatin is not the same in every cell type and gets slowly established during embryonic development. Moreover, the cell cycle plays a pivotal role in the establishment and maintenance of epigenetic features (see below).

In the next chapter I will describe how the cell cycle influences chromatin and how it is intrinsically connected with the regulation of gene expression during embryonic development. We will then focus on the regulation of H4K20 methylation and finally we will see why these information is relevant for the sake of this dissertation.

1.8 Histone modifications through the cell cycle

When a cell enters the cell cycle the genetic information present in the DNA is passed to the daughter cells without any alteration. This cannot be the case for histone modifications. In fact, new histones devoid of any modification are synthetized and inserted into the newly replicated and the parental genomes during the S-phase of the cell cycle. As a result, the epigenetic information contained in the nucleosomes gets diluted by this newly incorporated histones (Annunziato 2005, Probst, Dunleavy et al. 2009). It has been proposed that modifications such as K9me3 and K27me3 present on parental histones may self-propagate via direct recruiting of their own writers (Hansen, Bracken et al. 2008, Margueron, Justin et al. 2009). Using a SILAC-pulse in vivo labelling approach to distinguish old from new histones, the group of Anja Groth has demonstrated that upon replication, histone modifications are diluted by a factor of two-fold. The majority of these histone modifications bounces back to the original amount within one cell cycle via active modification of the newly synthetized histones. The restoration of the original levels is achieved gradually and that argues against a co-replication establishment. A special case is found for H3K27me3 and H3K9me3.

The restoration of these two marks is slower, affects both newly synthetized and old histones and the complete restoration of the level of these marks can take up to 3 cell divisions. So the authors conclude that there are 2 general mechanisms of propagation of the epigenetic information during the cell cycle. The first one, that applies to the majority of the histone modifications, involves the copying of the epigenetic information contained in the parental histones into the newly synthetized ones. The second one, demonstrated for K9me3, K20me2 and K27me3, sees a slow and constant deposition of these marks on both newly synthetized and old histones (Alabert, Barth et al. 2015). Given the role that K27 methylation plays in directing animal development this result intrinsically suggests that regulation of the cell cycle speed plays a role in embryonic development via direct regulation of the amount of K27 methylation in blastomeres. In Xenopus and Zebrafish, the first 12 embryonic divisions are very fast and lack G-phases. In this context K27 methylation is undetectable. After MBT the embryonic mitotic divisions slow down and that correlates with the appearance of proper G1/G2-phases. In line with what we have said before it is not surprising then that K27 methylation becomes detectable only after the MBT (Akkers, van Heeringen et al. 2009) (Vastenhouw, Zhang et al. 2010).

In this dissertation, I have been investigating the impact of H4K20 methyl states on developmental programs. In this regard, it is important to review our current knowledge on the fact that H4K20 methylation is regulated in a cell cycle-dependent manner (Figure 5). When chromatin structure is reestablished during S-phase on the

two DNA helices, an enzyme called Pr-Set7 or Set8 mono-methylates K20 on the newly synthetized nucleosomes. Pr-Set7 travels along the replication machinery bound to PCNA, thus mono-methylation is happening in concert with the replication of DNA. This modification affects all the genome and K20 stays mono-methylated throughout the G2 phase of the cell cycle (Jorgensen, Schotta et al. 2013). Interestingly H4K20me1 is considered to be a repressive histone mark (Nishioka, Rice et al. 2002, Oda, Okamoto et al. 2009). The modification seems to directly control the expression of cytoskeletal and cell adhesion proteins; moreover, it affects cell proliferation, neuronal survival and craniofacial development (Qi, Sarkissian et al. 2010, Asensio-Juan, Gallego et al. 2012). A Jumonji-domain containing protein called PHF8 can erase K20 monomethylation. Otherwise during the G0/G1 phases of the cell cycle the Suv4-20h1 and Suv4-20h2 enzymes di and tri-methylate K20. The expression of these two enzymes is restricted to this phase of the cell cycle. H4K20me2 is considered to be the most abundant histone modification in vertebrates, covering more than 85% of the H4 molecules (MacAlpine and Almouzni 2013). Given such abundance the modification is not restricted to any particular area of the genome and plays no clear effect in regulation of gene expression. In the mouse, Suv4-20h1 seems to be the main responsible for K20me2 deposition. On the other hand, H4K20me3 is considered to be a repressive histone mark.

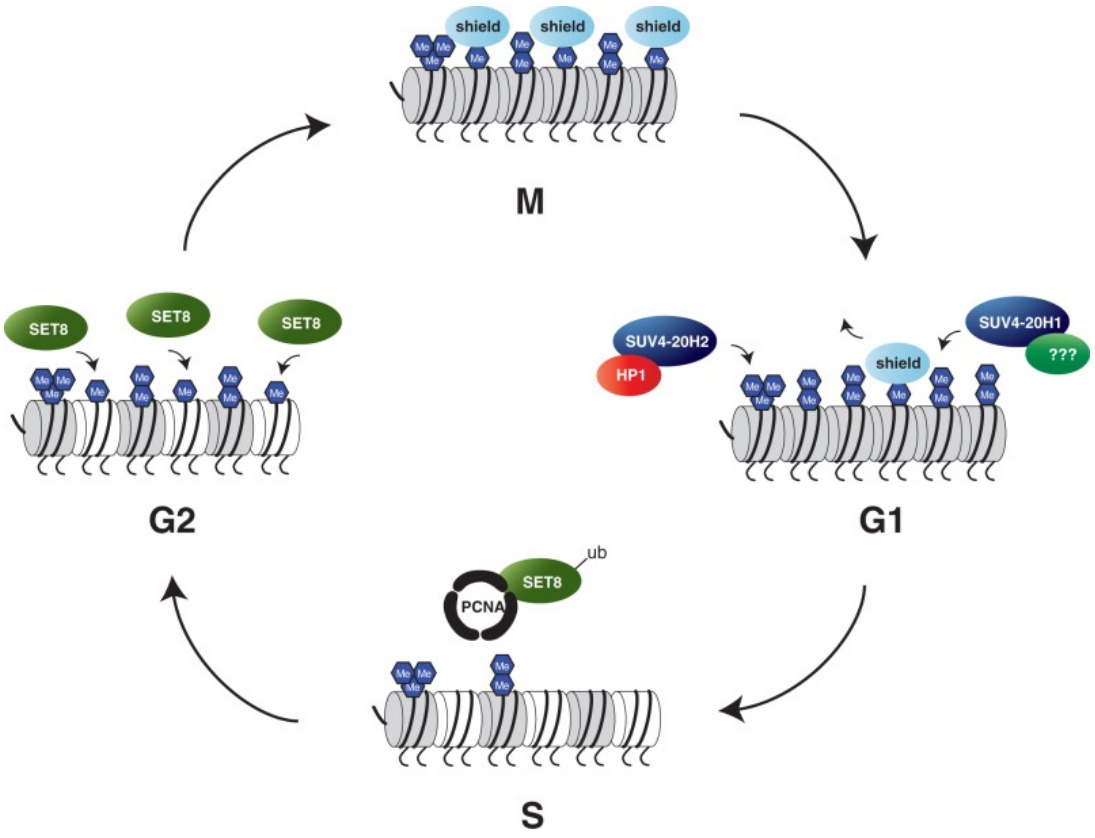


Figure 5: Cell-cycle regulation of H4K20 methylation. Source: Jorgensen, Schotta et al. 2013

It accumulates constantly throughout development (Schneider, Arteaga-Salas et al. 2011) and it localizes to specific chromosomal domains. In fact, K20me3 belongs to the Green heterochromatin and decorates the centromere, telomere, peri-centormeric and peri-telomeric regions (Filion, van Bemmel et al. 2010). In the mouse, K20me3 is deposited by Suv4-20h2. Mice that lack both Suv4-20h enzymes lose completely H4K20me2 and H4K20me3 and die perinatally. The loss of both enzymes induces also genome instability and provokes DNA rearrangements (Schotta, Sengupta et al. 2008). Consistent with these results, inhibition of the Suv4-20h enzymes with a specific small molecule inhibitor affects genome stability

(Bromberg, Mitchell et al. 2017). In the next chapter we are going to discuss in more detail the effects of manipulating K20 methylation state during embryonic development.

1.9 H4K20 methylation and development

As I said before, H4K20 monomethylation is written by Pr-Set7 and erased by PHF8. Interfering with Pr-Set7 function in embryonic stem cells leads to delays in the cell cycle, improper chromosome condensation and DNA damage. In Drosophila, mutations in this enzyme lead to lethality at the second instar larval stage (Beck, Oda et al. 2012) (Nishioka, Rice et al. 2002). Also mutations in the eraser of K20me1 have strong effects on development. Remarkably mutations in PHF8 are associated in human beings with the insurgence of X-linked mental retardation and craniofacial defects (Koivisto, Ala-Mello et al. 2007). Similarly, in Zebrafish, knock down of PHF8 provokes craniofacial abnormalities and impairs brain development in manner dependent on its catalytic activity (Qi, Sarkissian et al. 2010). The observed defects in brain development and the correlation with X-linked mental retardation can be explained by the fact that K20me1 regulates the expression of cytoskeletal genes such as RhoA, Rac1 and GSK3b. In fact, upon depletion of PHF8 many genes involved in cytoskeletal dynamics stay repressed. In neuronal cell lines model this results in impairment in the outgrowth of neurite and in their elongation, explaining partially how mutations in this gene can affect brain homeostasis (Asensio-Juan, Gallego et al. 2012).

The Suv4-20h enzymes are thought to convert sequentially the mono-methylated into di- and trimethylated states. Point mutations in Suv4-20h1 have been associated in humans with autism spectrum disorders in two independent genetic screenings (Iossifov, Ronemus et al. 2012) (Jiang, Han et al. 2015). In the mouse, depletion of the 2 enzymes leads to perinatal lethality. On the other hand, their up-regulation induces developmental arrest in the pre-implantation mouse embryo (Eid, Rodriguez-Terrones et al. 2016).

The situation is even more interesting in Xenopus embryos. Upon knock-down of the two enzymes via translation-blocking morpholino antisense oligonucleotides embryos fail to differentiate proper neuroectodermal derivates. The depletion predominantly affects the ectodermal germ layer. In fact, most genes, misregulated in double morphant embryos, are expressed in this germ layer. In consequence, neuronal differentiation is strongly reduced in these embryos. Moreover structures derived from cranial neural crest such as the skull, jawbones and melanocytes are defective (Nicetto, Hahn et al. 2013). These phenotypes go in concert with a depletion of K20me2 and K20me3 and an up-regulation of K20me1 and superficially resemble the ones described for Zebrafish PHF8 depletion (Qi, Sarkissian et al. 2010). Moreover, in Xenopus these phenotypes can be rescued via overexpression of the murine Suv4-20h enzymes that restore K20 methylation levels, but not by catalytically dead enzymes (Nicetto, Hahn et al. 2013). Last another phenotype arises upon Suv4-20h double knockdown and affects ciliogenesis. Ohnmar Hsam from our lab has demonstrated that the formation of multiciliated cells, both on the embryonic epidermis and in specialized structures of the embryonic kidney, is severely compromised in Suv4-20h depleted

frog embryos (Hsam 2015). The formation of ciliary axonemes, as revealed by acetylated alpha-tubulin staining and electronic microscopy, is strongly reduced or even abolished, despite master transcription factors like Multicilin and Foxj1 are only mildly affected. The expression of these two genes results to be unaffected when measured by qPCR, it has nonetheless to be normalized on the number of MCCs that in Suv4-20h-depleted epidermis are approximately 1.3 folds more abundant than in control ones. This increase in the MCCs number can be justified by the observation that Delta 1, a Notch ligand, is up-regulated at the mid-gastrula stage (NF12.5) in the double morphant prospective epidermis. However, the defect in ciliogenesis is not dependent on the slight down regulation of *mci* and *foxj1*. In fact, Julian Berges from our lab (personal communication/thesis in preparation) has demonstrated that the Suv4-20h-induced ciliogenic defect cannot be rescued by concomitant overexpression of Mci and Foxj1.

1.10 Objectives

The main objective of this dissertation was to define the transcriptional targets of the Suv4-20h enzymes responsible for the defect in multiciliogenesis. In order to identify such targets we took advantage of embryonic microinjections, to perform gain and loss-of-function experiments, RNA in situ hybridizations and Immunocytochemistry, to visualize the expression domains of particular genes, and high-throughput methodologies such as RNA and ChIp-seq, to study gene regulation.

In order to keep our focus on the embryonic skin we often made use of the animal caps organoid system. We performed RNA-seq in developmental time-course on these explants. In this context, we soon realized that our datasets were the deepest transcriptomic analysis ever performed on prospective embryonic epidermis. For this reason the result section of this dissertation will be composed of two parts. In the first part I am going to detail what is the transcriptional output of ACs and what can we learn about development from this embryonic system. In the second part I am going to take advantage of this newly developed resource to solve the developmental riddle of why Suv4-20h depletion affects multiciliogenesis.

2.0 Materials and Methods

2.1 Soutions for embryos production and handling

- Solutions:

- MBS 1X

Compound	Concentration
NaCl	880 mM
KCl	10 mM
$NaHCO_3$	24 mM
$MgSo_4$	8.2 mM
$Ca(NO_3)_2$	3.3 mM
$CaCl_2$	4.1 mM
Hepes	100 mM

Adjust pH to 7.6 with NaOH. Upon dilution add 10 ug/mL of gentamycin

- Steinberg´s Solution (SS)

Compound	Concentration
NaCl	580 mM
KCl	6.7 mM
$CaNO_3$	3.4 mM
$MgSO_4$	8.3 mM

Tris	50 mM
Kanamycin	0.1 g/L

Adjust pH to 7.4 with NaOH.

- Testes storage solution

Compound	**Concentration**
MBS	0.8X
Chicken serum	20%
Penicillin	200 U/mL
Streptomycin	200 U/mL

Store at -20°C and thaw upon use.

2.2 Superovulation & testes removal

Xenopus *leavis* females are injected with approx. 600 units of human Chorionic gonadotropin (Ovogest, MSD animal health) into the dorsal lymphatic sac. Ovulation starts 14-18 hours later. Xenopus *tropicalis* females are primed with 10 units of hormone. 48-12 hours later they are boosted with 150 units of hormone. Ovulation begins generally within 5 hours from the boosting (depending on the temperature, see below).

Xenopus males are anesthetized with a lethal dose (5g/L) of Tricaine. Usually within 30 minutes the animals are unconscious. They are then decapitated and testes can be excised from the abdominal cavity. A single *laevis* testis is enough for about 15 fertilization rounds while a *tropicalis* testes can be used for max. 3 fertilizations.

The testes are stored into the testes storage solution and remain vital for about two days (X. tropicalis) and 10 days (X. laevis).

2.3 In vitro fertilization and handling of embryos

Once the female frogs have started ovulating, eggs can be collected into petri dishes. Usually laevis *females* are squeezed gently, while *tropicalis* females are very fragile and cannot be squeezed. The occasional water around the eggs is to be removed. A small piece of testis (usually one half for X.*tropicalis*) is then homogenized with the help of a mortar into 1X MBS. The solution containing the sperms is then spread over the eggs. The mixture is left for 5 minutes (3 min for *X. tropicalis*) at room temperature and then flooded with freshly prepared 0.1X MBS. In this moment the fertilization begins and embryos are moved into incubators. *Laevis* embryos are held at temperatures ranging between 16 and 23 degrees Celsius while tropicalis embryos between 23 and 25°C.

2.4 Injection of the embryos

The embryos are singularly enveloped into a jelly coat that is to be removed in order to proceed with the injections. To do so, embryos are incubated with a solution of 0.1X MBS with 2% Cysteine at a pH of approx. 7.5. The dissolution of the coat can be easily followed by eye, indicated by touching-contacts between neighbouring embryos. Once the coat has been dissolved, the embryos are washed 3 times (4 for tropicalis) with 0.1X MBS and then transferred into 0.1X MBS +

gentamycin. Once at the right stage for the injections the animals can be placed on agarose grids in order to keep them still while injecting. The capillary needles required for the injections are generated using a Microneedle Puller (Sutter Instrument, model p-87). The needles are placed into the needle handling system (Medical System Pi-100) connected to a Picoliter Injector (Warner Instrument PLI-100A). Needles are manually broken with the help of a stereomicroscope (Stemi SV6, Zeiss) in order to obtain droplets of 1-5 nL. Injecting pressure was kept constant at 30.0 psi and injection time was modulated in the range of 0.05 to 0.4 sec. Injection volumes were adjusted not to overcome a maximum of 5 nL/embryo for X. *tropicalis* and 10 nL/embryo for X. *laevis.* Once the injection is completed the embryos are usually let to rest for 30 minutes and are then moved into a fresh, agarose-coated petri dish, where they continue to develop until the desired stage.

For morpholino injections in X. *tropicalis* we injected 20 ng of each morpholino per blastomere. X. *tropicalis* morpholino injections were lineage traced with Alexa 488 dextran (final concentration of 5ng/nl) in order to sort them into right or left injected for the acetylated alpha-tubulin staining. Morpholino injections on X*. tropicalis* were also used for the generation of double-morphant animal caps used for the transcriptomic analysis and qPCR. For Xenopus *laevis* we injected 40ng of each morpholinos per blastomere in order to generate animal caps for the *oct25* in situ hybridization, for the Western Blot analysis of K20 methylation and for the qPCR. For the confocal analysis of the ciliary phenotype X. *laevis* embryos were injected at the 8-cell stage in just one dorsal blastomere with 5ng of

each moprholino (control morpholino=10 ng) together with 150pg of Hyls1-GFP synthetic RNA in 2.5nl. X. *laevis* was injected with 300pg of synthetic Suv4-20h1 and Suv4-20h2 mRNAs for the Western Blot analysis of K20 methylation state.

2.5 Animal caps preparation

Animal caps were dissected between stage NF8 and NF9. The dissection is manually operated using a pair of forceps (Dumont #5, Fine Surgical Instrument). First the vitelline membrane was removed and subsequently the upper most part of the animal pole was excised. The explants were generated and kept into 1X SS + gentamycin solution. For the RNA-Seq analysis, each cap was cultivated in a single well of a 96-well plate containing 100 µL of 1X SS agar and 300 µL of Steinberg solutions. Animal caps were staged accordingly to sibling embryos. Embryos were staged accordingly Nieukoop and Faber (Nieuwkoop 1967). Animal caps showing elongation, which is indicative of mesodermal contamination, were discarded.

- Nucleic Acids

In Situ constructs. All the constructs are cloned into pCS2+ except Otogelin (pBS-SK) and Atp6v1e1 (pGEMTeasy).

Name	Linearization	Probe	Availability
Otogelin	EcoRI	T7	Axel Schweickert
Atp6v1e1	SacII	T7	Edgar Pera

Oct25	EcoRI	T7	Walter Knöchel
BMP7.2 RNA13254	HindIII	T7	de novo cloning
BMP7.2 RNA 13257	HindIII	T7	de novo cloning
CPNE1 RNA0071	HindIII	T7	de novo cloning
CPNE1 RNA 0068	HindIII	T7	de novo cloning
EVPL X1	ClaI	T7	de novo cloning
LTRX1-LTR_Xt	HindIII	T7	de novo cloning
ERV1-4-LTR_Xt	HindIII	T7	de novo cloning
Rem2b	HindIII	T7	de novo cloning

Cloning primers

Name	Direction	Sequence 5´-3
Circ6 F3	forward	5´gaaactctccagctctgtgaagtgg3´
Circ6 R3	reverse	5´gggcgtatgtggagctttgcttggc3´
Circ6 F4	forward	5´gccaagcaaagctccacatacgccc3´
Circ6 R4	reverse	5´ccacttcacagagctggagagtttc3´
Ext.6F3	forward	5´CTTGTTCTTTTTGCAGGATCCCATCGATTCgaaactctccagctc3´
Ext.6R3	reverse	5´CTATAGTTCTAGAGGCTCGAGAGGCCTTgggcgtatgtggag3
Circ27 F1	forward	5´gagagacaccggtctgtggtttgg3´
Circ27 R1	reverse	5´ggtccattttaagctcaatccgctc 3´
Circ27 F2	forward	5´gagcggattgagcttaaaatggacc3´
Circ27 R2	reverse	5´ccaaaccacagaccggtgtctctc3´
Ext.27F1	forward	5´CTTGTTCTTTTTGCAGGATCCCATCGATTCgagagacaccggtc3
Ext.27R1	reverse	5´CTATAGTTCTAGAGGCTCGAGAGGCCTTgagcttaaaatggacc3´
Circ1 F3	forward	5´ccttagactgctggaccgacttc
Circ1 R3	reverse	5´cggaaagatggagacatgtctgc
Circ1 F2	forward	5´gactttaggttgggaaggcagagtc 3´
Circ1 R2	reverse	5´ggatgtgcttggagaaggatcagtg 3´

cpne1 0071rna F	forward	GGATCCCATCGATTC tgcacattgggaactgccccttcc
cpne1 0071rna R	reverse	GGCTCGAGAGGCCTTctctatccattttccatctccaac
cpne1 0068rna F	forward	GGATCCCATCGATTC ctcatggctgggagcagcaactc
evpl X1 F	forward	GGATCCCATCGATTCcactccagcacagacctgtcc
evpl X1 R	reverse	GGCTCGAGAGGCCTT gctctgccctttaaacatggtg
BMP7.2 rna57 F	forward	GGATCCCATCGATTC ggagtttatttcttatcggcgctc
BMP7.2 rna57 R	reverse	GGCTCGAGAGGCCTT gtatttgatgagatggaactcaggg
BMP7.2 rna54 R	reverse	GGCTCGAGAGGCCTT gtatttgatgagatggaactcaggg
BMP7.2 rna54 F	forward	GGATCCCATCGATTC tgttcactctggcttaaacatgaatc
ERV1-4-I F	forward	5´GGATCCCATCGATTCgcatctggttgtctggtccaaac3
ERV1-4-I R	reverse	5´GGCTCGAGAGGCCTTgttctacagtatagggtggcaaagag3
Rem2b F	forward	5´GGATCCCATCGATTCcgcataagtcagtgtaggaactgg
Rem2b R	reverse	5´GGCTCGAGAGGCCTTgcttcaatatacatttagccaagggac
LTRX1-LTR F	forward	5´GGATCCCATCGATTCgtaagaaatcgtgtatttgggtgcc3
LTRX1-LTR R	reverse	5GGCTCGAGAGGCCTTcacccctcttactagggttccgg3
Xt Suv420h1 F	forward	5´GGATCCCATCGATTCATGAAGTGGTTGGGCGAATCCAAG3´
Xt Suv420h1 R	reverse	5´tactactcctactccTGCATTGAGTCTCAAGGATTGATC3´
Xt Suv420h2 F	forward	5´GGATCCCATCGATTCATGGGTTCAAATCGCTTGACGGCCA3
Xt Suv420h2 R	reverse	5´tactactcctactcc AGCGGTTTCTTCGCTCGATGTGGAT3

qPCR primers Xenopus *tropicalis*

Name	**Direction**	**Sequence**
Odc f	Forward	5´ccctggttcagaggacgtta3
Odc r	Reverse	5´AGTATCTCCCAGGCTCAGCA3´
Oct25 f	Forward	5´Agagtccccagaacccaa3´
Oct25 r	Reverse	5´aagggctaccagtccatgtg3´

Morpholino Oligonucleotides

Name	Sequence
Xl, Xt Suv4-20h1	5´ggattcgcccaaccacttcatgcca3´
Xl Suv4-20h2	5´ttgccgtcaaccgatttgaacccat3´
Xt Suv4-20h2	5´ccgtcaagcgatttgaacccatagt3´
Control	5´cctcttacctcagttacaatttata3´

Indexes/Indices for RNA-Seq Libraries

Sample Name	Index sequence
10.5 wt 1	CGATGT
10.5 wt 2	CGTACG
10.5 wt 3	ACAGTG
16 wt 1	GCCAAT
16 wt 2	CAGATC
16 wt 3	CTTGTA
24 wt 1	AGTCAA
24 wt 2	GTCCGC
24 wt 3	ACTGAT
10Co-Mo 1	ACTGAT
10.5 h1h2 1	ATTCCT
10Co-Mo 2	CGATGT
10.5 h1h2 2	TGACCA
10.5Co-Mo 3	GTCCGC
10.5 h1h2 3	ACTGAT
15Co-Mo 1	CAGATC
15 h1h2 1	CTTGTA
15Co-Mo 2	ACAGTG
15 h1h2 2	GCCAAT
15Co-Mo 3	AGTCAA

15 h1h2 3	GTCCGC
24 Co-Mo 1	ACAGTG
24 h1h2 1	CAGATC
24 Co-Mo 2	CGATGT
24 h1h2 2	AGTCAA
24Co-Mo 3	GCCAAT
24 h1h2 3	CTTGTA

2.6 RNA libraries preparation and sequencing.

For the library preparation each biological replicate derives from a different mating pair. Approximately 30 ACs were collected for each sample, vortexed in TRIzol (Ambion) until complete dissolution and snap-frozen in liquid nitrogen. Each developmental stage is represented by 3 biological replicates; total RNA was extracted from the animal caps using TRIzol (Ambion) and phenol/chloroform extraction. On-column RNA clean-up, including DNA digestion, was performed using RNeasy Mini-Kit (Qiagen). RNA quality was checked on Agilent Bioanalyzer 2100 using Agilent RNA 6000 Pico Kit. Ribosomal RNA was depleted using Ribo-Zero Gold rRNA Removal Kit (Human/Mouse/Rat) from Illumina using the manufacturer's instructions. The resulting Ribosomal-depleted RNA was cleaned and concentrated using RNA Clean & Concentrator™-5 from Zymo Research. Ribosomal depletion was verified using again the Agilent Bioanalyzer 2100. Libraries were prepared using NEBNext Ultra™ Directional RNA Library Prep Kit for Illumina (New England Biolabs) following the manufacturer´s instructions. Indexes and adaptors

come from NEBNext Multiplex Oligos for Illumina both primer sets 1 and 2. DNA purification and size selection was performed using Agencourt AMPure XP from Beckman Coulter. Quality and size of the libraries was verified on the Agilent Bioanalyzer using Agilent DNA 1000 Kit.

2.7 Sequencing data analysis

Next generation sequencing of AC RNA derived libraries (triplicates from 3 time points) were performed on the Illumina Hiseq4000 platform at BGI (Hong Kong). At least 30 million reads per sample were generated. We used additional datasets of RNA-seq samples from whole embryos made publically available by Owens et al (Owens, Blitz et al. 2016) and deposited at Sequence Read Archive (https://trace.ncbi.nlm.nih.gov/Traces/sra/) under following ids: SRR1795649, SRR651127, SRR1795663. All sequencing data were aligned to version 9.0 of Xenopus tropicalis genome as provided by Xenbase (Karimi, Fortriede et al. 2018) using STAR algorithm (https://www.ncbi.nlm.nih.gov/pubmed/23104886).

Genetic isoform analysis followed the established annotation from NCBI Xenopus tropicalis Annotation Release 103 which relies on genome annotation 9.1 (RSEM). To study the differences between time points and reference samples we make use of DSEQ2 package (Love, Huber et al. 2014). For analysis of repetitive elements, we use RepBase (Bao, Kojima et al. 2015). Transcription factor analysis was performed using the Mogrify algorithm as described by Rackham et al (Rackham, Firas et al. 2016). The list of transcription factors was obtained from Blitz et al (Blitz, Paraiso et al. 2017). We calculated the

interactions of the transcription factors using the STRING database (Szklarczyk, Morris et al. 2017).

2.8 Preparation of RNA in situ hybridization probes

For the RNA in situ hybridizations (In Situs) embryos were anesthetized in 0.05% benzocaine (N24), fixed in MEMFA (0.1M MOPS 2mM EGTA 1mM MgSO4 3.7% Formaldehyde) and stored in Ethanol. Antisense RNA probe sequences for analysis of splicing isoforms and repetitive DNA elements were selected from the top entries in Supplementary Tables 5 and 7. Embryonic RNA from mixed developmental stages (10.5-16-24) was extracted as described above. Total RNA was then reverse transcribed using SuperScript III (Invitrogen) by random priming. The primers used to amplify the targets present overhang homologies to the pCS2+ vector. Primers sequences are given in the supplementary Material (Supplementary table 8). The resulting amplicons were gel purified and cloned into pCS2+ using the Gibson Assembly Method (Gibson, Young et al. 2009). Otogelin (*otog*), *atp6v1e1*, epidermal keratin (*xk81a1*) and oct25 (*pou5f3.1*) plasmids were kindly provided by Drs. Axel Schweickert (Universität Hohenheim), Edgar Pera (Lund University) and Walter Knöchel (University of Ulm) respectively. The plasmids containing BMP7.2 RNA 13254, BMP7.2 RNA 13257, CPNE1 RNA0071, CPNE1 RNA 0068, ERV1-4-LTR_Xt, LTRX1-LTR_Xt and Rem2b were linearized using HindIII, otogelin, xk81a1 and oct25 with EcoRI, for EVPL X1 we used ClaI while for *atp6v1e1* we used

SacII. The linearized plasmids were subsequently transcribed overnight using T7 (antisense probe) and Sp6 (sense control) Polymerases (Promega) and cleaned-up with RNeasy Mini-Kit (Qiagen).

2.9 RNA In situ hybridization

Working solutions

- proteinase K: 10 mg/mL in H_2O and store at -20°C
- 1X MAB: 11,61 g Maleic acid, 8,76 g NaCl bring to pH 7,5 and filter
- Antibody solution: 2% BMB (Blocking reagent, Roche) in 1X MAB
- Alkaline phosphatase buffer: 100 mM Tris/HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$, 0,1% Tween 20.
- SSC (20X): 3M NaCl, 0.3 M sodium citrate. Bring to pH 7.0
- Staining solution: 4,5 µL NBT (100 mg/mL in 70% Dimethylformamide) 3,5 µL BCIP (50 mg/mL in 100% Dimethylformamide) in 1 mL of AP buffer. NBT= nitro blue tetrazolium, BCIP= 5-bromo-4-chloro-3-indolyl-phosphate
- Hybridization solution: 1% Boehringer blocking solution, 50 mg torula RNA, 0.01 Heparin, 0.1% Tween 20, 0,1% CHAPS, 5X SSC, 50% formamide, 5 mM EDTA. Store at -20°C and thaw upon use.

Bleaching solution: 1% Hydrogen Peroxide, 5% Formamide, 0,5% SSC

- MEMFA: 01M MOPS, 2 mM EGTA, 1 mM $MgSO_4$, 3.7% formaldehyde

- Procedure

Upon fixation embryos have to be kept at least over night in ethanol to solubilize and remove cell membranes. Then embryos were rehydrated into PBSw (= PBS + 01% Tween) in three steps (25, 50 and 75 %) on a roller. Thus they were washed 3 times x 5 minutes in PBSw. They were subsequently incubated in PBSw + 10 ug/mL of proteinase K for 20 minutes at RT. This step allows the degradation of eventual proteins associated with the RNAs. They were then washed twice for 5 minutes in PBSw. Upon this step embryos were refixed in PBSw + 4% of paraformaldehyde for 20 minutes on a shaker (gentle shaking). The paraformaldehyde was then washed with a PBSw rinsing and 5 washes in PBSw lasting 5 minutes each. The solution was then exchange with a solution of 50% PBSw and 50% hybridization solution. After 3 minutes the solution was discarded and embryos were incubated into 0.5 mL of 100% hybridization solution. After 3 minutes the hybridization solution was substituted. Endogenous alkaline phosphatases were inactivated via a 1h incubation in a water bath at 65°C followed by 2-6 hours of pre-hybridization at 60°C. 3-5 µL of dUTP probe were added to 100 µL of hybridization solution and heated to 95°C in order to release any secondary structure. The solution containing the probe was then added to the embryos in the water bath and let hybridize O/N. On the following day the solution was removed, exchanged with fresh hybridization solution and kept in the water bath at 60°C for 10 minutes. The solution is then rinsed with 2% SSC and the embryos were then incubated with three washes (20 minutes each) of 2% SSC at 60°C. Subsequently the vials were rinsed with 0.2% SSC and incubated in 2 washing of 30 min each of 0.2% SSC at 60°C. This step is critical because it removes the aspecifically bound probes. After

the second wash the solution was promptly substituted with MAB buffer for a total of 2 washes lasting 15 minutes each at RT. The MAB buffer was exchanged with 1 mL of antibody buffer and embryos were incubated for 1% on a rocking platform. After that the ab buffer was removed and substituted with 0.5 mL of new antibody buffer with 1:2000 dilution of anti-dig antibody (Anti digoxigenin-AP Fab fragment, Roche) that was left reacting with the embryos for not less than 4 hours. The embryos were then rinsed and washed in MAB buffer overnight on a roller.

On the following day the specimens were washed for 3-4 times in MAB: Each wash lasts 1 hour. They were then rinsed and incubated in alkaline phosphatase buffer for 15 minutes on a roller at RT. AP buffer was replaced with NBT/BCIP staining solution and the color reaction was checked by eye until the staining emerges. Washing with PBS stops the color reaction and embryos are refixed for at least 90 minutes in MEMFA. Last, bleaching of pigment granules was achieved in the following way: embryos were dehydrated in 75%b EtOH in PBS for 15 minutes and then moved into bleaching solution on a visible-light source, followed by re-fixation in MEMFA. Embryos were photographed using Leica M205FA stereomicroscope. Pictures were obtained as Z-stack acquisitions.

2.10 Detection of Circular RNAs

CircRNAs were identified with the method described by Westholm and colleagues (Westholm, Miura et al. 2014). Primer sequences are listed in the table below. We used diluted libraries as template. The resulting amplicons were visualized using Agilent Bioanalyzer using

Agilent DNA 1000 Kit. In order to reach an appropriate length for sequencing a second PCR with primers containing SP6 promoter and pCS2+ derived backbone DNA sequence was performed for selected candidates (primer sequences in the table below labeled as ext.). Sequencing was conducted at MWG Eurofins.

2.11 qPCR

qPCR was performed on the LightCycler® platform using multi well plates 384/white provided by Roche. The reaction mix was obtained mixing 1 µL of cDNA template, 5 µL Fast SYBER Green Master Mix supplied by Applied Biosystem, 1 µL of 3 µM Primers mix and 3 µL of H_2O. The thermal cycle is standard and is composed in the following way

Initial denaturation	95°C	5 min	1X
Denaturation	95°C	10 sec	45X
Annealing	60°C	20 sec	45X
Elongation	72°C	30 sec	45X
Melting	95°C	5 sec	1X
Melting	65°C	1 min	1X
Cooling	40°C	30 sec	1X

- Protein Methods

2.12 Immunocytochemistry (ICC)

Working solutions

- PBT: PBS, 2 mg/mL BSA, 0.1% triton x-100
- Antibody buffer: PBT + 10% of heat inactivated lamb serum
- Alkaline phosphatase buffer: 100 mM Tris/HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$, 0,1% Tween 20 + 25,5 mg/100mL Leavamisole
- Bleaching solution: 1% Hydrogen Peroxide, 5% Formamide, 0,5% SSC

Procedure:

After fixation in MEMFA embryos were and stored in methanol. This step removes the cell membranes. Embryos were then slowly rehydrated by increasing the PBT/Methanol ratio. The rehydration was followed by 2 washes of 15 minutes each in PBT. The embryos were subsequently incubated for 1 hour in antibody buffer in order to block possible aspecific antibody binding sites. Following the embryos were incubated O/N with antibody buffer + the primary antibody (1:500 Monoclonal Anti-Acetylated Tubulin antibody Sigma-Aldrich, T6793).
On the following days the samples were washed for 5-6 times in PBT. Each wash lasted between 30-60 minutes. Then they were incubated overnight with the secondary antibody in AP-buffer (Shp x Mouse Fab IGg Alk Phos, Chemicon 1:1000). On the following day embryos were washed 5-6 times in PBT for 30-60 min per wash. Last they were incubated with the alkaline phosphatase buffer containing the endogenous alkaline phosphatase inhibitor levamisole for 15 minutes. Last the color reaction was performed in AP buffer + BCIP-

NBT. The color reaction is usually extremely quick (below 15 minutes) and is stopped by washing the samples with PBS. After the color reaction the embryos were refixed in MEMFA for a minimum of 90 minutes, dehydrated in 75% EtOH in PBS and bleached in bleaching solution under a visible light bulb.

2.13 Nuclei Extraction

Working solution

- E1 complete: 90 mM KCl, 50mM Tris-HCl pH 7.4, 5mM MgCl, 0.1 mM EDTA, 10 mM sodium-butyrate, 0.1 mM PMSF, 0.02 mM Leupeptin, 0.1 mM DTT.
- add 0.25M or 1.25M sucrose for the relative solutions.

Procedure:

We collected approximately 80-100 injected NF18 embryos per sample. MBS was removed and the embryos were washed 3 times into E1+0.25M sucrose and centrifuged at 600rpm in a table top centrifuge. We removed as much solution as possible and we incubated the embryos for 20 min at RT into 1 mL of E1+0.25M sucrose. Embryos were then transferred into a 5 mL glass douncer (Braun, Melsungen) with 3 mL of E1+0.25M sucrose and homogenized with a total of 20 gentle strokes. The lysate was then moved into a 15 mL falcon and centrifuged for 10 min at 1000 rpm at 4°C in a swing out rotor. This step separates the cytoplasmic fraction from the nuclear one that sediments as a pellet. The pellet was

subsequently resuspendend in 3 mL of E1+0.25M sucrose with 0.5% triton x-100. The resuspended nuclei were then incubated on ice for 20 minutes. AT this point we prepared 50 mL falcons with a cushion of 5 mL of E1+1.25M sucrose. The nuclei solution was then carefully overlaid on the cushion and the falcons were centrifuged for 30 minutes at 1000 rpm at 4°C in a swing out rotor. The nuclei are at this point in the pellet; we then discarded the supernatant and resuspended the nuclei into 1.3 mL of E1 without sucrose. The sample was then moved into normal 1.5 mL tubes and centrifuged for 5 minutes at 5000 rpm at 4°C. We then washed the pellet with 1.3 mL of fresh E1 solution and repeated the centrifugation step. Finally the pellet can be resuspended in Laemmli buffer in the amount of 2.5 µL/embryo.

2.14 SDS-PAGE and Western blot analysis

SDS-PAGE and Western Blot analysis were performed accordingly to Sambrook et al 1989. We used precast 4-16% gradient polyacrylamide gels (Serva-Gel TG Prime) and blotted on PVDF membranes (Roti-PVDF 0.45 µm). The chemiluminescent substrate used is Amersham ECL and ECL Plus (GE-Healthcare). Membranes were exposed to on X-rays films (Super-RX Fuji) with different exposure times. In some instances the signals were recorded using ChemiDoc (Bio-Rad).

Epitope	**Provider**	**Dilution**	**Reference**

H4K20me1	Hiroshi Kimura	1:2500	Hayashi-Takanaka et al 2015
H4K20me2	Gunnar schotta	1:500	Schotta et al, 2008
H4k20me3	AbCam	1:1000	Ab9053
panH3	AbCam	1:10000	Ab1791
Anti-mouse	Jackson Immunoresearch	1:2500	115-035-003
Anti-Rabbit	Jackson Immunoresearch	1:10000	111-035-144

3.0 Results

3.1 Animal caps differentiate into epidermis

The differentiation of Animal Cap (ACs) explants recapitulates that of the embryonic epidermis (Hamburger 1955, Asashima and Grunz 1983, Jones and Woodland 1987). To verify this assumption ACs were isolated at the blastula stage, harvested at NF24 and stained with a series of RNA in situ hybridization (in situ) and immunocytochemistry (ICC) directed against specific epidermal markers (Figure 1A). Acetylated alpha-tubulin (ICC) stains multiciliated cells, otogelin stains goblet cells, xv-atpase specifically recognizes small secretory cells while keratin is a pan-epidermal marker. The staining patterns on the animal caps is comparable to the one observed on the embryonic epidermis and confirms that ACs differentiate into a mucociliary epithelia (Figure 1A). Consequently, my first goal was to characterize factors that are relevant for the differentiation of the embryonic epidermis. For this reason, I wondered how the transcriptome maturates during ACs' differentiation. In order to generate transcriptomic datasets, ACs from X. *tropicalis* were generated at the end of the blastula stage (NF9). 30 ACs were harvested at the gastrula (NF11), neurula (NF16) and tailbud (NF24) stage. Staging was performed accordingly to sibling embryos. Three biological replicates were produced for each time point (Figure 1B). RNA was extracted and used to prepare paired-end RiboZero libraries. Approximately 30 million reads per sample were generated, summing up to almost 100 million reads per developmental stage. Bioinformatic analysis was conducted in collaboration with Pawel Smialowski.

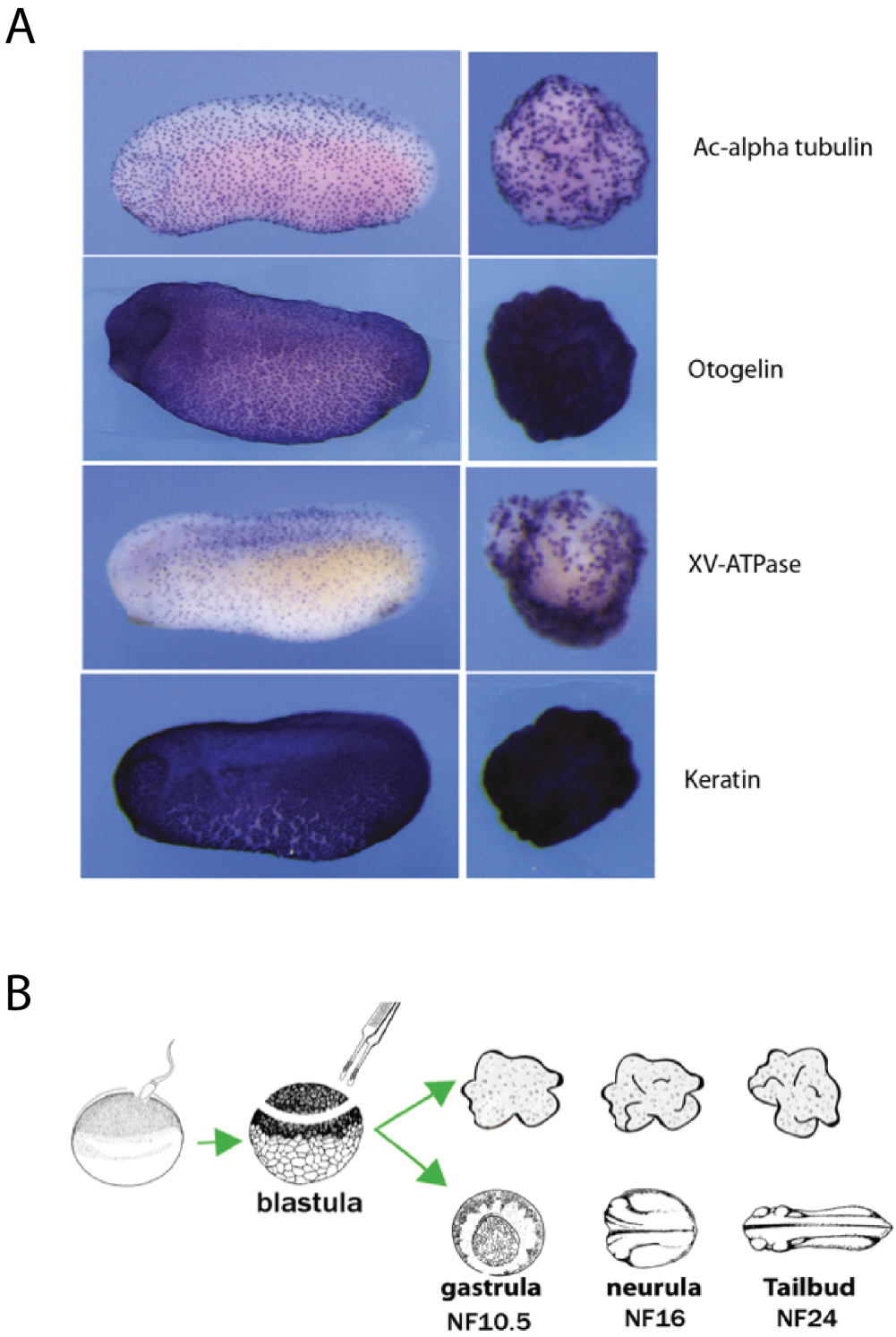


Figure 1: acetylated alpha-tubulin ICC and otogelin, xv-atpase and Keratin in situs on NF24 embryos and ACs derived from X. *tropicalis.* In the left column are embryos at nf24 (tailbud) stage, on the right are the animal caps. Fig 2B: X. *tropicalis* ACs were cut at the blastula stage NF9 and harvested when sibling embryos reached the gastrula (NF10.5) neurula (NF16) and tailbud stage (NF24).

The epidermal character of the ACs was also assessed by looking at non-epidermal genes. The expression of selected potential pluripotency associated genes (Fig. 2A), mesendodermal genes (Fig. 2B) and ectodermal genes (Fig. 2C) was plotted at the three timepoints. We observed a decay in the expression levels of the former two groups, coupled to a strong up-regulation of ectoderm specific transcripts. Overall, these data indicate that our AC explants reliably describe the differentiation of the larval epidermis from germ layer determination to a differentiated epithelium.

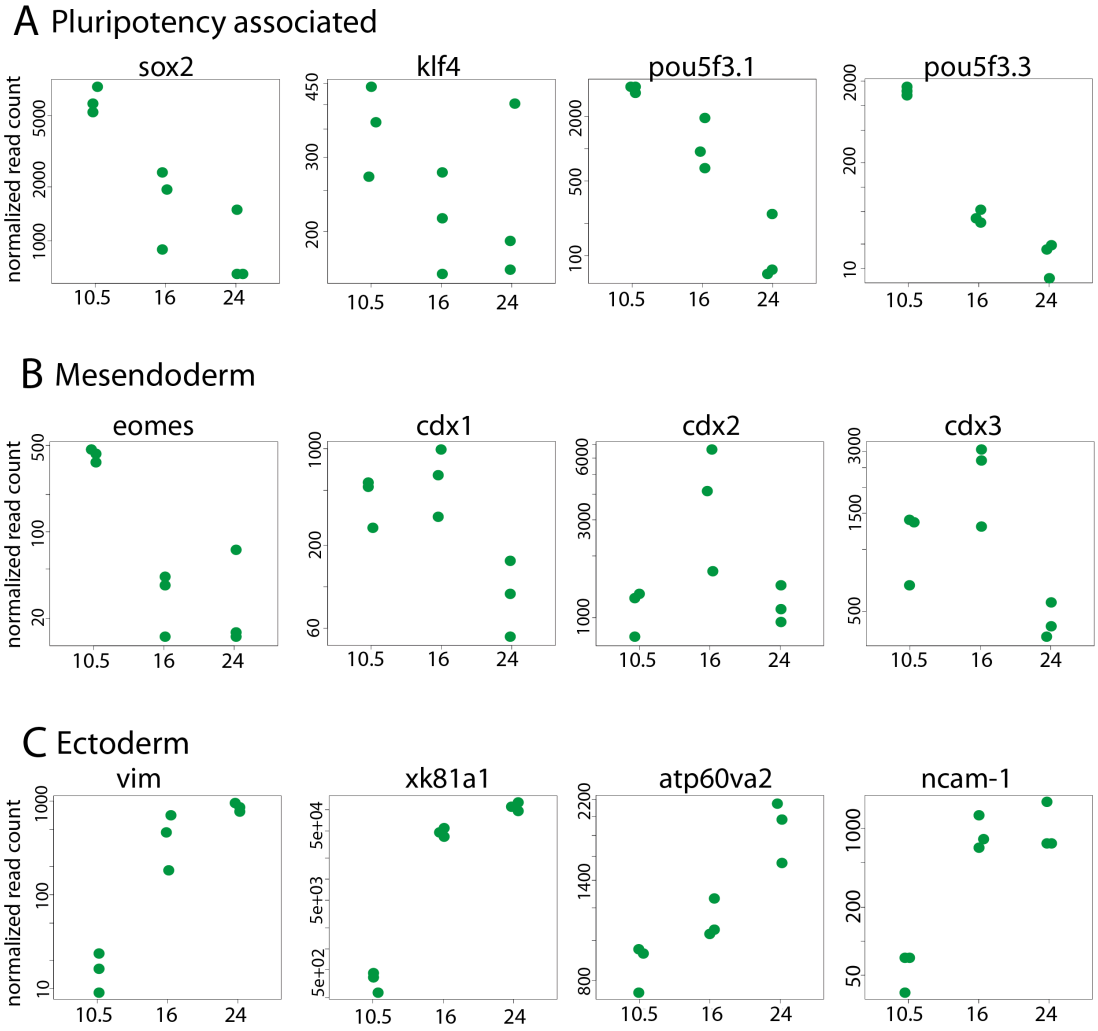


Figure 2: Expression of germ-layer specific genes at each given timepoint. Each circle represents a single sample. The Y-axis represents the normalized read

count while on the X-axis the single timepoints are represented. A)Pluripotency-associated genes sox2, klf4, pou5f3.1 and pou5f3.3. B) Mesoderm associated genes eomes, cdx1, cdx2 and cdx3. C) Ectoderm associated genes vim, xk81a1, atp60va2 and ncam-1.

3.2 Vast changes in gene expression accompany the differentiation of ACs

To obtain a first impression on the quantitative and qualitative RNA features of the isolated non-neural ectoderm we retrieved from the literature RNA-seq tracks generated from the whole embryo (Owens, Blitz et al. 2016) for the same stages we analyzed. We then compared the transcriptome of the animal caps to that of the whole embryo at the neurula stage (Figure 3A, Supplementary table 1)

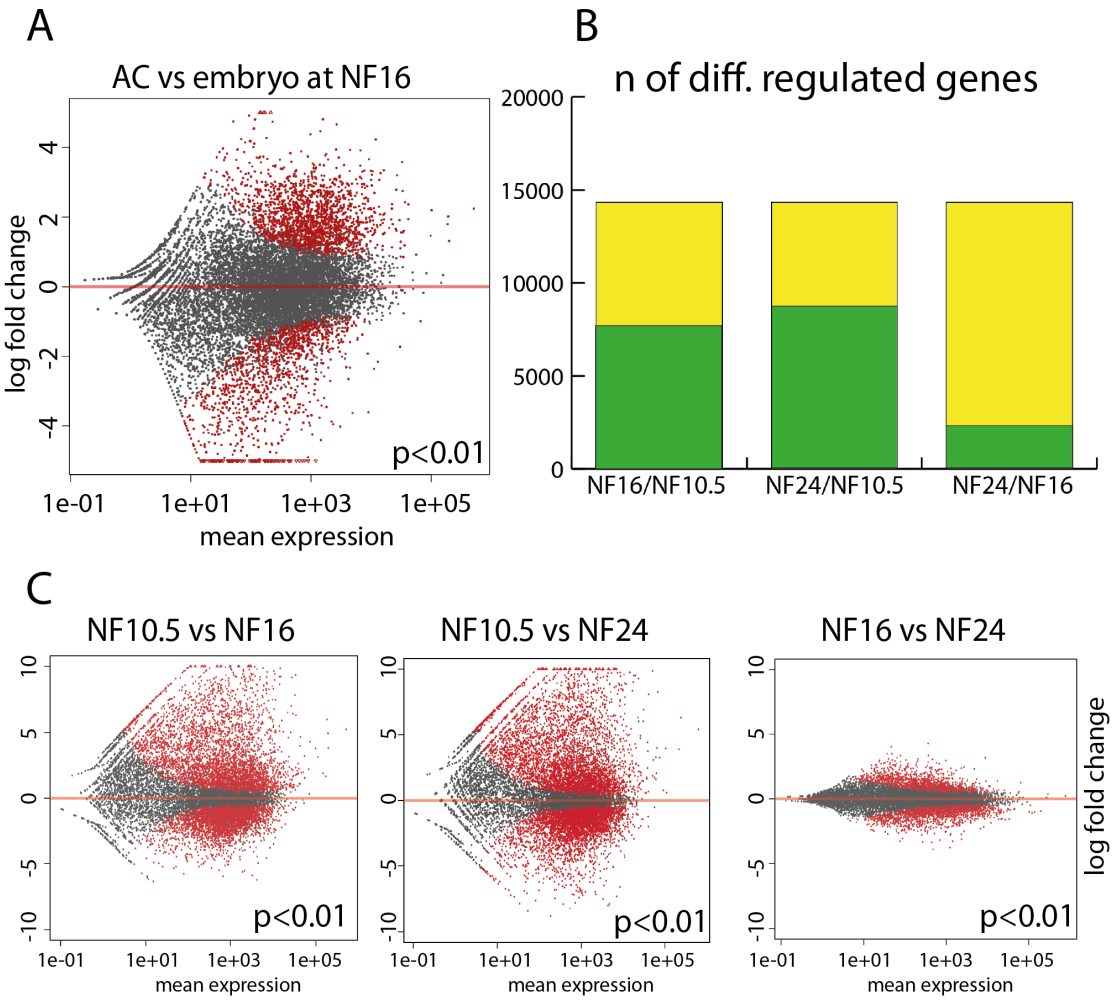


Figure 3: Gene expression in the ACs. A) Gene expression in the animal caps normalized to the whole embryo at the neurula stage (NF16). Gene expression was obtained as normalized read count (number of reads mapping to the locus adjusted for the size of the library) and normalized to the expression in the whole embryo. Each dot represents a gene. In red genes that are significantly differentially regulated with a p-value threshold of p<0.01. The Y-axis represents the log2 fold change while the X-axis the mean expression. B) Plotting the extent of differential gene expression in AC. Yellow: total number of genes detected in the RNAseq. Green: genes that are differentially regulated between the given developmental stages (p<0.01). The maximal number is defined by the annotated gene models present in XenTro9.1. (approx 15 thousand) C) Scattered plots representing pair comparisons between the analyzed stages. Each dot represents a gene, in red genes that are significantly differentially regulated with a p<0.01.

If on one hand, genes from non-ectodermal lineages are not expressed in the animal caps and cluster in the lowest part of the scattered plot (GO, data not shown). On the other, all the genes that up-regulated in the ACs are expressed in the embryo as well, reflecting the fact that the embryo is composed as well of epidermis. This is also reflected by the smaller extent of the up-regulation of these genes.

Then we proceeded comparing the gene expression in ACs of different age (Figure 3B, Supplementary table 2). To our great surprise we found that the majority of the mRNAs is differentially regulated between the gastrula (NF10.5) and the neurula stage (NF16), with 7224 genes being differentially regulated with a p – value cutoff smaller than 0.01, while a much smaller difference was observed between the neurula and the tailbud stage (NF24). The scatter plots in Figure 3C show the distribution of the differentially regulated genes and can be used to evaluate the extent of the differential expression. From these observations we concluded that the majority of the changes in gene expression in the developing epidermis are happening between NF10.5 and NF16. This is documented by the fact that more than half of the annotated genes are differentially regulated between these two developmental stages, while only a minor portion of the transcriptome is differentially regulated between NF16 and NF24.

Last, we intended to have an overlook at cell-type specific genes. To our advantage, the laboratory of Marc Kirschner has recently published a detailed catalogue of X. *tropicalis* cell states derived from single cell RNA-seq (Briggs, Weinreb et al. 2018). They identify marker genes and potential fate regulators associated with every and

each epidermal cell type. We filtered our datasets for genes that are up-regulated in the ACs compared to the embryo and that are differentially regulated during the time course. In this way we have been able to retrieve 91% (147/162) of the potential fate regulators (Supplementary table 3, section Briggs1) of epidermal cell types and an identical percentage of the potential cell markers (352/390) (Supplementary table 3, section Briggs2) belonging to the 4 cells types described above. Finally, Quigley and colleagues (Quigley and Kintner 2017) have published a list of core multiciliogenesis associated genes. The list consists of approximately 800 genes that have been linked to ciliogenesis by a series of gain of function and loss of function experiment. Our analysis has revealed that 56% of the genes on the list is indeed positively regulated in the course of ACs´ differentiation (supplementary Table 3, section Quigley). Another 28% of the genes listed by the authors is expressed in ACs but their expression is not regulated. Last 16% of the genes on the list are either not expressed or not detected in our explants.

3.3 Alternative exon usage shapes the ACs' transcriptome

The massive differences in gene expression, which we derived from pairwise comparisons of ACs along the process of differentiation, could be fueled by de novo gene activations and by differential exon usage. To examine the latter possibility in more detail, we compared our read database to already defined splicing isoform-models contained in the "NCBI Xenopus *tropicalis* Annotation Release 103". First we compared the expression of splicing isoforms in the ACs with the whole embryo at NF16 (Figure 4A, Supplementary table 4).

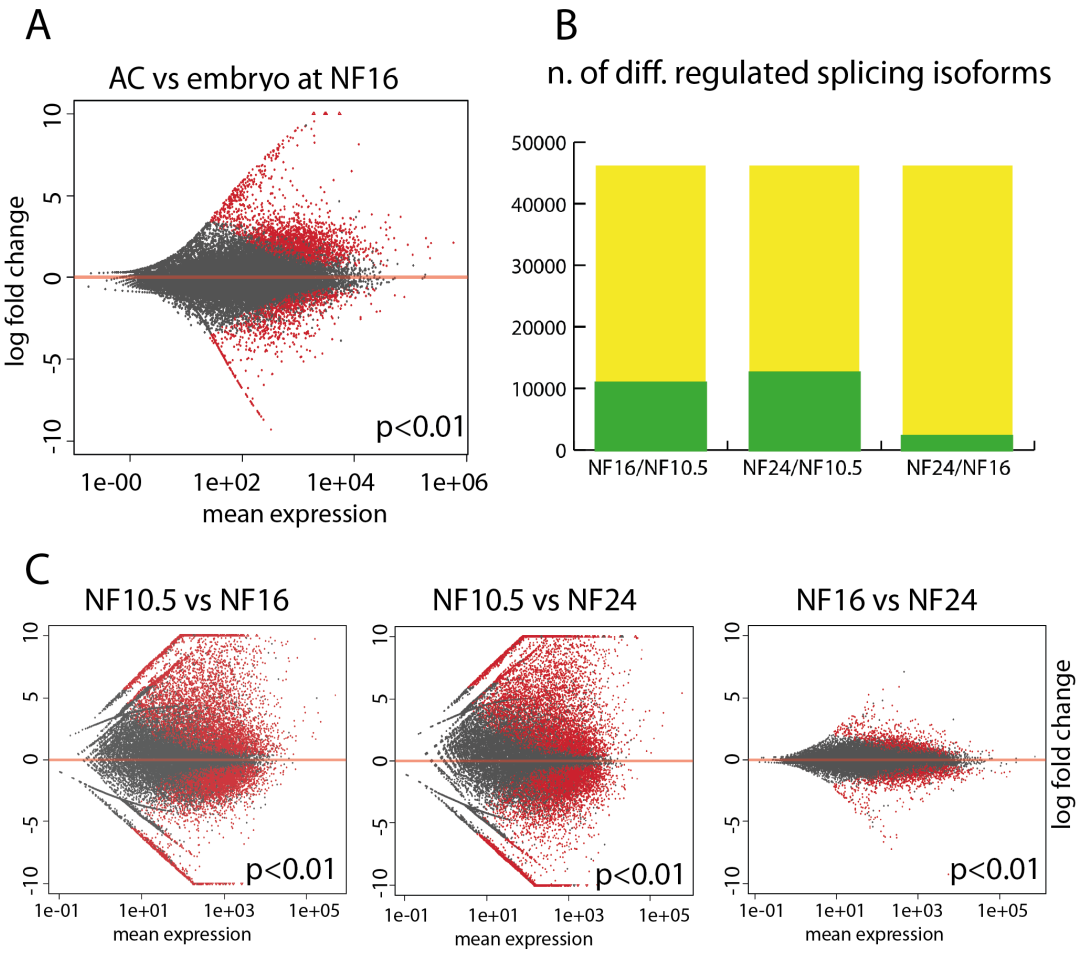


Figure 4: Differential exon usage during ACs´ differentiation. A) Splice isoform-expression in the ACs compared to the whole embryo at NF16. Each dot represents a splicing isoform. Each red dot is a splicing isoform whose expression is differentially regulated between embryo and AC with a p value cutoff of p<0.01. The Y-axis represents the log2 fold change while the X-axis the mean expression. Figure 4B: Numbers of splicing isoforms that is differentially regulated between 2 given stages: in yellow the total repertoire of splicing isoforms models, in green those whose expression differs in the analyzed pair with a p-value cutoff <0.01 C) Scatter plots representing pair comparisons between the analyzed stages. Each dot represents a splicing isoform; red dots - splicing isoforms that are differentially regulated (p<0.01).

We found out that 3168 splicing isoforms are differentially regulated between the embryo and the ACs with a p-value cutoff of $p<0.01$. Interestingly the extent of the differential regulation of the splicing isoforms is extremely wide reaching in few cases a log2 fold change of 10 despite the fact the whole embryo is composed also of epidermis. This suggests that in these instances genes that are widely or strongly expressed in the embryo are transcribed in the ACs as epidermis-specific isoforms, carrying or lacking epidermis-specific exons. We then wondered how pervasive the differential splicing is between the different developmental stages we analyzed.

Unexpectedly 11054 splicing isoforms are differentially regulated between the gastrula (NF10.5) and the neurula stage (NF16) when applying a cutoff of $p<0.01$. Moreover, the extent of their differential regulation is impressively wide spanning from a minimum of -13.5 to a maximum of 15.6 log2 fold change. Similar numbers are observed when comparing the gastrula (NF10.5) to the tailbud stage (NF24). However, similarly to what we have seen for gene expression, only minor differences are observed between the neurula (NF16) and the tailbud stage (NF24) (Figure 4 B-C, Supplementary table 5). Provided that alternative splicing affects coding exons in the mRNAs, this observation may have a strong impact on the ectoderm-specific proteome.

To address this hypothesis systematically, one would need a comprehensive database of the Xenopus proteome. Unfortunately, the available Xenopus protein annotation on Uniprot, wwPDB or Xenbase are too limited to be used for this purpose. As second best approach, we matched each splicing isoform to its gene of origin. For

the comparison between the gastrula ACs (NF10.5) and the neurula (NF16) stages we retrieved 11054 differentially expressed splice isoforms, which match to 8467 different genes. Following the existing annotation we derived that 1859 genes are not protein coding, 2819 code for at least 1 protein while 3789 code for at least 2 protein isoforms. Numbers are similar for the other comparisons. This tells us that pervasive alternative splicing accompanies the massive switch in gene expression we observed and it models the transcriptome of the prospective epidermis.

In some cases, although the total normalized read numbers for a given gene do not change much, exon usage is very different between the two samples. This means that the embryo and the ACs despite expressing the same genes at similar level produce RNAs with different exonic composition. In line with it, comparing splicing isoforms expression in the ACs with the embryo allowed us to identify embryo and AC-specific splicing isoforms. We generated a list of genes that posses a splicing isoform that is prevalent in the ACs and another one that is prevalent in the embryo (Supplementary table 6). As proof-of-case, we analysed the two genes of this list that exhibit the maximal variance (*evpl* and *cpne1*) and another gene (*bmp7.1*) that was manually picked. Each one of these genes exhibits particular splicing dynamics. The first case is *evpl,* a gene involved in keratinocytes maturation. It possesses two splicing isoforms, one of which is dominant in the ACs and another one that is prevalent in the embryo. The second case is well documented by *cpne1* a calcium-dependent phospholipid binding protein. This gene has many variants that are co-expressed but only one being prevalent either in

the embryo or in the ACs. We then decided to validate this data via RNA *in situ* (*in situ*) hybridization. For this reason, we cloned splice variant-specific exons for the three genes mentioned above. For *evpl* we cloned a single exon that is expressed only in the ACs, while for *bmp7.2* and *cpne1* we cloned 2 exons each, one specifically expressed in the embryo and the other in the ACs. We *in vitro* transcribed the constructs to produce digoxigenin-labeled RNA probes and generated embryos and ACs for the RNA *in situ.* If our calculations are valid one would to see the ACs specific probes to give staining only in the ACs and vice versa for the embryo.

As expected the skin specific probes give staining both in the AC explants and on the embryonic skin, while the embryo specific probes did not stain the ACs.

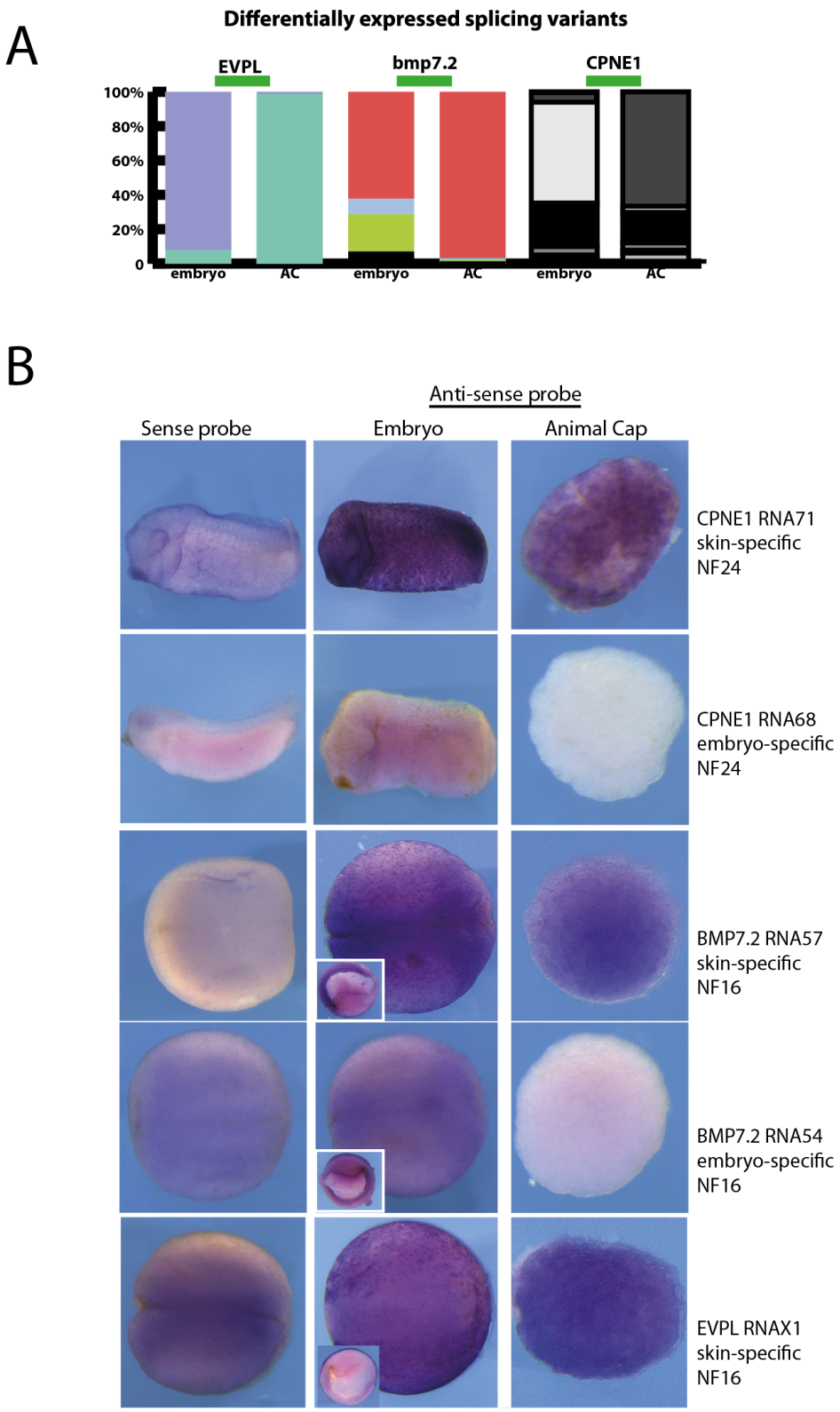


Figure 5: Spatial expression of embryo or AC-specific splicing isoforms. A) Representation of the number and expression level of splicing isoforms in embryo and ACs for the indicated genes. B) RNA *in situ hybridization* on ACs or embryos at the indicated developmental stage hybridized with either the embryo

specific probes (CPNE1 RNA0068, BMP7.2 RNA13254) or the skin specific ones (CPNE1 RNA0071, BMP7.2 RNA13257 and EVPL X1). Inserts represent sagittal sections. Hybridizations have been conducted in parallel for each row.

3.4 Rpetitive DNA elements are dynamically expressed in ACs

Another intriguing class of transcripts expressed during early embryogenesis is repetitive DNA elements (REs). REs are non-coding regions of the genome. Among the REs classes transposons and transposons-derived sequences are the most abundant family, constituting more than a third of the Xenopus genome. These elements are in Xenopus, as well as in mammals, decorated with repressive histone modifications such as H3K9me3 and H4K20me3. Their expression is thus silenced, except during early embryonic development. The function of these transcripts is unclear. Transcription of repetitive loci during embryogenesis could just reflect the dynamics of chromatin maturation, given the absence of repressive chromatin marks in early embryos. However, these observations collide with the fact that historically transposons expression has been considered to be deleterious as provoking mutagenic insertional events. This contradiction lead, in recent years, several authors to hypothesize that REs and transposon in particular may play a role in normal embryogenesis. If their expression is not transcriptional noise, e.g. resulting from the lack of chromatin-induced repression, but functional to normal embryonic development, then it should be subjected to regulation. In this scenario one would expect REs expression to differ between the embryo and a pure isolated tissue as the ACs. In order to validate this

hypothesis we compared REs expression between the ACs and the embryo at the gastrula stage (NF10.5). 724 REs are currently annotated for X. tropicalis on the REs database RepBase. We found out that 137 of them are differentially expressed between the embryo and the ACs with a cutoff of $p<0.01$ (Figure 6A, Supplementary table 8). Intriguingly, in the comparison some REs have a log2 fold change expression lower than -4 or bigger than 4. We interpreted this information as evidence of a differential regulation of REs transcription in different embryonic tissues. This observation suggests that REs expression is regulated and thus could possibly be functional to normal development. Our analysis continued by looking at differences in expression between the different developmental stages. We found that more than a fourth of the total repertoire of repetitive DNA elements is differentially regulated between the gastrula and the neurula stage, reaching 261 elements when comparing the gastrula to the tailbud. Notably, differences between the neurula and the tailbud stage are negligible (Figure 6 B, Supplementary table 7). Scatter plots of pairwise comparisons (Figure 6C) reveal that the biggest variance in expression belongs to repetitive elements that are active at NF10.5 and become silenced at later stages.

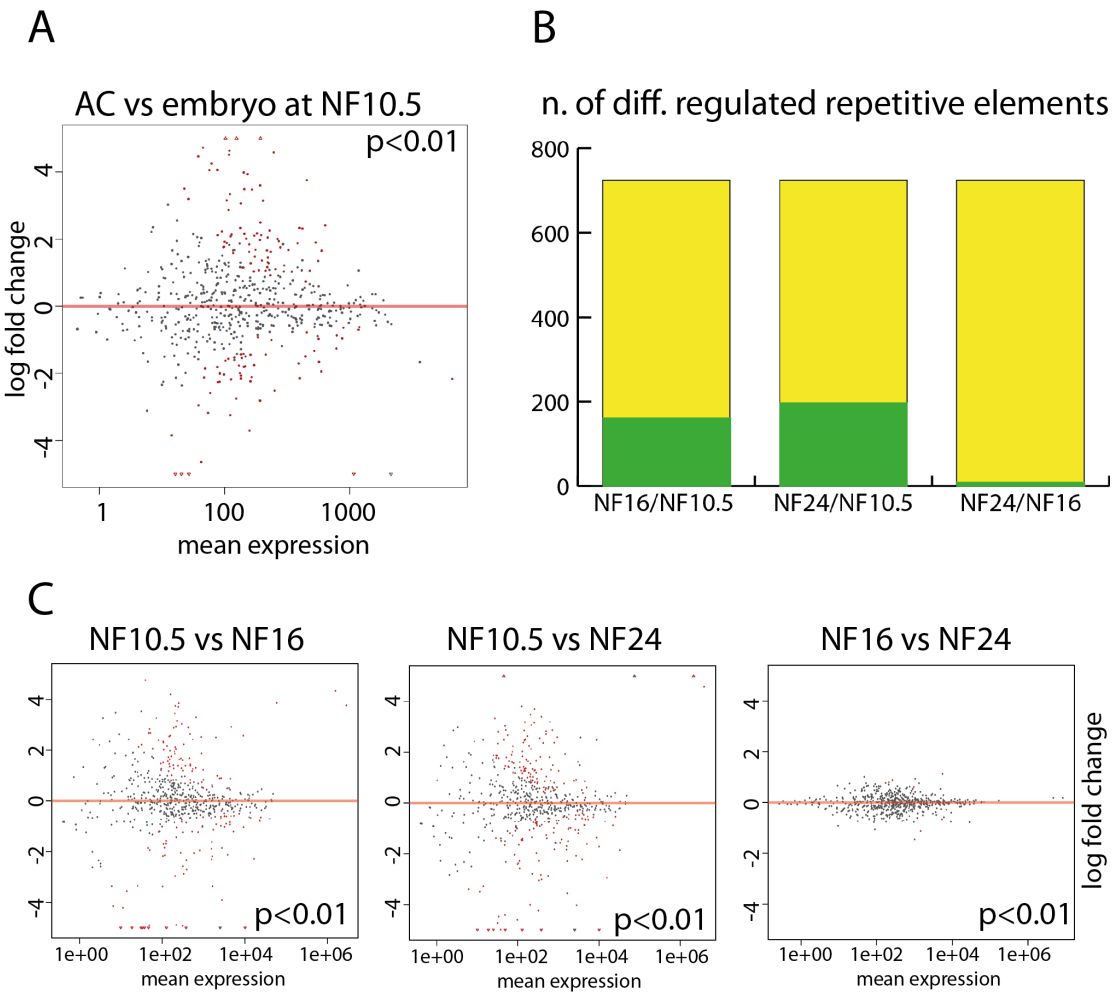


Figure 6: REs expression in the animal caps explants. A) Differential expression of Repetitive DNA elements between ACs and embryo at early gastrula (NF10.5). Each dot represents a repetitive DNA element. In red REs whose expression is differentially regulated between embryo and AC with a p value cutoff of p<0.01. The Y-axis represents the log2 fold change of normalized read, while the X-axis details the mean expression. 6B) Numbers of repetitive DNA elements that is differentially regulated between 2 given stages: in yellow the total repertoire of annotated repetitive DNA elements, in green those whose expression differs in the analyzed pair with a p-value cutoff <0.01. C) Scatter plots representing pair comparisons between the analyzed stages. Each dot represents a RE, in red REs that are differentially regulated with a p<0.01.

These observations imply that RE transcription is non-uniformly regulated within the embryo, a conclusion that can be tested by RNA *in situ* hybridization. We selected few elements among those whose expression is high in the ACs at NF10.5 but gets silenced by NF24. We chose 3 elements belonging to specific subclasses of repeats. ERV1-4_LTR_Xt is an endogenous retrovirus with long-terminal repeats. LTRX1_LTR is a standard long-terminal repeat retrotransposon. Last REM2b is the transcript arising from the centromeric satellite repeats. We cloned the cDNA of each one of these transcripts and performed *in situ* analysis.

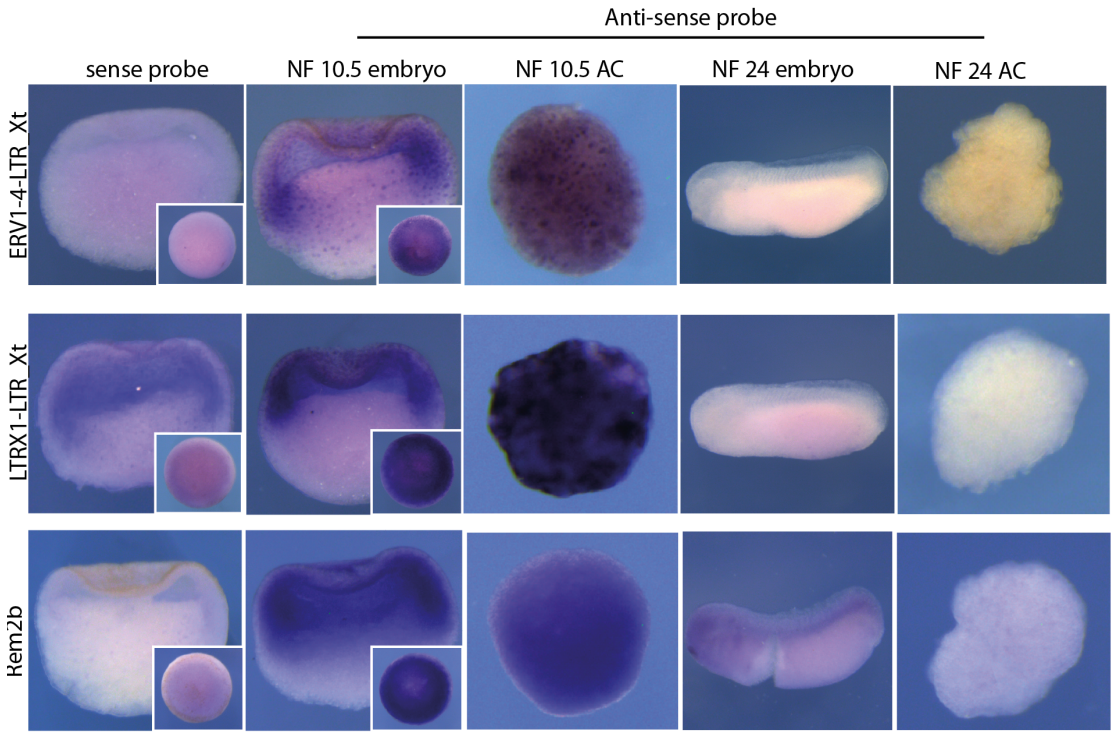


Figure 7: In situ analysis of repetitive DNA elements expression. The left column shows the sense control where the probe as been transcribed in sense orientation. Inserts show sagittal sections. Signals in this column represent background staining or antisense transcripts arising from the RE element. In the remaining columns, embryos were hybridized with RNA probes complementary to the REs transcripts. In the second column are embryos at the gastrula stage.

Inserts show embryos that were bisected before the hybridization. In the third column are the ACs at the gastrula stage. In the fourth column are the embryos at the tailbud stage and in the last column the ACs at the tailbud stage.

The experiment reveals that all three elements are expressed at the gastrula stage both in embryos and in the ACs. The sense probe control suggests that their transcription is mainly unidirectional. Staining on bisected embryos reveals a sub-regionalization of expression. Rem2B is expressed in the animal pole and in the equatorial region of the embryo in a uniform manner, while it is not expressed in the outer cell layer and in the vegetal pole area. LTRX1 probe stains predominantly the involuting mesoderm (represented by the deep cell layers of the equatorial region) and the superficial ectodermal cells in a salt and pepper manner. Last, ERV1-4-LTR_Xt probe stains all three germ layers of the embryo in a punctuated manner but it highlights predominantly the involuting mesoderm. By the tailbud stage, RE transcripts are not detectable anymore both in embryos as well as in ACs, except for Rem2B, whose expression can still be appreciated, though at low levels, in the neural tube of the tailbud embryo.

It appears clear that the expression of these REs is regulated not only in a temporal manner, but also in a spatial one. Our results strongly indicate that different REs families are subjected to specific transcriptional regulation, supporting the emerging view of a functional role for REs in normal embryonic development.

3.5 Inter and intra group comparison of transcripts' families

Can the transcriptional state of the ACs be predictive of their developmental state? Is the transcription of the three classes of transcripts we analyzed so far (genes, splicing isoforms and REs) tuned in a similar manner throughout development? We attempted to answer to these issues by carrying out a principal component analysis (PCA) for each class of transcripts we previously detailed. The PCA unbiasedly positions single samples (or biological replicates) along the 2 principal components that can explain the majority of the variance between measured parameters. The results of our analysis are shown in Figure 8. For what regards gene expression the PCA is able to tell the three developmental stages apart. In fact, the single biological replicates of each time point cluster together and the different time points are well resolved. This is not the case for the PCA for splicing isoforms and repetitive elements. In these two cases in fact, the PCA is able to resolve only the gastrula (NF10.5) stage from the later two. Samples from NF16 and NF24 for these 2 classes of transcripts cannot be told apart and tend to overlap along the same coordinates. This result suggests that gene expression can be predictive of the developmental stage of the ACs explants. Moreover, it indicates that gene expression is more finely tuned that splicing isoforms and REs expression.

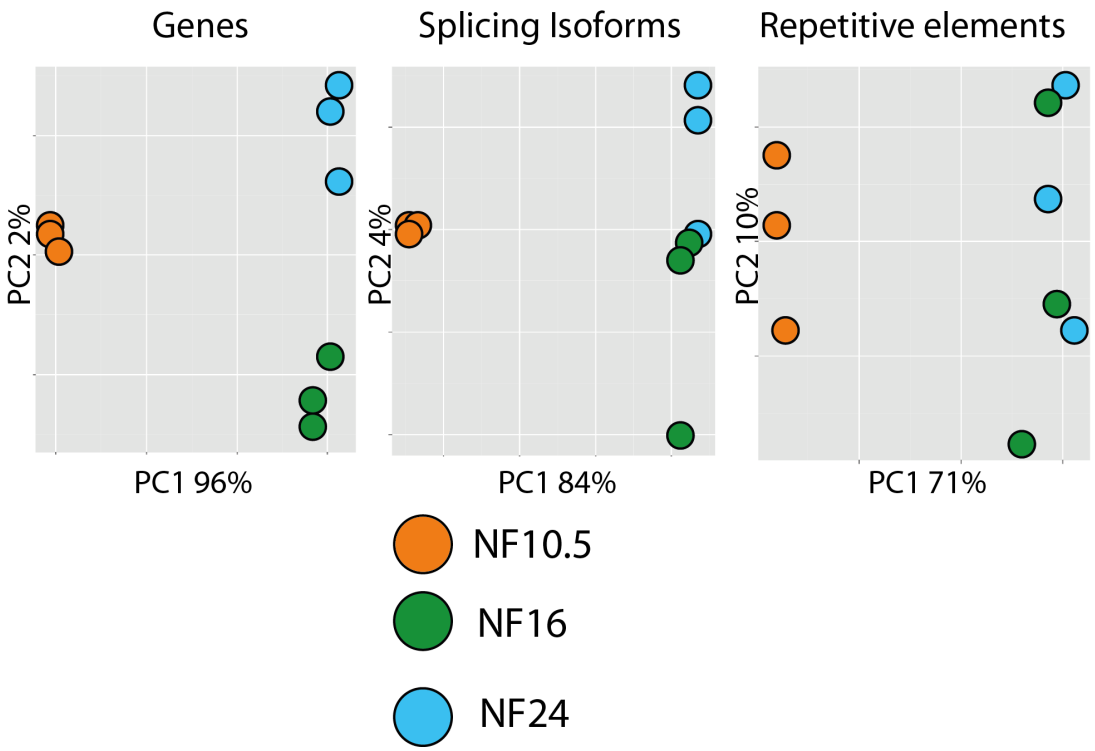


Figure 8: Principal component analysis of the three classes of transcript analyzed in the study. Each dot identifies a different biological replicate. Different colors identify different developmental stage. On the Y and X-axis, the two principal components that can explain the maximum of the variance.

3.6 Animal caps express exonic circular RNAs

The last class of transcripts we analyzed is circular RNAs (circRNAs). CircRNAs are covalently closed RNA molecules arising via back-splicing of pre-mRNAs (Huang, Yang et al. 2017). The mechanisms regulating their productions are still poorly understood. The same is true for their functions: some authors have proposed they may function as miRNA sponges; some others have described circRNAs that are translated (Legnini, Di Timoteo et al. 2017, Zlotorynski 2018). In Xenopus authors have identified the presence of circRNAs in the egg cytoplasm. However, these transcripts were intronic lariats produced by normal splicing events (Talhouarne and Gall 2014).

CircRNAs of exonic origin have been described in Xenopus adults. Moreover, authors have shown that their expression can be mis-regulated by chronic atrazine exposure (Sai, Li et al. 2018). Our interest for novel regulators of epidermis differentiation lead us to wonder whether circRNAs of exonic origin are expressed in the developing ACs. In order to answer to this question we used the method described by Westholm et al, (Westholm, Miura et al. 2014). First we identified all the circular junctions in our data sets. For example, if exon 7 is at the 5' and is followed by exon 1 for a given gene that can be called a circular junction. If the circular junction is flanked by a GT-AG splicing site that junction can be attributed to a circularRNA. One example is shown for 1 single biological replicate of NF16 (Figure 9A). The algorithm identifies 4770 circular junctions in this dataset. 382 of them are flanked by a GT-AG splicing site, thus originated via a back-splicing event. 314 align to a gene model in the X. tropicalis gene annotation 9.0. This method is scarcely sensitive. In fact, to be identified reads must span the circular junction and satisfy the requirement of having at lest 15bp on each side of the junction. As a result, each circRNA we identified is represented by a small number of reads. On average approx. 10 circRNAs per sample have more than 5 reads, the majority being identified by just 1 read. The element with the highest number of reads in the 9 datasets has 48. On average we identified 250 circRNAs per developmental stage (Figure 9B, Supplementary table 9). Some elements are constantly expressed throughout the developmental stages and seem to accumulate constantly (Figure 9C). To validate the physical existence of these elements we chose 3 among those with the highest read number for the three developmental stages (circ27 for NF10.5, circ1

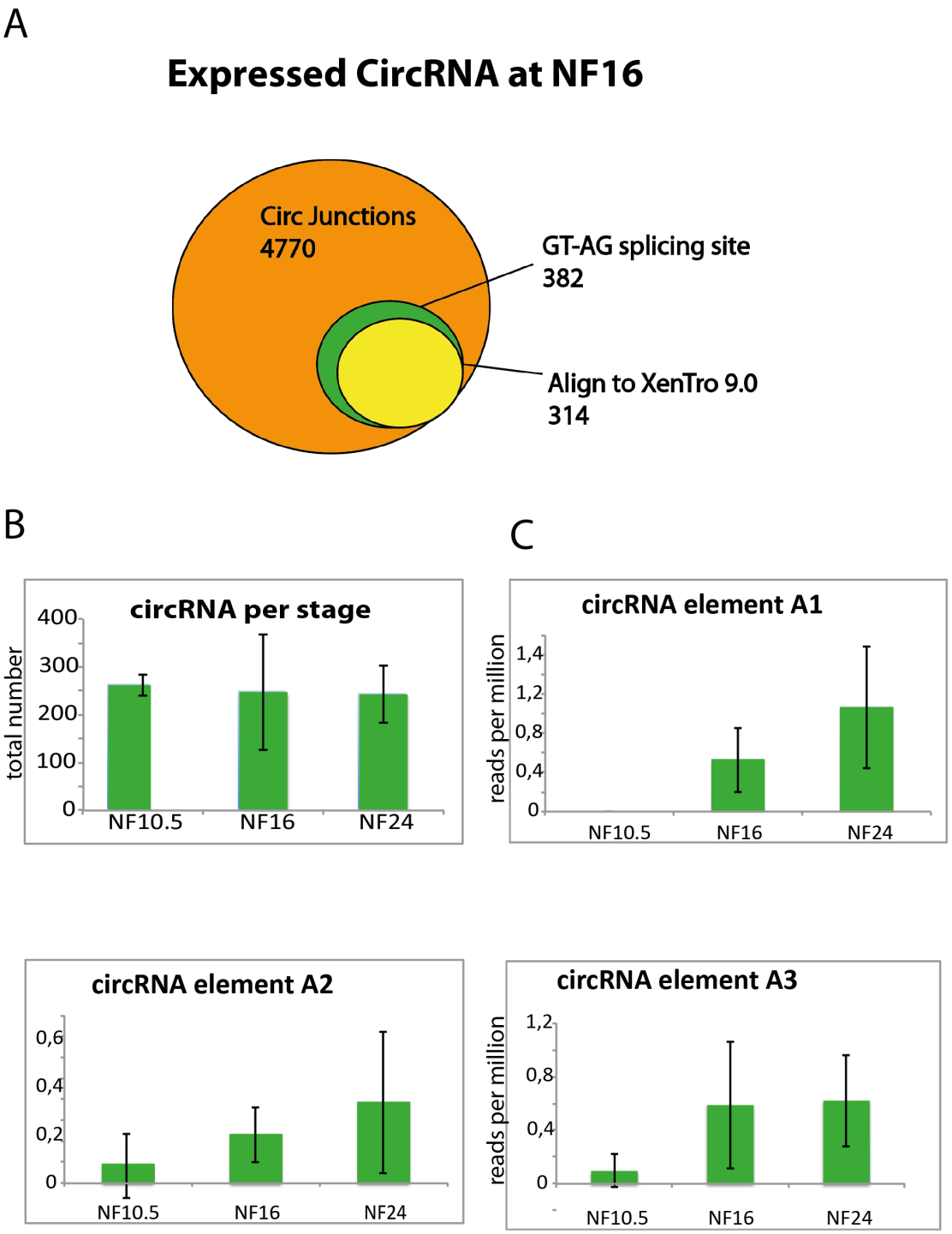


Figure 9: Expression of circRNAs in ACs explants. A) An example of filtering of the circular junctions for 1 single biological replicate at NF16. B) Mean value of the total number of circRNAs identified in each developmental stage. Error bars represent standard deviations. C) Relative abundance of 3 of the most expressed circularRNAs.

for NF16 and circ6 for NF24) and performed end-point PCR on them. We designed primers upstream and downstream of the junction so that forward primer is on exon 7 while the reverse primer is on exon 1. This couple of primers should amplify across the circular junction and give amplicons of short size (circ1= 188 bp, circ6= 100bp, circ27=93 bp). As a control we designed a primer pair amplifying in the opposite direction so that the forward primer is on exon 1 and the reverse on exon 7 (Figure 10A). These primers should not span the junction and in case they managed to amplify such long product, they would give very long amplicons. The primers spanning the junctions have given amplicons of exactly the expected size, while the control primers gave either very long bands or multiple bands (possibly indicative of alternative splicing) (Figure 10B). The amplicons for circ1 and circ6 have been successfully sequenced and overlap with the junction and with the flanking exons. That is the first time that exonic circular RNAs are identified and tracked in Xenopus and in general in a pure isolated epidermal tissue.

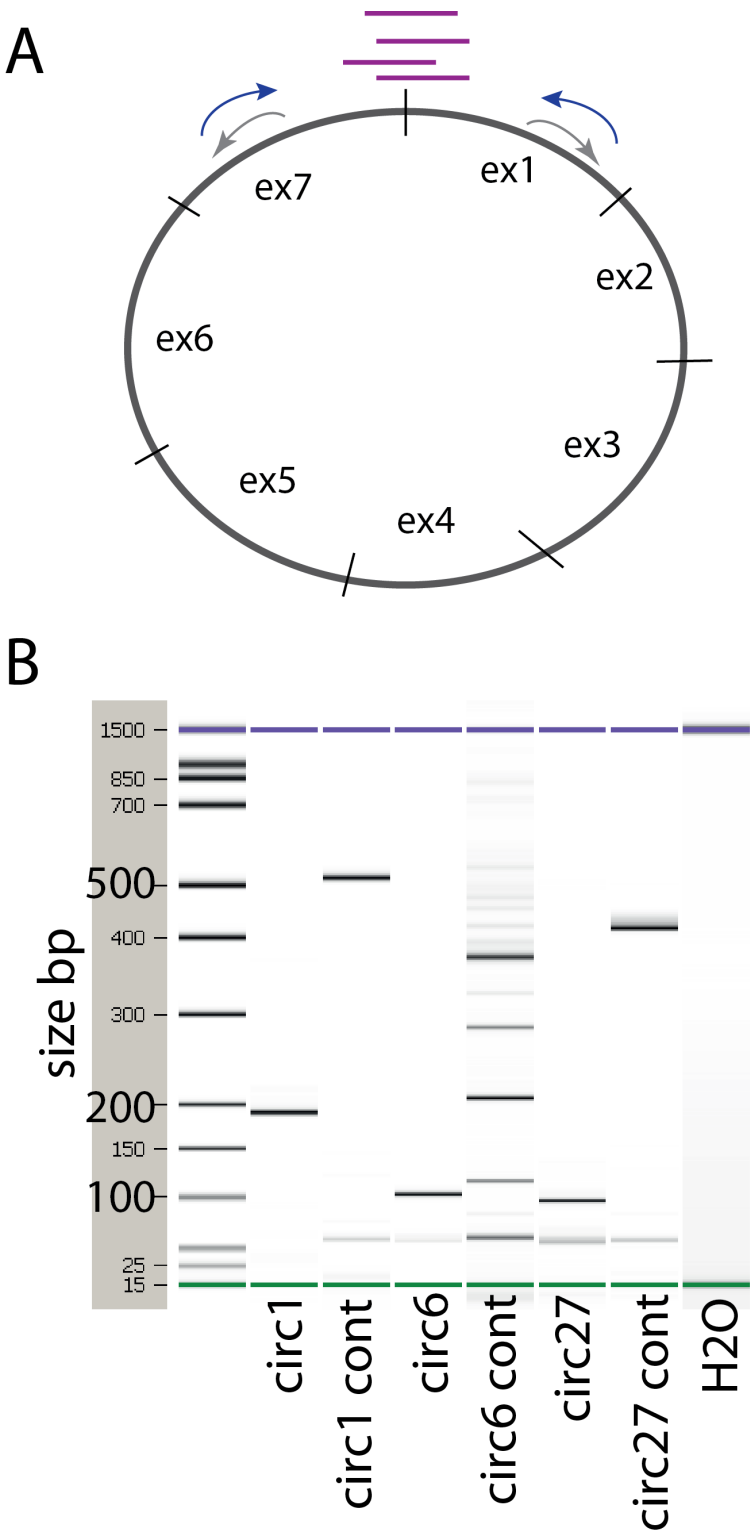


Figure 10: circRNAs validation. A) A graphical sketch that illustrates a model of circRNA. In magenta are shown the reads spanning the circular junction that can be uniquely attributed to a circRNA. In blue are shown diagnostic primers that produce amplicons spanning the circular junction. The control primers are shown in grey. B) Bioanalyzer analysis of the circular amplicons. 3 circRNAs have

been amplified (circ1, circ6, circ27) and they have been run in parallel with their negative controls (label with “cont”) were primers were designed in the opposite orientation.

3.7 Modeling the impact of transcription factors

Transcription factors (TFs) directly control gene expression. Historically, the majority of the best-understood TFs have been identified in forward genetic screenings, where a mutation was connected to a function. Later on, reverse genetics approaches have allowed authors to expand the repertoire of known TFs identifying entire families via sequence similarity. However, just a minor portion of the overall TFs in Xenopus has been functionally tested. Our initial aim was to investigate which TFs direct the transcriptional program of ACs. At the same time we wanted to avoid slipping into the bias of giving more importance to functionally characterized factors. In order to do so we took advantage of the Mogrify algorithm described by Rackham et al (29). The algorithm has been developed and used to predict factors required for direct reprogramming. This method is not perfect, but it is the best we have to do it so far. The algorithm integrates the STRING interaction network of each given TF expressed in the datasets together with its expression dynamics of expression in the pairwise comparisons between developmental stages. In this way TFs are ranked based on the measured change in their sphere of influence. For each comparison the algorithm produces two lists of transcription factors that are predicted to control the transcriptome of either the earlier (neg_rank) or the later (pos_rank) stage of the comparison (Supplementary table 10).

A graphical representation of the result for the comparison between NF10.5 and NF24 is shown in Figure 11. In this case we analyzed the top 10% of factors for each list (total number of ranked TFs oscillates between 40 and 250 among all comparisons).

Figure 11A shows us that the overall transcriptome of NF10.5 to be largely controlled by TFs that are associated with pluripotency and the early phases of embryonic development. In the shown comparison, on top of the list (#1 and #3 where #=position) the algorithm places *zic1* and *zic2* TFs involved in the establishment of the left/right asymmetry and regulation of pluripotent stem cells. In position #6 dmtf1 involved cellular proliferation, in #7 it places *sp5* a TF involved in embryonic patterning. Moreover, both homeologs of *sox17b* make it to the top of the list in positions #10 and #13. On the other hand the transcriptome of NF24 seems to be dominated by lineage specific TFs (Figure 11B). On top of the list #1 we fund *eth* (epithelium-specific Ets transcription factor) a TF involved in epithelia differentiation, in #2 hoxa1 a gene involved in differentiation and morphogenesis with a extremely high expression in the skin, in #3 we found *tp63* the major regulator of multiciliated cells differentiation and in #6 *fos*. *Fos* clusters in #2 together with rel in #3 for NF16 also when considering the comparison between NF10.5 and NF16. This is of interest because *fos* and *rel* are two of the 5 transcription factors used to transdifferentiate fibroblast into keratinocytes, the major components of the differentiated skin (Rackham, Firas et al. 2016). The method is for sure not perfect. In fact, for NF24 it puts on top of the list several TFs that are normally expressed in hematopoietic cell lineages. A closer look to these

transcription factors reveals that they are characterized by a very small number of reads and very low p-values, which have to be considered responsible for their high positioning on the list.

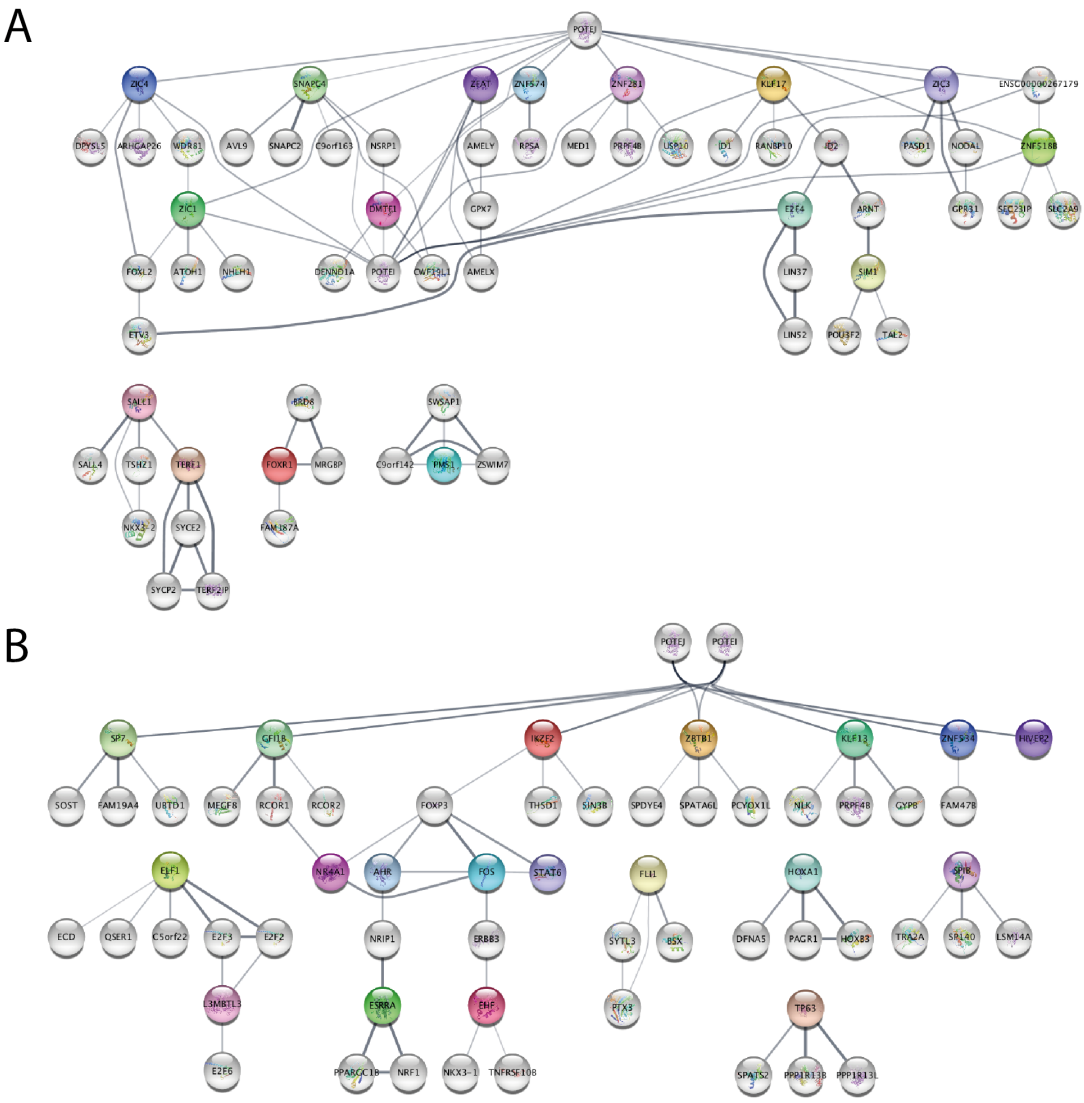


Figure 11: Transcription factor analysis for the comparison between NF10.5 and NF24. A) TFs that are specific for NF10.5, in B) those specific for NF24. In different colors the TFs identified by the algorithm, in grey their main interactors. The thickness of the connection lines represents the relative strength of the interaction.

3.8 Epigenetic regulation of epidermal cell types

The regulation of the ACs transcriptome is orchestrated at different levels. We already discussed about the impact of transcription factors on the overall tanscriptome. Each cell type on the embryonic epidermis can be identified by the expression of master TFs regulating its development. It is known that *foxj1* (forkhead box j1) and *mci* (multicilin) are indispensable regulators of multiciliated cells differentiation (Thomas, Morle et al. 2010, Stubbs, Vladar et al. 2012). Ectopic expression of *mci* in embryonic cells is sufficient to differentiate them into multiciliated cells (Stubbs, Vladar et al. 2012). *Foxa1* controls the development of the small secretory cells (SSC), *foxi1* directs the differentiation of the ionocytes as well as *grhl1-3* do in goblet cells (Esaki, Hoshijima et al. 2009, Gao, Vockley et al. 2013, Dubaissi, Rousseau et al. 2014). Moreover, the deep layers of the ectoderm contain a population of stem cells from which the epithelial cell types arise. The self-renewal features of these cells are transcriptionally controlled by tp63 (Briggs, Weinreb et al. 2018).

However, little is known about the epigenetic regulation of transcription in the prospective epidermis. Previous insights from our lab had shown that the Suv4-20h histone methyltransferases seem to specifically regulate the multiciliogenic differentiation program in the prospective epidermis without affecting the fate decisions of the other three epidermal cell types (Nicetto 2012, Hsam 2015). Thus, we identify in the Suv4-20h enzymes a possible key to further our understanding of how the animal caps´ transcriptome is orchestrated at the epigenetic level during development.

3.9 *Suv420h* write K20me3 which is mostly intergenic but regulates oct25 expression

Suv4-20h1 and Suv4-20h2 are responsible for the deposition of the di and tri-methylation marks on the lysine 20 of histone H4. To test the specificity of the 2 enzymes in laying the K20 marks we knocked down their expression singularly or in pair, in X. *laevis,* using published (Nicetto et al., 2013) antisense translational-blocking morpholino oligonucleotides (morpholinos). We injected 40ng per blastomere of each morpholino in both blastomeres at the 2-cell stage; we then collected the embryo, extracted the histones and performed western blot analysis of the methylation state of K20 (Figure 12A).

The KD of the enzymes alone or in couple leads to an accumulation of the K20me1 mark. That is coherent with a scenario were both proteins use K20me1 as a substrate. Interestingly the K20me3 is abolished only when both enzymes are knocked down in concert, suggesting a partially overlapping function of their methyl-transferase activity. We then assessed the same issue in another way: we cloned the X. *tropicalis* Suv4-20h enzymes and triple flagged them. To make sure the constructs were actually translated we transcribed synthetic mRNAs, injected them onto embryos and checked for their expression. The actual functionality of the 3 enzymes was demonstrated using antibodies against the 3XFlag-Tag by WB, immunocytochemistry and confocal analysis which confirmed that the 2 synthetic constructs are translated and transported into the nucleus where they associate with DAPI-dense foci (data not

shown). We then injected 300pg of each synthetic RNA alone or in couple and checked for K20 methylation states (Figure 12B). In this case the result is diametrically opposite to the KD. Both RNAs reduce the amount of mono-methylated K20 while increasing the di and tri-methylated states.

To summarize, we have seen that if on one hand we strongly deplete K20me3 just when both Suv4-20h enzymes are depleted on the other, the gain of function of both reduces K20me1. Thus, this western blot analysis strongly suggests that both enzymes can write the di and tri-methylation mark.

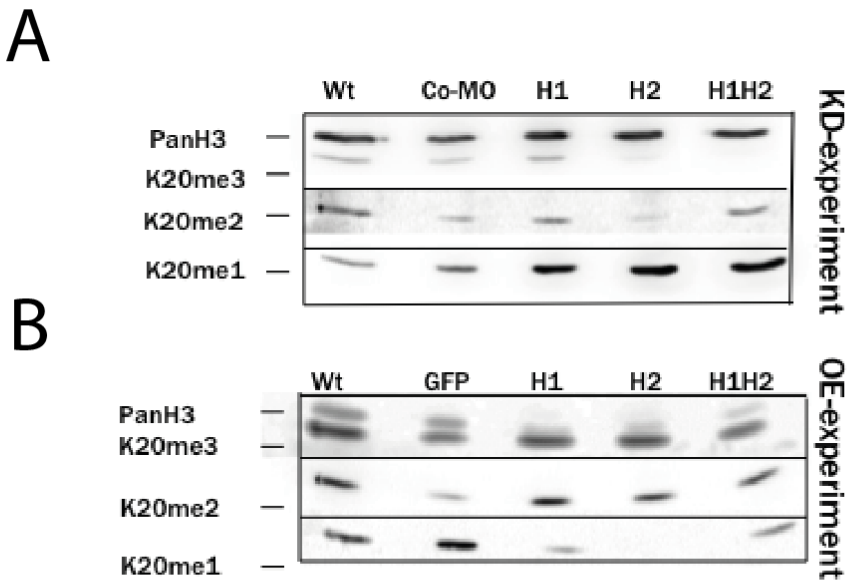
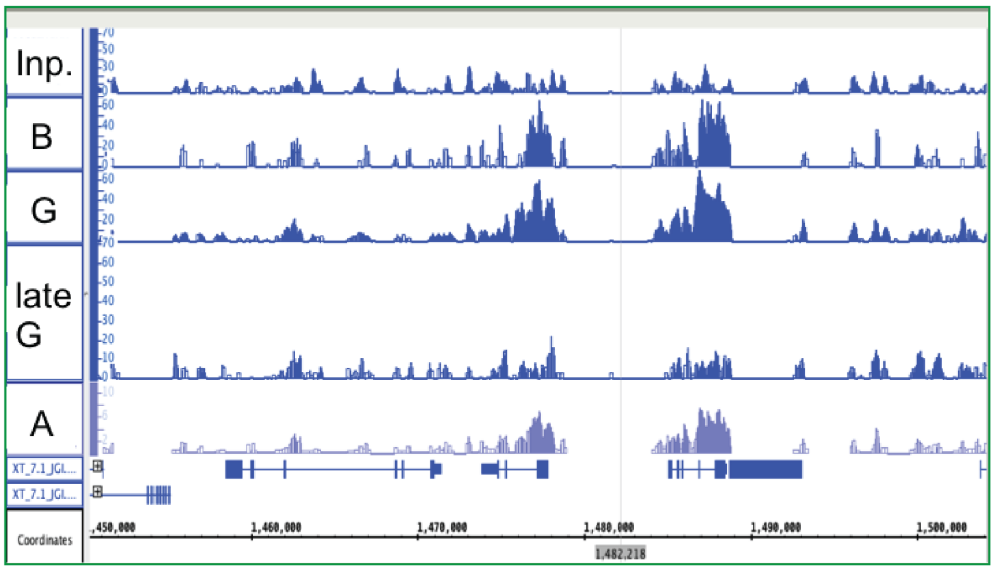


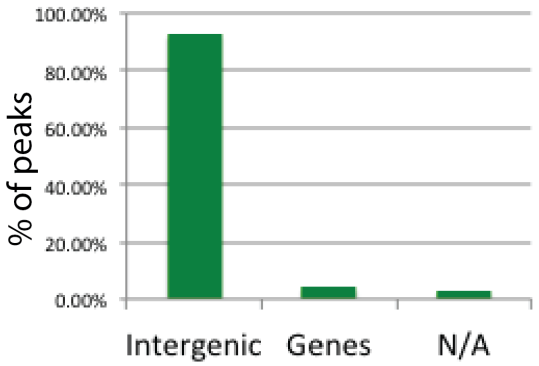
Figure 12: A) Representative Western blot for the different methylation states of H4K20 upon Suv4-20h knock down in radially injected X. *laevis* embryos (n=3. Each biological replicate comes from a different mating pair run on a different gel). B) Western blot for the different methylation states of H4K20 upon overexpression of the Suv4-20h enzymes in radially injected X. *laevis* embryos n=3.

We then looked at overall K20me3 occupancy in whole embryo ChIP-seq. We retrieved ChIP-seq datasets in biological triplicates for K20me3 from the literature (van Kruijsbergen, Hontelez et al. 2017) for the blastula, gastrula and late gastrula stage produced from wild-type X. *tropicalis* embryos. Thus we performed peak calling for each dataset. We soon realized that many peaks appear to be stage specific; this can be appreciated for the *oct* genes locus (Figure 13A). A peak is clearly visible at the blastula and gastrula stage on the *oct25* and *oct91* genes but it disappears by the late-gastrula stage. This observation could reflect a plastic deposition in different cell types or developmental stages of K20me3. Alternatively, it could simply reflect a reduced accessibility of the chromatin to the K20me3 antibody, in line with the notion that more accessible regions of the chromatin are immune-precipitated more easily (Jain, Baldi et al. 2015). For this reason, in order to focus on the most stable peaks, we averaged together the three developmental stages. In this way we obtained a profile that is more resistant to the stage specific peaks fluctuations that we discussed above. We used this combined track for our subsequent analysis. In total we retrieved 11758 peaks. K20me3 seems to be predominantly intergenic with 92.9% of the peaks falling into this category while only 4% of the peaks is associated with genes (counting together exons, introns, TSS and promoters) (Figure 13B). N/A stands for non-

A



B



C

Name	Log2(Ip/Input)
LSU-rRNA_Hsa	3,618
SSU-rRNA_Dme	3,502
SSU-rRNA_Hsa	2,581
Gypsy-8-LTR_XT	2,444
Gypsy-16-I_XT	2,28
DIRS-4B_XT	2,22
Gypsy-8-I_XT	1,968
Gypsy-27-I_XT	1,938
ERV1-4-I_XT	1,809
ERV1-4-LTR_XT	1,601
tRNA-Ile-ATA	1,601
Gypsy-19-I_XT	1,591
DIRS-16_XT	1,504

D

	Mus musculus (REF)	Client Text Box Input (Hierarchy NEW!)					
GO molecular function complete	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
Ras GTPase binding	377	11	2.30	4.78	+	2.58E-05	2.33E-02
↳small GTPase binding	393	11	2.40	4.58	+	3.72E-05	2.80E-02
↳protein binding	8675	78	53.00	1.47	+	1.91E-05	2.15E-02
↳binding	12778	112	78.06	1.43	+	8.60E-10	3.89E-06

Figure 13: A) Genome browser view of K20me3 occupancy at the *oct25* locus. Inp. stands for input, B, G and late-G stand for: blastula, gastrula and late gastrula stage. A stands for averaged. For each track n=3 except Input where n=1 and A where n=9. B) Bar chart indicating the genomic localization of H4K20me3 peaks relative to genes. C) List of repetitive DNA elements (mainly transposons) enriched for K20me3 (more than 1.5 log2(Ip/Input). D) Panther GO analysis

(molecular function complete) of the function of the genes that are associated with K20me3.

assigned and refers to those peaks that cannot be attributed to either category. This result is in line with what we know from the literature. It is known in fact that K20me3 decorates the centromeric satellite repeats and the major satellite repeats. Moreover, we have been able to demonstrate an enrichment of K20me3 on specific classes of transposable elements partially overlapping with what was already described in the literature (Figure 13 C) (van Kruijsbergen, Hontelez et al. 2017). Notwithstanding that, the result could reflect the complexity of the entire embryo. In fact, if we use a less stringent peak definition we can see that up to 30% of the total number of peaks associates with genes (data not shown). K20me3 could decorate different loci in different embryonic tissues and these differences could go averaged out in the whole embryo ChIp-seq, while its affinity for transposable elements could be consistent among different tissues.

In our analysis less than 300 genes are decorated with K20me3. We thus ran a gene ontology analysis on them using Panther statistical enrichment test against the mouse gene ontology annotation. The only category that is significantly enriched is “Ras-GTPase binding” (Figure 13D). At this point we did not consider this observation relevant, but as I will show it will acquire meaning by the end of this dissertation.

Among the few genes that are decorated by K20me3 there are *oct25* and *oct91.* In Figure 13A the ChIp-seq tracks are shown on IGB

genome browser for the *oct60/oct25/oct91* locus. In this specific case K20me3 is enriched on the *oct25* and *oct91* genes, while it does not decorate *oct60*. The decoration of this locus fluctuates; it is high at the blastula and gastrula stage and fades out by the late gastrula stage. Still, when averaging the three time-points together (Figure 13A, A-track) we get an overall better noise to signal ratio. This confirms our earlier ChIp-PCR results (Nicetto et al., 2013) and extends them for oct91.

What happens when we remove K20me3 by inhibiting the Suv4-20h enzymes? We took as an example *oct25* a gene that is highly decorated with K20me3. To answer to this question we injected X. *laevis* embryos radially either with the control or with the Suv4-20h morpholinos, we cut animal caps, harvested them at NF18 and processed them through an *oct25 in situ* hybridization (Figure 14A).

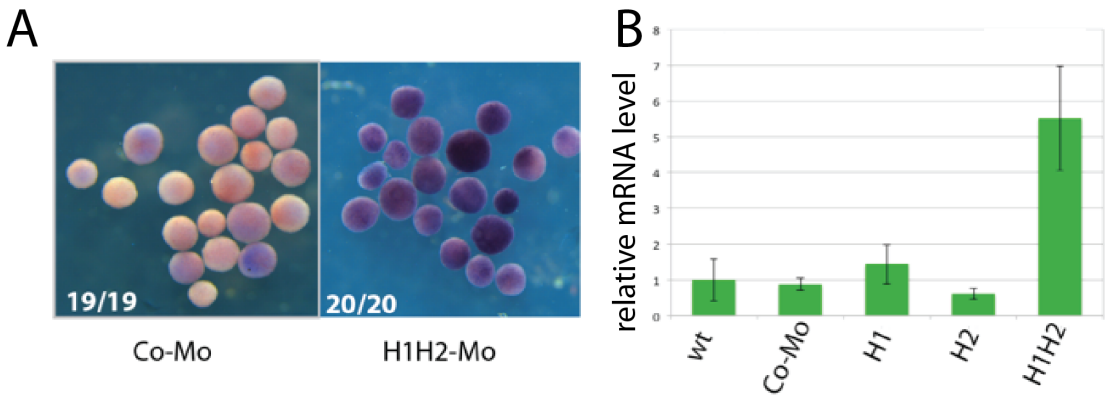


Figure 14: A) RNA *in situ* hybridization against *oct25* mRNA on ACs at NF18 injected with a control (Co-Mo) or both Suv4-20h morpholinos (H1H2-Mo). B) *oct25* mRNA abundance relative to *odc* in radially injected embryos n=3.

The result clearly indicates that upon removal of K20me3 *oct25* is strongly up regulated in the animal caps explants, recapitulating previous observations from our lab (Nicetto et al., 2013).

We then wondered: are both enzymes required for the silencing of *oct25*? To answer to this question we induced knockdown of the Suv4-20h enzymes with single morpholinos or both in combination and determined by RT-qPCR the relative *oct25* mRNA levels in AC explants (Figure 14B). While the double-knockdown condition induced a significant up-regulation of *oct25* mRNA, we see here for the first time that the single-knockdown is not sufficient to influence *oct25* transcription.

3.10 Suv4-20h1 depletion affects multiciliogenesis

Recent unpublished data from our lab has indicated a specific epidermal phenotype in Suv4-20h double knockdown embryos. The cellular composition of the mucociliary epithelium in such embryos is fully maintained, showing normal abundance for goblet cells, ionocytes and small secretory cells. The abundance of multiciliated cells is slightly enriched in the double-morphant epithelium, consistent with the observed upregulation of *dll-1* mRNA, the Notch ligand responsible for MCC fate selection (Stubbs, Davidson et al. 2006). Most notably, however, the cilia production in these MCC cells is severely compromised. This cell type-specific defect impairs a major function of the mucociliary epithelium, as shown by the reduced fluid flow on the surface of the double morphant embryos. At the current state of our analysis, the ultimate ciliogenic defect occurs at the level of axoneme formation at apically docked basal bodies. Taken together, this data indicates that Suv4-20h1/h2 enzymes are required in MCC cells to form a cilia tuft (personal communications by Dario Nicetto, Ohnmar Hsam and Julian Berges).

In order to distinguish between requirements for H4K20me2 or -me3 marks in the process of multiciliogenesis, we injected unilaterally X. *tropicalis* embryos with Suv4-20h morpholinos and stained them at NF16 for acetylated alpha-tubulin on the epidermal surface. This modified tubulin is part of the ciliary axoneme and provides an abundant epitope to stain cilia tufts. We tested the suv4-20h morpholinos singly and in combination. The results, shown in Figure 15, indicate that the injection of control morpholino or Suv4-20h2 morpholino alone have no effect on acetylated alpha-tubulin levels on the injected side. However, the single injection of suv4-20h1 morpholino clearly reduced the axonemal stain. The reduction is even stronger, when the two Suv4-20h enzymes are knocked down simultaneously. This suggests that although Suv4-20h1 seems to be the main regulator of multiciliogenesis, Suv4-20h2 can partially compensate for its loss. This postulated compensation could explain why the ciliogenic defect is exacerbated, when both enzymes are knocked down in concert, similarly to what we observe for H4K20me3 levels (Figure 15A).

This finding has important implications. Unlike the derepression of the *pou5f3.2/oct25* gene, which requires the simultaneous knockdown of both Suv4-20h enzymes (Figure 14B; Nicetto et al., 2013), the ciliogenic defect in MCCs is principally elicited by the single knockdown of Suv4-20h1. Since the two enzymes are endowed with different tasks – Suv4-20h1 writing K20me2 marks genome-wide, while Suv4-20h2 writing K20me3 marks in constitutive heterochromatin and on few gene loci – the reported neural

differentiation (Nicetto et al., 2013) and multiciliogenic phenotypes in Xenopus may be rooted in misregulation of distinct genes.

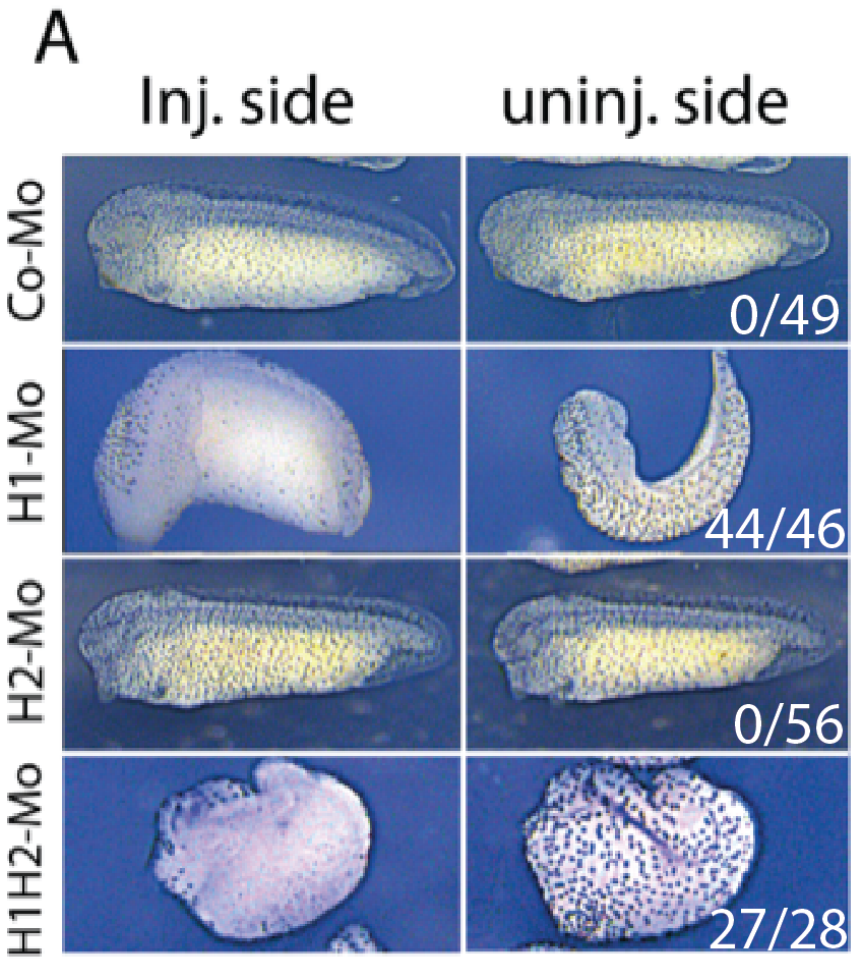


Figure 15: Multiciliogenesis is strongly affected upon Suv4-20h depletion. A) Embryos were unilaterally injected with either Control morpholino or Suv4-20h morpholinos alone or in couple, they were fixed around NF28 and processed through an ICC against acetylated alpha-tubulin. Each visible dot on the embryonic skin represents a tuft of cilia. Numbers represent the portion of

embryos that show down-regulation of Ac-alpha tubulin staining on the injected side. The experiment has been performed 3 times (n=3)

3.11 Suv4-20h depletion specifically affects ciliogenesis

The experiments described above have elucidated differences between single and double knockdown conditions for Suv4-20h enzymes with regard to corresponding histone modification states in bulk chromatin and a ciliogenic defect in the larval skin. These findings immediately prompt questions as to the molecular nature of the ciliogenic defect, which aspect of MCC differentiation is impaired, and - perhaps most astonishingly – how such abundant modifications like H4K20 dimethyl marks can be linked to a phenotype of such exquisite specificity?

To answer to these questions, we took again advantage of the AC explants system to obtain information on genome-wide transcriptional changes occurring in the Suv4-20h depleted condition during epidermal differentiation. Xenopus *tropicalis* embryos were radially injected, at the two cell-stage, either with control morpholino or with both Suv4-20h morpholinos, since this condition achieves the maximal ciliogenic defect. The animal caps were cut between NF8 and NF9 and harvested when the sibling embryos were reaching NF10.5, NF16 and NF24 (Figure 16). This allows to differentiate between transcriptional changes occurring in the context of wild type AC differentiation, as described in the first part of my thesis, from changes brought about by the loss of H4K20me2/-me3 modifications.

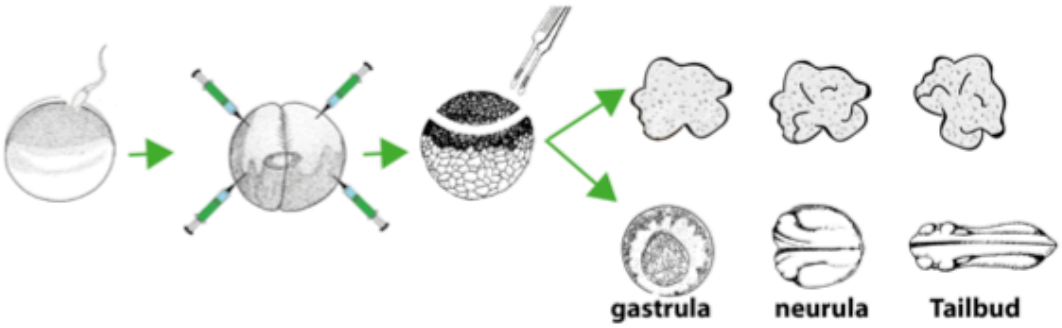


Figure 16: Scheme representing the generation of double-morphant animal caps. Embryos were generated and subsequently injected with the Suv4-20h morpholinos radially. When embryos reached NF8-9 (mid- to late blastula stage), we dissected the ACs manually and culture them in isolation until sibling embryos reach the desired developmental stages (NF10.5, NF16 and NF24).

Approximately 30 AC per stage were collected in triple biological replicates, both for the control and for the double morphant condition. Already at this point we noticed 2 relevant differences between the double morphant ACs and their control siblings: first, they do not spin around, propelled by the MCCs, an indication for the reduced axonemes, and second, they tend to lose cells around the neurulation stage and they completely disassociate by NF28 (data not shown). We subsequently extracted total RNA and prepared RiboZero libraries that have then been sequenced with a coverage of approx. 30 million reads per library. We first tried to confirm the correlation between K20me3 occupancy and gene expression in double morphant embryos. Figure 17A depicts the composite oct gene locus as an example. The RNA-seq tracks for the Suv4-20h double and control knock down animal caps are loaded together with whole embryo ChIP-seq tracks of K20me3 (the “A track” of Figure 13A

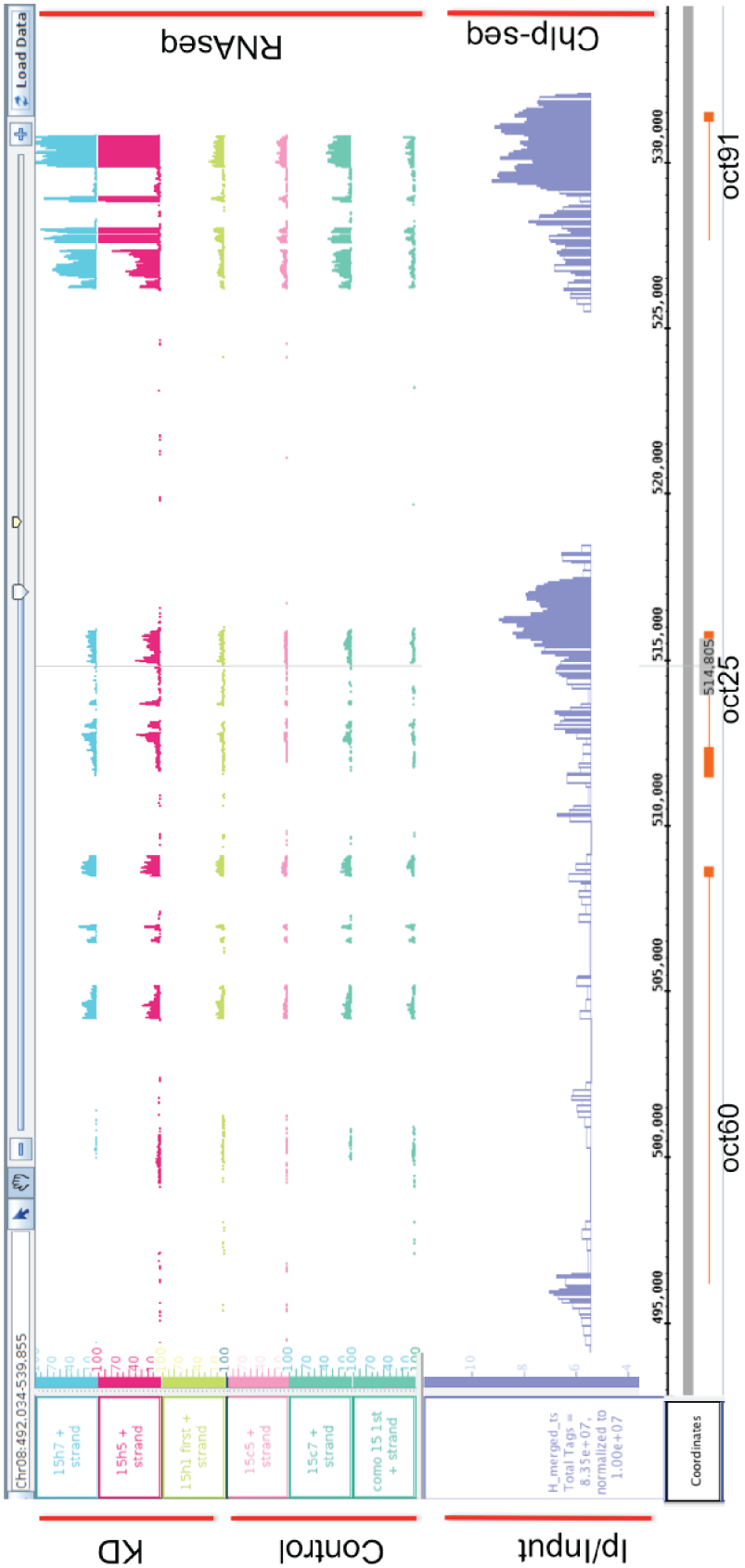


Figure 17: Transcriptional response upon K20me3 depletion at the oct60/oct25/oct91 genetic locus.
A) RNA-seq tracks for NF16 for the Suv4-20h double knock down animal caps (labeled with "KD"), followed by control knock down animal caps (named "Control") loaded together with the averaged (A) Ch-IP-seq track (labeled as Ip/Input) for H4K20me3 over the *oct* genes locus.

is shown only). The *oct60* gene is poorly decorated with H4K20me3 and its expression is not affected by Suv4-20h inhibition, while *oct25* and *oct91* are decorated with K20me3 and get up-regulated upon Suv4-20h removal. We then looked at differences in gene expression between Suv4-20h double knock down and control knock down animal caps in the three timepoint. At the mid-neurula stage the multiciliated cell fate has been specified. At this stage, nearly three thousand genes are misregulated in the animal caps with a p-value<0.05, remarkably the majority (1723/2997, i.e. 58%) being down regulated (Supplementary Table 12). This magnitude and the bias towards down regulation of gene expression is incompatible with the assumption that we are relieving a repressive chromatin mark from genes (Figure 18A). In the earlier stage we investigated (NF10.5) the KD of the Suv4-20h enzymes does not lead to a strong transcriptional mis-regulation. This observation may reflect the proximity of this stage to the mid-blastula-transition and so could witness the fact that ACs at this stage are still relying on maternal transcripts. Interestingly though, the most down-regulated gene in this cohort is Suv4-20h1. This result is quite unexpected given the fact that mopholino oligonucleotides are translational blocking compounds that are not supposed to affect mRNA stability. This observation may reveal the existence of a novel transcriptional feedback loop regulating the expression of Suv4-20h1.

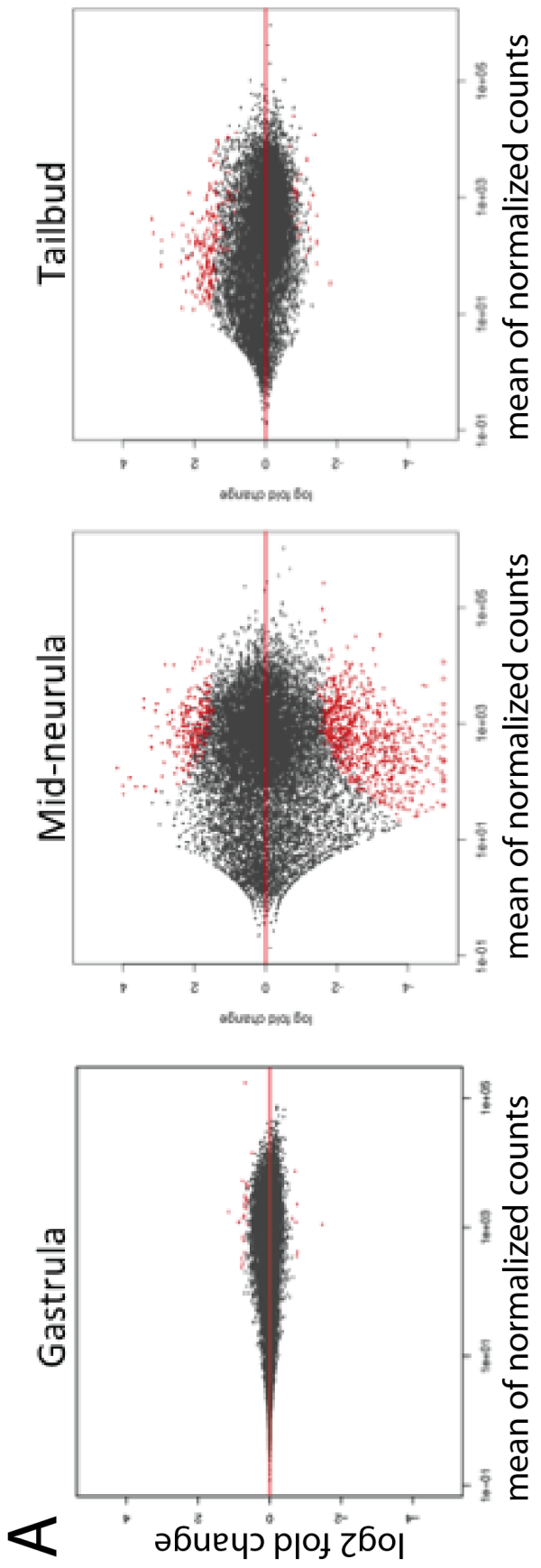
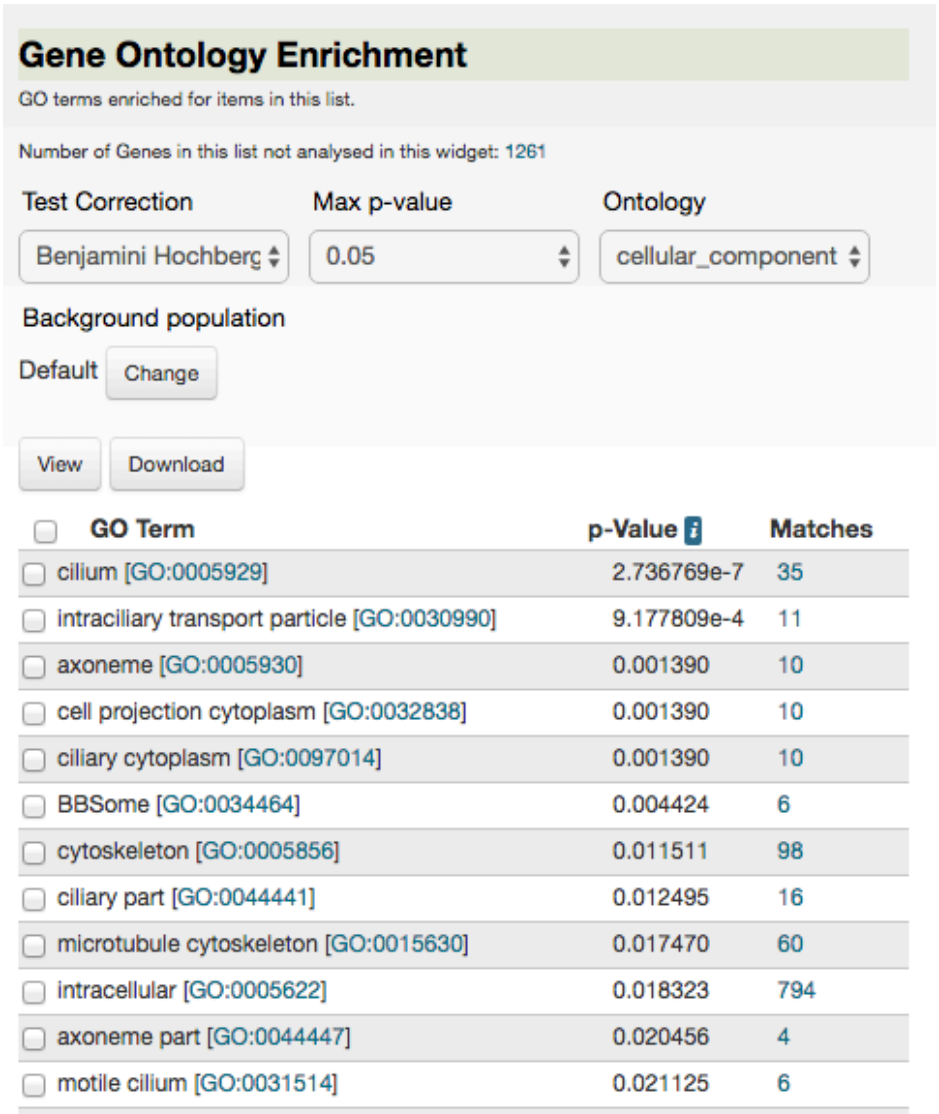


Figure 18: Gene expression in Suv4-20h KD ACs normalized to Control KD A) Scatter plot of the differences in gene expression between the control and the Suv4-20h knock down caps. Each dot represents a gene, in red genes that are differentially regulated between the 2 conditions. On the Y-axis is plotted the log2 fold change while the X-axis the mean of normalized count. Explant age: Gastrula (NF10.5), mid-neurula (NF16) and tailbud (NF24).

We decided to approach the data first by looking at gene ontology enrichment. We made use of two different publicly available algorithms: Xenmine (Reid, Karra et al. 2017) and Panther (Mi, Poudel et al. 2016). The first database is specific for Xenopus, while the second exploits the superior annotation of the mouse genome. Using Panther is possible because the Xenopus gene nomenclature copies the mammalian gene annotation for homologous genes.

A



GO Term	p-Value	Matches
cilium [GO:0005929]	2.736769e-7	35
intraciliary transport particle [GO:0030990]	9.177809e-4	11
axoneme [GO:0005930]	0.001390	10
cell projection cytoplasm [GO:0032838]	0.001390	10
ciliary cytoplasm [GO:0097014]	0.001390	10
BBSome [GO:0034464]	0.004424	6
cytoskeleton [GO:0005856]	0.011511	98
ciliary part [GO:0044441]	0.012495	16
microtubule cytoskeleton [GO:0015630]	0.017470	60
intracellular [GO:0005622]	0.018323	794
axoneme part [GO:0044447]	0.020456	4
motile cilium [GO:0031514]	0.021125	6

Figure 19: XenMine GO results for “cellular component” A) Snapshot of the results obtained by running only the cohort of down-regulated genes in the Suv4-20h KD NF16 ACs.

When we ran the entire cohort of down-regulated genes on Xenmine under the ontology category “cellular component” we found that basically all the significantly enriched categories have to deal with the formation of cilia (Figure 19). Specifically, the term cilium scores first with 35 genes associated with it, the second term is intraciliary transport particle, followed by axoneme with 11 genes, cell projection cytoplasm with 10 genes, ciliary cytoplasm 10 genes BBsome 6, cytoskeleton with 98 genes and so on. The selectivity of this result is even more striking when blasting exclusively the down-regulated genes cohort against Panther’s mouse gene ontology annotation. When we look at overrepresented cellular components (GO term = cellular component complete) and perform a statistical enrichment test on our list, the result is even more striking. The first enriched terms are connected to ciliogenesis, with the first one being again cilium with 144 genes belonging to this category, followed by cell projection with 310 genes. In lower positions the algorithm identified cytoskeleton with 282 genes, axoneme with 42 genes, motile cilium with 42 genes or ciliary basal body and centrosomes with 82 hits (Figure 20).

The results from GO enrichment analysis strongly support the specificity of the multiciliogenic phenotype observed upon Suv4-20h enzyme knockdown. But how is this specificity achieved? How can a genome wide histone modification such as H4K20me2 affect the expression of a functionally specific subset of genes?

A

	Mus musculus (REF)	Client Text Box Input (Hierarchy) NEW!				
GO cellular component complete	#	#	expected	Fold Enrichment	+/-	Δ P value
Unclassified	1426	22	88.24	.25	-	0.00E00
cilium	556	144	34.40	4.19	+	6.87E-43
cell projection	2163	310	133.84	2.32	+	2.83E-41
intracellular	13036	1044	806.65	1.29	+	3.54E-38
cytoplasm	10309	885	637.90	1.39	+	6.43E-38
plasma membrane bounded cell projection	1990	286	123.14	2.32	+	9.84E-38
cytoskeleton	1963	282	121.47	2.32	+	4.43E-37
intracellular part	12923	1032	799.65	1.29	+	3.24E-36
cell part	15203	1151	940.74	1.22	+	2.44E-35
cell	15207	1151	940.98	1.22	+	2.90E-35
ciliary part	404	110	25.00	4.40	+	5.85E-34
cytoskeletal part	1421	221	87.93	2.51	+	1.04E-32
organelle	12315	983	762.03	1.29	+	1.59E-31
cell projection part	1115	180	68.99	2.61	+	8.85E-28
plasma membrane bounded cell projection part	1115	180	68.99	2.61	+	8.85E-28
microtubule cytoskeleton	1053	173	65.16	2.66	+	1.99E-27
intracellular organelle	11286	880	698.36	1.26	+	4.87E-20
axoneme	104	42	6.44	6.53	+	1.02E-17
ciliary plasm	105	42	6.50	6.46	+	1.44E-17
microtubule organizing center	600	105	37.13	2.83	+	1.63E-17
motile cilium	161	49	9.96	4.92	+	6.11E-16
ciliary basal body	122	42	7.55	5.56	+	2.91E-15
cellular_component	20795	1353	1286.76	1.05	+	8.79E-15
centrosome	471	82	29.14	2.81	+	4.11E-13

Figure 20: Panther GO results for "cellular component complete". A) Snapshot of the results obtained by running only the cohort of down-regulated genes in the Suv4-20h KD NF16 ACs.

On theoretical grounds, three scenarios could explain these observations. First, similarly to the response found for Oct25/Oct91, the depletion in H4K20me2/me3 modifications could transcriptionally de-repress inhibitors of the multiciliogenic program. To validate this hypothesis we looked at the expression level of characterized negative regulators of ciliogenesis: *cdc20b*, miR449, *cep110, cep97* and *kif19a*. We found these genes not to be affected by the KD except in few instances (*cep97* and *kif19a*) where the genes are down regulated. Second, double-morphant embryos

may experience a disbalance of cilium formation and degradation, which enhances ciliary turnover. Indeed, degradation of motile cilia in MCCs is initiated by the caseine kinease 1ε (Ck1ε) which phosphorylates Dvl2 activating Hef-1, Aurora kinase A and Hdac6 (Pugacheva, Jablonski et al. 2007, Lee, Johmura et al. 2012, Shnitsar, Bashkurov et al. 2015). This results in deacetylation of tubulin and degradation of the cilia. Again also in this case the KD affected the transcription of none of these factors. Finally, loss of H4K20 di- and tri-methylation could bear indirect consequences that link epigenetic information to the ciliogenic program.

Before ruling out definitely the existence of a negative regulator we wondered whether an uncharacterized negative transcriptional regulator of ciliogenesis might exist. In order to assess this point we took advantage of the Homer algorithm to look for common motifs in the cis-regulatory regions (1 kb around TSS) of the down-regulated ciliogenic genes derived from the Xenmine list. Not surprisingly we identify the DNA binding motif of the Rfx transcription factor family as being enriched (in order: Rfx5, Rfx3, Rfx1, Rfx2, Rfx6) (Figure 21A). When we repeated the analysis taking into account the entire list of down-regulated genes at NF16 the analysis retrieved the same factors in a different order (in order: Rfx3, Rfx1, Rfx2, Rfx5 and Rfx6) (Figure 21B). Rfx factors are known positive regulators of the ciliogenic program both in primary and motile cilia (Thomas, Morle et al. 2010), thus they cannot be considered as negative transcriptional regulators.. Moreover, none of them is mis-regulated upon Suv4-20h depletion in the RNAseq results. Notably, also none of the other known positive transcriptional regulators of the

multiciliogenic program, such as *foxj1*, *mci*, *e2f4* and *e2f5* is significantly misexpressed in Suv4-20h double morphant condition.

A

Rank	Motif	Name	P-value	log P-pvalue	q-value (Benjamini)	# Target Sequences with Motif
1		Rfx5(HTH)/GM12878-Rfx5-ChIP-Seq(GSE31477)/Homer	1e-5	-1.273e+01	0.0027	22.0
2		RFX(HTH)/K562-RFX3-ChIP-Seq(SRA012198)/Homer	1e-5	-1.197e+01	0.0029	12.0
3		Rfx1(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-4	-9.862e+00	0.0160	15.0
4		Rfx2(HTH)/LoVo-RFX2-ChIP-Seq(GSE49402)/Homer	1e-4	-9.561e+00	0.0162	11.0
5		X-box(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-3	-7.955e+00	0.0647	11.0
6		Rfx6(HTH)/Min6b1-Rfx6.HA-ChIP-Seq(GSE62844)/Homer	1e-2	-6.328e+00	0.2745	26.0
7		AT1G69570(C2C2dof)/col-AT1G69570-DAP-Seq(GSE60143)/Homer	1e-2	-5.376e+00	0.6094	32.0
8		COG1(C2C2dof)/col-COG1-DAP-Seq(GSE60143)/Homer	1e-2	-4.791e+00	0.9566	32.0

B

Rank	Motif	Name	P-value	log P-pvalue	q-value (Benjamini)	# Target Sequences with Motif
1		RFX(HTH)/K562-RFX3-ChIP-Seq(SRA012198)/Homer	1e-14	-3.232e+01	0.0000	120.0
2		X-box(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-11	-2.735e+01	0.0000	143.0
3		Rfx1(HTH)/NPC-H3K4me1-ChIP-Seq(GSE16256)/Homer	1e-11	-2.682e+01	0.0000	192.0
4		Rfx2(HTH)/LoVo-RFX2-ChIP-Seq(GSE49402)/Homer	1e-11	-2.622e+01	0.0000	119.0
5		Rfx5(HTH)/GM12878-Rfx5-ChIP-Seq(GSE31477)/Homer	1e-11	-2.582e+01	0.0000	294.0
6		Rfx6(HTH)/Min6b1-Rfx6.HA-ChIP-Seq(GSE62844)/Homer	1e-4	-1.111e+01	0.0023	491.0

Figure 21: Enriched motifs in the cis-regulatory regions of the down-regulated genes. A) Motif enriched in a region of 1 kb surrounding the TSS of the ciliogenic genes previously identified by XenMine (GO category cilium). B) Motifs enriched in a region of 1 kb surrounding the TSS in the cohort of down-regulated genes in Suv4-20h depleted ACs at NF16.

Two out of three of our hypotheses were ruled out by lack of evidences. The last option left is a direct or indirect mechanism linking K20 methylation to the ciliogenic program. In order to assess this option, we investigated the cohorts of mis-regulated genes from the other two developmental stages.

By NF24 multiciliated cells specification and differentiation has already happened. At this stage the Suv4-20h depleted animal caps exhibit different properties respect to the control ones; they lose cellular adhesion and interestingly, they start to fall apart. Differential gene expression analysis reveals that only 224 genes are mis-regulated in Suv4-20h depleted ACs at this stage (Supplementary Table 13). The reduced number of mis-regulated genes probably reflects the fact that a principal component analysis is unable to cluster the double knock down apart from the control knock down samples. That is very different from what we observed for NF16 where the two conditions cluster separately from each other. Last, no gene ontology class is enriched in the cohort of mis-regulated genes.

More interesting is the case of the gastrula transcriptome (NF10.5). This is the earliest stage we analyzed; it proceeds by several hours the massive mis-regulation observed at NF16 and follows by a couple of hours the onset of the zygotic transcription. Moreover, it is the stage when the commitment of lineages starts to happen. At this stage only 30 genes are differentially regulated between the Suv4-20h depleted and the control animal caps with the majority being up regulated (23/30 genes) (Supplementary table 11). The majority of up-regulated genes (16/23) code for proteins involved either in the organization of cytoskeletal structures and/or in the regulation of the cell cycle. Arhgef17, Ralgapb, Rassf3 and Rasl11b are regulators of

the actin cytoskeleton stability, while Map7d3, Mapkbp1, Dtl and Gpaa1 regulate the dynamics of the microtubule cytoskeleton (Stelzer, Rosen et al. 2016). Very interesting is the co-mingling between the major cytoskeletal realms (Actin/Tubulin) and cell cycle regulation. For the actin cytoskeleton for example Arhgef17 is a guanine nucleotide exchange factor (GEF) of RhoA GTPases, but it is also an essential factor of the spindle assembly checkpoint (Isokane, Walter et al. 2016), similarly to Ralgapb and Rassf3. Both control actin cytoskeleton stability and cell cycle progression (Kudo, Ikeda et al. 2012, Personnic, Lakisic et al. 2014). For the microtubules Map7d3 is a protein that control the polymerization of the microtubules in the mitotic spindle (Kwon, Park et al. 2016) while Dtl is a microtubule bound protein that translocates into the nucleus and regulates cell cycle progression (Ueki, Nishidate et al. 2008). Among the regulators of cell cycle we found Rif1 a replication timing regulatory factor, Pds5b that is required for the accumulation of AuroraB at the centrosomes.

3.12 K20 methylation links cell cycle progression to cytoskeletal dynamics.

The transcriptome of the ACs at NF10.5 has revealed us that the majority of the genes being up-regulated at this stage play a function in regulating cytoskeletal dynamics during the S-G2-M phases of the cell cycle. At NF10.5 cells are rapidly dividing and proper G1 phases are starting to appear in the cell cycle. However, at NF16 genes involved in cytoskeletal dynamics are down-regulated in block. At this stage, embryonic cells have a proper G1 phase in their cell cycle

(Iwao, Uchida et al. 2005). It is interesting to notice that these genes being up-regulated do not belong to the mitotic spindle but fall into gene ontology categories representing post-mitotic cytoskeletal structures such as the cilium (obviously enriched in the epidermal compartment) and cell projections (Mi, Poudel et al. 2016, Stelzer, Rosen et al. 2016). Why are cytoskeletal genes down regulated in block at NF16? A possible answer to this question can be deduced from the literature. The lab of Martinez-Balbaz (Asensio-Juan, Gallego et al. 2012) has revealed that K20me1 is a repressive histone mark that specifically controls the expression of cytoskeletal genes. Knock-down of the K20me1 specific de-methylase (Phf8) induces persistence of K20me1 on genes´ promoters leading to a down-regulation of these genes. Phenotypically in neurons this affects the production of cellular projections such neurite outgrowth because of the lack of structural cytoskeletal proteins. We have seen in figure 12A that upon KD of the Suv420h enzymes we have a massive increase of K20me1.

In light of what we said above the lack of ciliary axonemes in multiciliated cells that have been correctly specified could be due to the persistence of K20me1 on cytoskeletal genes similarly to what the Martinez-Balbas´ lab has described. If this hypothesis was true we should be able to rescue the cilliogenic defect in Suv4-20h double morphant epidermis by reducing K20me1 levels. We can do so by co-injecting the Suv4-20h morpholinos together with a synthetic RNA coding for the Xenopus K20me1 specific de-methylase Phf8. Results of this experiment are shown in Figure 22A. Embryos were unilaterally injected with either the Co-Mo, the Suv4-20h

morpholinos, 500 pg of Phf8 RNA or three rescue conditions where the Suv4-20h morpholinos were co-injected with growing amounts of Phf8 RNA (150, 300 and 500pg). Embryos were then fixed around NF28 and processed through an anti-Acetylated tubulin ICC. Phf8 injections alone increase significantly the amount of acetylated-alpha tubulin staining on the injected side (Figure 22B). But most importantly we were able to rescue the ciliogenic defect in Suv4-20h DMO epidermis via the coinjection of Phf8 RNA. The rescue is dose dependent and follows a ventral-dorsal gradient (Figure 22C). Taken all together our results clearly indicate that the ciliogenic defect observed in Suv4-20h double-morphant epidermis is due to persistence of the repressive histone mark K20me1 on genes involved in the formation of cilia. This situation would recapitulate what has been described in the literature for neurite outgrowth (Asensio-Juan, Gallego et al. 2012).

In line with this finding, our previous observations regarding the defective maturation of neurons in Suv4-20h double morphant embryos (Nicetto et al., 2013) acquire new significance. We speculate that the mechanism we have just proposed would apply also to other post-mitotic cytoskeletal structures, such as neuronal projections. If this is true, why is the link with ciliogenesis so strong? Why we do not find other cytoskeletal structures being enriched in the GO? We reckon that the strong enrichment for ciliogenic factors among the down-regulated genes cohort in our ACs explants can be attributed to the choice of the model tissue under investigation. In fact, ciliary axonemes are the prevalent post mitotic cytoskeletal structures we

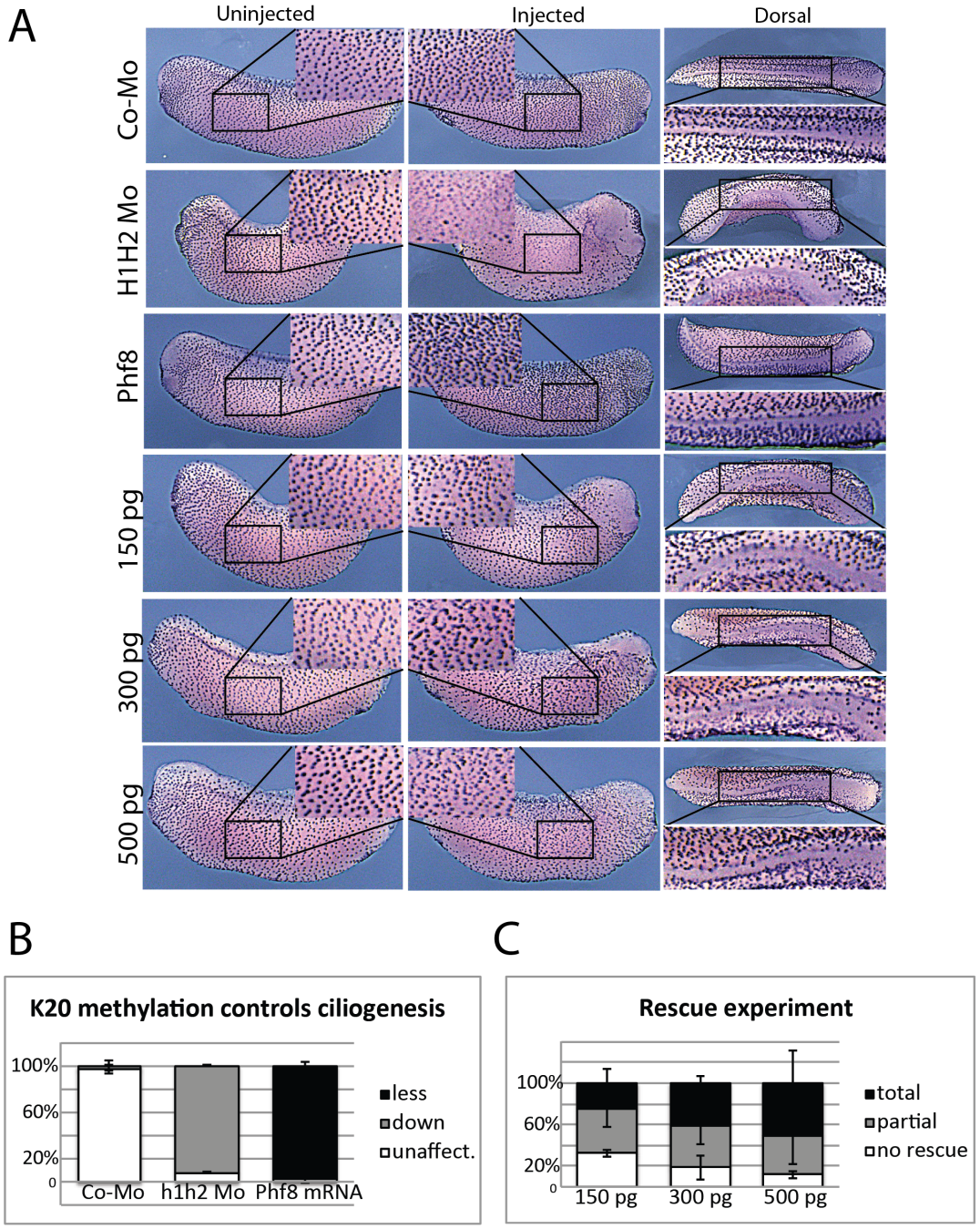


Figure 22: Rescue experiment. A) Embryos were unilaterally injected with either: Control morpholino, Suv4-20h morpholinos, 500 pg Phf8 mRNA, suv4-20h morpholinos + 150 pg of Phf8, suv4-20h morpholinos + 300 pg of Phf8 and suv4-20h morpholinos + 500 pg of phf8 mRNA. The right-most panel shows dorsal pictures of the injected embryos. B) Quantification of the number of embryos showing either less, more or unaffected amounts of Acetylated alpha-tubulin staining following control morpholino, suv4-20h morpholinos or Phf8 mRNA. Error bars represent standard deviations. C) Embryos were sorted in the three

rescue conditions into three groups depending on the grade of rescue of the ciliogenic phenotype: no rescue, partial rescue, where the ventral side is completely rescued while the dorsal is not, and total rescue, where the cilia staining is restored on the dorsal side as well.

find in mucociliary epithelia. Moreover, it did not escape our notice that in our GO analysis on the down-regulated genes cohort, “cell projection” and “cytoskeleton” are among the most enriched categories encompassing respectively 310 and 282 genes. We speculate that if our analysis were conducted in a different tissue, e.g. neural tissue, we would have found a correlation with neurite outgrowth.

In the discussion chapter of this dissertation I will describe in fine details a model that put all this information together. As a result, we speculate that K20 methylation has evolved as a mechanism that links the cell cycle to the dynamics of cytoskeleton.

4.0 Discussion

4.1 More than 50% of all the genes are differentially regulated during epidermis differentiation

We all derive from a fertilized egg. The intricacies of our neural network, the complexity of our body plans as well as the intellectual output of our individuality have all began with a zygote. How is this complexity achieved? How does a single cell become a fully-grown adult? Our field of interest and ultimately the interest of the developmental biology community is finding answers to those questions. Many key players of the embryonic development have been identified like small signaling molecules, growth factors, signaling cascades and transcription factors. We identified a lack of knowledge regarding the systematic transcriptomic analysis of purely isolated developing tissues.

The Xenopus animal caps (AC) epidermal organoids have allowed us to address questions that are impossible in other model organisms, especially mammalians. First, we managed to safely isolate the prospective ectoderm from blastula embryos without the risk of carrying along underlying germ layers. Second, we managed to follow their development outside the animal body. Third, they allowed us to collect a great number of individuals at the same exact developmental stage. Finally, their size and number enabled us to collect enough RNA to perform transcriptomic analysis with RiboZero selection.

In this dissertation I have initially defined the transcriptomic output of the prospective epidermis at 3 key developmental stages: gastrula, neurula and tailbud. Our initial hypothesis was that not only the differential genes usage, but also other mechanisms contribute to the establishment of organismal complexity. Hereby, we examined the use of alternative splicing, the transcription of repetitive DNA elements, the production of exonic circular RNAs and the interaction networks of expressed transcription factors through the developmental stages. Our results indicate that all these factors are transcriptionally regulated in the skin organoids, suggesting their functional involvement in the differentiation of the embryonic epidermis. This part of our project was thought to be a resource to the Xenopus community. For each dataset we have generated we produced spreadsheets that will be publicly accessible (Supplementary material). Moreover, for reasons that will be elucidated further in this discussion, we generated an online web resource where users can easily access the information about the expressed splicing variants for each given genes at the analyzed time points.

Subsequently we wondered which factors upstream of the transcriptional output guarantee an epidermis-specific regulation of transcription and what is the reason of their specificity. We focused our attention on Suv4-20h histone methyl-transferase. These enzymes write the H4K20me2 and H4K20me3 epigenetic marks. Previous observations from our lab (from O. Hsam and D. Nicetto) had shown by depletion of these 2 proteins, how the differentiation of multiciliated cells on the embryonic epidermis is strongly

impaired. In our observations multiciliated cells are the only epidermal cell type being affected by the depletion, in the terminal stages of their differentiation, when they are to produce bundles of ciliary axonemes. By taking advantage again of the AC organoids we were able to demonstrate that K20 methylation state controls the expression of cytoskeletal genes involved in the formation of specialized post-mitotic structures such as ciliary axonemes and cell projections. Last, we think our results suggest a possible unifying theory for the function of K20 methylation. We propose a model were K20 methylation has evolved as a mechanism to phase cytoskeletal dynamics in concert with the cell cycle.

4.2 Animal Cap explants are pluripotent and differentiate into epidermis

ACs are derived from blastula embryos. At this time point, the prospective ectodermal cells are still pluripotent. In Figure 2 we showed that the pluripotency-associated genes *klf4*, *sox2*, *oct91* and *oct60* are highly expressed at NF10.5 and get shut off by NF24. However, the pluripotency of this tissue is witnessed as well by the fact that ACs can be induced to transdifferentiate into all three germ layers derivates (Ariizumi T et al 2009).

When cultured in growth factor free medium, ACs differentiate by default into epidermis (Cibois, Luxardi et al. 2015). This can be inferred by the fact that mesodermal genes (*eomes, cdx1, cdx2* and *cdx3)* are being repressed by NF24 while ectoderm specific ones (*vimentin, xk81a1, atp60va2* and *ncam1)* are by this stage

strongly up-regulated. In Figure 1, we clearly showed that the whole surface of the ACs explants stains for markers specific for particular epidermal cell types, with a pattern that is undistinguishable from the embryo. We used acetylated alpha-tubulin antibodies to stain multiciliated cells, otogelin in situ probes to stain goblet cells and xv-atpase probe to stain small secretory cells. The only skin cell type we did not investigate directly is small secretory cells. We lacked a proper marker for them, though their presence can be deduced from 2 simple observations; first, keratin, a pan-epidermal marker, stains the entirety of the explants. Second, ACs spin around (videos not showed) propelled by the stroke actionof their cilia tufts multiciliated cells. This behavior could not be elicited without the presence of serotonin-producing cells (Dubaissi, Rousseau et al. 2014). This set of data was meant to demonstrate that ACs are proper and functional skin organoids that correctly recapitulate embryonic epidermis differentiation.

4.3 Regulation of gene transcription in the developing epidermis

The epidermal default differentiation of ACs provides an outstanding opportunity to follow the differentiation of a simple tissue, made of only 4 cell types (Briggs et al., 2018) with no further perturbation after explantation. This extreme reduction of cell types should become reflected in transcriptomic analysis, when compared to the transcriptome of the whole embryo, which by NF22 consists of more than 120 cell types (Briggs et al., 2018). To investigate the differences in RNA expression, we compared in depth the transcriptomes of ACs and whole embryos at neurula stage. We found that 3.8 thousands of

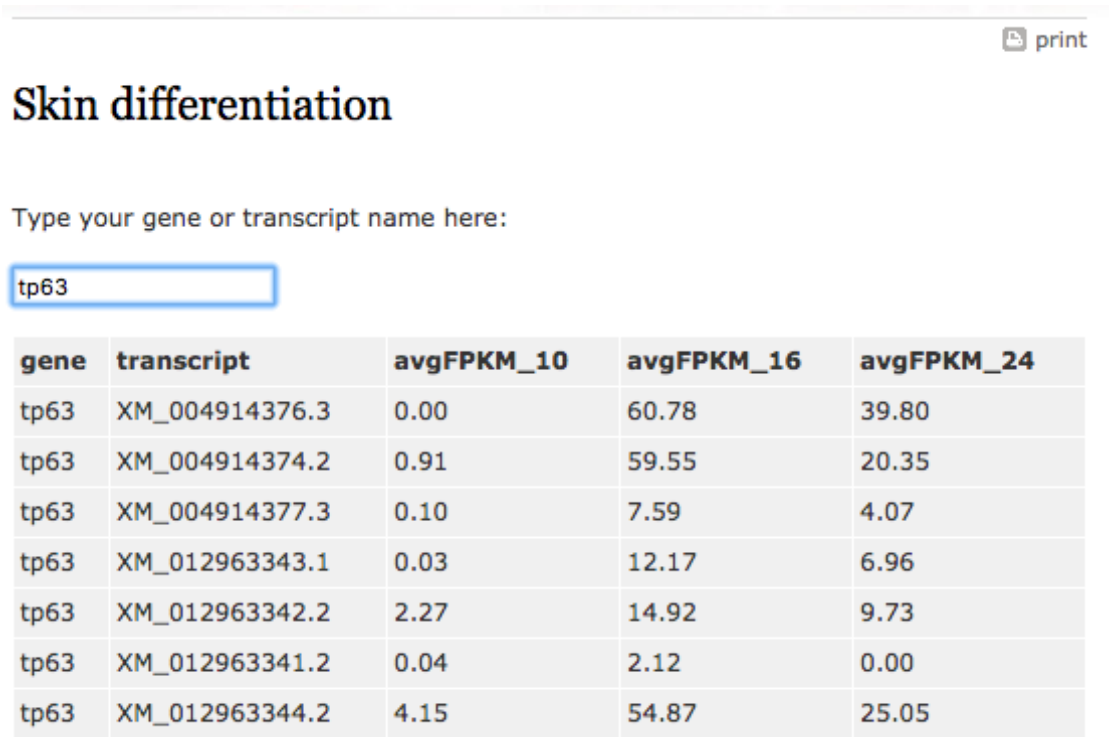
genes are differentially regulated in the ACs (more than 25% of the annotated gene models). Those that are up-regulated in ACS exhibit a smaller extent of variation. This is definitely due to the fact that the prospective epidermis is definitely a component of the whole embryo. In contrast, genes that are down-regulated exhibit a greater variation. Not surprisingly in this cohort belong all the lineage specific genes that are expressed in the embryo but not in the skin organoids (like CNS-, mesoderm and endoderm specific genes). This analysis has identified for the first time the entire repertoire of genes that are specifically expressed in the prospective epidermis at the time of its commitment. We then wondered to what extent gene expression is modulated in the prospective ectoderm to guarantee its transition from pluripotency to terminally differentiated epidermis. For this reason, we compared the transcriptomic datasets of the different stages among themselves. Unexpectedly, we found out that the majority of the frog's transcriptome is differentially expressed during AC´s development. Between NF10.5 and NF16 stages, more than 7000 genes are differentially expressed. In the comparison between NF10.5 and NF24 the number raises up to over 8000. Remarkably little differences are observed between NF16 and NF24, suggesting that the fate decision on a genetic level is taken between the gastrula and the neurula stage and after neurulation the transcriptome of the ACs stays relatively stable. Considering that the total number of annotated transcripts in X. *tropicalis* is approx. 15 thousand we conclude that most of the genes are differentially regulated between the pluripotent and the terminally differentiated stage.

But gene expression alone does not account for the variety of expressed proteins. In fact, we noticed that in some cases the total number of reads that align to a gene does not vary in the different developmental stages. Although in many instances the proteins that are derived from the same genetic locus can have very different structural and functional properties. This is because alternative splicing can account for the differential usage of exons and thus for the production of different protein isoforms. We wondered how pervasive differential splicing is in the ACs through their development. To do so we took advantage of the latest X. *tropicalis* annotation release (103). This database sums up an average of 1.84 splicing isoforms per gene. We then aligned our transcriptomic data to this latter annotation and compared the expression of the splicing isoform in the ACs versus the whole embryo at NF16. This method allows us to recognize existing splicing isoforms but it does not permit to identify novel ones. We found out that 3.3 thousand splicing isoforms are differentially expressed between the two datasets. We verified the validity of this observation via a set of *in situs* hybridization where we showed that truly we could easily distinguish differences in the expression of specific exons for a given gene between the embryonic epidermis and the whole embryo.

More remarkable was the result when we compared the different developmental stages among each other. We found out that more than 11 thousand splicing isoforms are differentially expressed between NF10.5 and NF16, while the number is above 12 thousand when comparing NF10.5 to NF24. Is this splicing affecting the coding capacity of the affected mRNAs? Following the existing annotation we derived that 1859 genes are not protein coding, 2819 code for at

least 1 protein while 3789 code for at least 2 protein isoforms. Our data revealed that the majority of these splicing isoforms align to annotated protein models, thus the answer being yes. This result teaches us two novel things. First, a previously undocumented amount of variation between the pluripotent and the differentiated states is achieved through alternative splicing. Second, the variation observed in exon usage is greater than the variation observed in gene usage. This new result has important implications for the Xenopus community. In fact, in our studies we often perform overexpression and *in situ* experiments using synthetic RNA constructs. Until now, our community has always made use of the transcripts annotated on Xenbase to build the synthetic construct required for the aforementioned experiments. Our data clearly indicate that an enormous amount of genes is expressed in the epidermis as specific splicing isoforms, different from the one that is predominant in the embryo and annotated on Xenbase. Thus, our result suggests that many gain of function studies have used the wrong isoform, certainly complicating the interpretation of the results, and potentially even leading to neomorphic effects, when the overexpressed protein isoform is not identical with the prevailing tissue-specific protein. The same can be said for the RNA in situ Hybridization experiments – the presence of alternative exons may reduce the length of the DNA sequence, over which the as RNA probe is complementary to the mRNA, which could alter ultimately the signal strength of this assay. In order to avoid such complications in the future, one should determine the prevailing splice isoforms in the tissue of interest. To assist the community in this goal, at least for the non-neural epidermis, we are not only publishing the entire list of splicing

isoforms for every given gene and their relative abundance, but we have also created an interactive web resource, where this information can be retrieved instantaneously. An example of how it works is shown in Figure 15. The web resource contains a typing box where users can type the name of their gene of interest. In this case we have chosen the proto-oncogene *tp63*, which is expressed in epidermal basal progenitor cells. The web resource retrieves a list of all the annotated splicing variants for the gene and tells us their expression level, as Fragments per Kilobase of transcript per Million of mapped reads (avgFPKM), at the different developmental stages (10, 16, 24). The sequence information for each splicing variant can be retrieved by browsing the NCBI accession name (e.g XM_004914376.3) on the NCBI website.


print

Skin differentiation

Type your gene or transcript name here:

tp63

gene	transcript	avgFPKM_10	avgFPKM_16	avgFPKM_24
tp63	XM_004914376.3	0.00	60.78	39.80
tp63	XM_004914374.2	0.91	59.55	20.35
tp63	XM_004914377.3	0.10	7.59	4.07
tp63	XM_012963343.1	0.03	12.17	6.96
tp63	XM_012963342.2	2.27	14.92	9.73
tp63	XM_012963341.2	0.04	2.12	0.00
tp63	XM_012963344.2	4.15	54.87	25.05

Figure 23: Snapshot of the web resource we have generated to retrieve informations about the expression of splicing isoforms for a given gene in ACs at the analyzed developmental stages

4.4 Temporal and spatial expression of Repetitive DNA elements

As we mentioned in the introduction increasing evidences support a functional involvement of transcripts arising from non-coding DNA elements in the regulation of developmental programs. These sequences are spread over the genome as relics of ancient viral invasions. Most them are decorated with classical repressive histone modifications reflecting the fact that their expression could still lead to mutagenic insertional events. We decided to try to identify lineage specific signatures of their expression. Therefore, we compared the expression of this class of transcripts between the embryo and the AC explants at the gastrula stage. We identify more than 100 elements that are differentially expressed with high significance. Some of them seem to be highly enriched in the animal caps versus the embryo, and vice versa. Then we compared the different developmental stages among themselves. We found out that between the gastrula and the neurula stage hundreds of repeats are differentially expressed, accounting for more than a third of the total number of annotated repeats in Xenopus. To better understand the functional relevance of this result, we decided to take some of the top responders among the differentially regulated repeats between the early and late stages of

animal caps development. We chose 3 repeats that are expressed in ACs at high levels during gastrulation and get shut off by the tailbud stage. When we looked at the *in situ* pattern of expression of these repeats we immediately noticed something unexpected. The transcripts are expressed in well-defined embryonic regions. The first one is mainly enriched in the involuting mesodermal tissue, the second seems to have a neat ectodermal expression while the last one, the centromeric repeat transcript localizes preferentially in the ectoderm and in minor extent in the mesoderm. All of them are expressed at high level at the gastrula stage and get shut off in the tailbud. The centromeric repeats is still on in the neural tube of the embryo at the tailbud stage. What does it witness to us? We have speculated that if repetitive elements expression is just background noise reflecting the immature formation of heterochromatin (Schneider, Arteaga-Salas et al. 2011) at early developmental stages their expression should be uniform among the different embryonic districts. On the other hand, if they play a functional role in directing developmental programs we should see them expressed in a localized manner, similar to tissue specific genes. Our observations support the second hypothesis. This result is the to be meant in the context of the discussion on the function of repetitive DNA elements as the latter evidence of their involvement in normal cellular functioning.

4.5 Intra- and Inter-group transcripts clustering and identification of circularRNAs

How tight is the regulation of the transcription of the various classes of transcripts we have analyzed so far during the development of the prospective epidermis? A way to answer to this question is to perform an unbiased clustering of our various datasets. Each single biological replicate of the three analyzed stages was plotted into a principal component analysis. The result displayed in figure 8 is our answer to the question above. When we consider gene expression the three biological replicates of each time-point cluster among themselves and the different developmental stages are clearly resolved on the 2 principal components. The situation is similar for the expression of the splicing isoforms, although for this class of transcripts we do not reach the same resolution observed for gene expression. The case is completely different for the repetitive DNA elements. If, on one hand, the NF10.5 dataset separates from the others on the first principal component, on the other hand, the samples from NF16 and NF24 intermingle on both principal components, therefore they cannot be told apart. This analysis strongly suggests that genes first and splicing isoforms' expression second are under tighter regulation than repetitive elements during embryonic skin differentiation. Moreover, this analysis gives us a tool to approach further developmental issues. In fact, it suggests that among the three classes of transcripts we analyzed so far mRNAs can be predictive of the developmental stage of an animal cap.

To conclude our analysis of the expressed transcripts in the ACs we turned our attention to the most exotic class of RNAs described so far: the circularRNAs. We said in the introduction that the presence of circularRNAs in the Xenopus model had already been described (Talhouarne and Gall 2014). Our analysis however focused on the

developing prospective epidermis and has revealed for the first time the presence of circularRNAs of exonic origin. Moreover, we have shown that the number of circularRNAs stays pretty much constant during the phases of development while some of them seems to accumulate steadily. Little is known so far about the function of these transcripts. As we said before anecdotal evidences have highlighted specific functions for some of them, although their general significance is still under debate (Huang, Yang et al. 2017) (Kumar, Shamsuzzama et al. 2017). Do they play a functional role in the post-transcriptional regulation of gene expression for example by sponging miRNAs? Are they extremely stable translated messengers? Or are they simply a byproduct of normal transcription? For sure our analysis cannot answer to these questions. What we accomplished so far has been identifying for the first time circularRNAs deriving from the coding sequences of genes in Xenopus. We confirmed the existence of the circular junctions and ultimately we have generated and made publicly available lists of these transcripts, accessible to everyone who wants to address those questions in the Xenopus model system.

4.6 Prediction transcription factors that modulate the epidermal transcriptome

We have characterized so far the transcriptomic output of the embryonic epidermis at key developmental stages. Our analysis up to now has been merely descriptive and was meant to constitute a resource for the developmental community. The next step of our investigation aimed to identify the regulators of the observed

transcriptomic output. Which factors regulate the transcription of specific loci in different stages of the embryonic development? As we all know, regulation of gene expression happens at different levels including every step from the binding of RNA Pol II to a promoter until the translation of the mRNAs is regulated. We want here to focus on the regulation of transcription of specific DNA loci. Above all we think transcription is regulated by two mechanisms: First, the binding of transcription factors to the DNA, which by virtue of their DNA-binding and protein-protein interaction domains select their target genes (Todeschini, Georges et al. 2014)- Second, via epigenetic regulation of DNA accessibility, which makes the DNA information available to the binding of TFs and of the PolII machinery (Jaenisch and Bird 2003). In this last paragraph of the RNAseq analysis we are going to discuss about transcription factors.

We have previously introduced Mogrify. An algorithm that combines between two conditions the fold change in mRNA expression of TFs and it's p-value with information from the STRING interaction networks to prediction a weight of each given TF in regulating the entire transcriptome (Rackham, Firas et al. 2016). We feed the algorithm with the data relative to all the annotated TFs in Xenopus that were expressed in our dataset. Although our analysis is only bioinformatical and thus predictive, some of the results are surprisingly congruent with published experimental information. For example, we have seen that for the gastrula stage the algorithm identifies as top influencers of the transcriptome TFs that are involved in pluripotency maintenance and embryonic patterning. It matches with the fact that at NF10.5 (gastrula stage) cells are still pluripotent and embryonic patterning is taking place. As well as for

the later stages we have seen that it identifies the master regulator of the multiciliogenic cell fate as well as 2 of the factors required to transdifferentiate adult human cell types into keratinocytes. Multiciliated cells are 1 of the 4 cell types that make up the epidermis (Brooks and Wallingford 2014) while all the epidermal cell types in Xenopus exhibit keratinocyte-like properties (In figure 1 keratin staining stains 100% of the epidermis). Our interpretation of this result is that our application of Mogrify is, at least to some extent, effective and exhibits predictive capacities. In fact, we were able to confirm the importance and weight of some of the top transcription factors based on the published literature. One draw back has been the identification of hematopoietic specific TFs as top responders of the Mogrify analysis for NF24. When we addressed the data manually we immediately realized that the number of reads of these factors was very low thus their high ranking may constitute a bias of the analysis. It is important to notice though that no other exotic lineage gets identify by Mogrify (for example liver or kidneys specific TFs). Most probably in the process of cutting the explants occasional mesodermal cells were carried along. This would likely justify the observed low levels of expression of hematopoietic marker genes in the explants. Notwithstanding that, there is still room for a functional explanation of this finding. The transcription factor analysis is supposed to be predictive and we have already said that the best outcome of this methodology would have been identifying novel regulators of ACs´ transcriptome. Possibly, some of the TFs identified as being specific for the hematopoietic lineage might also play a role in ACs specification.

It would be extremely interesting in the future to perform functional analysis on the other TFs whose role in epidermal differentiation has never been described. A fast way of assessing novel regulators of epidermis differentiation would be knocking down these transcripts with the use of morpholinos or engineering dominant negative isoforms. One could check if the differentiation and/or gene expression is affected in specific epidermal cell types as a read out. For this reason, as we did for all the previous dataset, also the predictive TF analysis is made publicly available to the developmental community. Still it is extremely stimulating to think that such a bioinformatic algorithm could help solving such sophisticated issues as identifying major drivers in cell types specification.

To conclude we reckon that the identification of Tp63, Foxj1, Foxi1, Foxa1, Rfx1-2-3 and Grhl1-3 in the TFs analysis is really exemplar. This is because all these factors have been identified since many years as being essential factors for the differentiation of epidermal cell types in Xenopus (Esaki, Hoshijima et al. 2009, Vidarsson, Westergren et al. 2009, Choksi, Lauter et al. 2014, Dubaissi, Rousseau et al. 2014, Pan, Adair-Kirk et al. 2014). The regulation of the multiciliogenetic differentiation program will be the subject of the next paragraphs of this dissertation.

4.7 Epigenetic regulation via Suv4-20h histone methyl-transferases

As we said previously much is already known about the transcription factors that regulate the differentiation of the different epidermal cell types. Our analysis up to now was meant to explore further possible

regulators of epidermal differentiation by looking at canonical as well as exotic factors such as splicing isoforms, repetitive DNA elements and circRNAs. Given our interest for such regulators, it did not escape our notice that previous observations from our lab suggested another level of regulation of epidermal cell types differentiation; the epigenetic regulation.

In fact, previous observations from Hsam, Nicetto and Berges have shown that upon depletion of Suv4-20h histone methyl-transferases multiciliogenesis on the embryonic epidermis is strongly impaired (Nicetto 2012, Hsam 2015)(Julian Berges personal communication). In the model we illustrated in the introduction, Suv4-20h1 seems to be responsible for the di-methylation of K20, while Suv4-20h2 writes the tri-methylated state. We decide to validate this model in Xenopus. We found out that upon depletion of the enzymes singularly the embryos strongly accumulate K20me1, while the tri-methylation of the residue is lost only upon double knockdown in concert of the 2 enzymes. In mouse the situation is very similar. MEFs that lack either of the two enzymes strongly accumulate K20me1 and the max abundance is reach in the double-null situation. Similarly, when K20me3 is abolished completely only in the double-null condition (Schotta, Sengupta et al. 2008) This suggests that their enzymatic functions are at least partially promiscuous and both enzymes are to some extent responsible for both the di- and tri-methylation of this residue in mouse as well as in Xenopus. The situation is diametrically opposite when we overexpressed the two enzymes. In fact, both of them significantly reduce K20 mono-methylation and increase the higher methylated states. This result does not solve completely the

issue of which is the favorite substrate of the two enzymes but clearly indicate that their function can be redundant.

The function of H4K20me2 in mammalian cells is still not well understood despite several studies have tried to shade light on this modification (Botuyan, Lee et al. 2006, Schotta, Sengupta et al. 2008, Yang, Pesavento et al. 2008). On theoretical grounds, H4K20me2 is considered to be a neutral histone mark. In fact, it is unlikely that it can elicit specific functions given the fact that approx. 80% of H4 is di-methylated in embryos with mixed cell composition (mitotic/non-mitotic cells). We thus speculated that Suv4-20h enzymes might control gene expression via K20 tri-methylation. So, we decided to look at K20me3 occupancy genome wide in the whole embryo. By conservative peak calling, we found out that K20me3 enriched regions are almost completely intergenic. Similarly to H3K9me3, (Bulut-Karslioglu, De La Rosa-Velazquez et al. 2014) in mouse ES cells, only a few hundred genes are decorated with this modification, while more 5500 are decorated with K27me3, the best studied gene-associated repressive histone modification (Young, Willson et al. 2011). Among the genes that are decorated with K20me3 the only ontology group being significantly enriched is that of the Ras-associated G-proteins. Initially this data did not mean much to us but as we will see later in the final model it acquired a completely new significance while proceeding with our study. Among the genes that are decorated with K20me3 we identified the *oct91* and *oct25* genes, consistent with the data from Nicetto et al. (Nicetto et al., 2013). We demonstrated that upon removal of either of the protein the expression of *oct25* is not affected. However, when we knockdown both genes in concert with strongly up regulate oct25 expression. We

then verified that the de-repression of *oct25* in the double knockdown is an actual phenomenon in the ACs as well. Our analysis clearly shows a functional redundancy of the two enzymes, at least with respect to the *oct25/pou5f3* gene locus. We must say that the *oct25* gene locus localizes in the sub-telomeric region of chromosome 8, in fact it falls in the first half million bp of chromosome 8 (Figure 24).

Figure 24: screenshot taken from XenBase showing the sub-chromosomal location of oct25. The position of the oct25 gene is indicated on a model of chromosome 8 by a red bar. Its genomic coordinates are shown in the panel below.

This is relevant because, as we said in the introduction, it is known that the Suv4-20h enzymes take part in the formation of heterochromatin in the centromere, pericentromeric regions, telomere and subtelomeric regions. In light of these observations the de-repression of these locus could be attributed to a position effect variegation-like phenomenon. Position effect variegation (PEV) is a phenomenon first described in Drosophila for which repressive histone marks present at the centromere can spread in the pericentromeric regions and influence the expression of genes. The search for of mutations that abolish the PEV have led to the initial

discovery of SU(VAR)3-9 the histone methyl-transferase responsible for the deposition of H3K9me3 (Elgin and Reuter 2013). Several HMT have strong affinity for the modification they write. It is thought that when they bind to already modified histones they will modify neighboring nucleosomes leading to a spread of the modification. This phenomenon has been described already for H3K9me3 and H3K27me3 in different species (Bannister, Zegerman et al. 2001, Lachner, O'Carroll et al. 2001). These molecular mechanisms are conserved from yeast to human (Towbin, Gonzalez-Aguilera et al. 2012) and so, we deduce, they must exist as well in Xenopus. In mammalian systems it has been shown that H3K9me3 recruit HP1 to pericentromeric and subtelomeric regions, which subsequently recruits Suv4-20h2, which will write the H4K20me3 mark (Schotta, Lachner et al. 2004, Schotta, Sengupta et al. 2008). Moreover, Suv4-20h2 recruits via direct interaction condensins to induce chromosomes condensation (Hahn, Dambacher et al. 2013). The behavior of Suv4-20h2 is thus very different from the ones of Pr-Set8 and Suv4-20h1 that modify K20 in a genome wide manner. For this reason, taking into consideration the subtelomeric localization of *oct25* we propose that the transcription of this gene is controlled by the spreading of the heterochromatin in the subtelomeric domain. However, this model works only if we assume that Suv4-20h1 in Xenopus can compensate for the lack of Suv4-20h2 in the subtelomeric region, given the fact that *oct25* is de-silenced only when both genes are knocked down.

4.8 Suv4-20h depletion strongly impairs multiciliogenesis in the embryonic epidermis

As we have said before our interest for the Suv4-20h enzymes derives from our knowledge of their effect on multiciliogenesis. In Figure 12 we have shown that down-regulation of Suv4-20h1, but not Suv4-20h2, leads to an impairment in cilia tufts formation on the embryonic epidermis. This is the first time that we identify a functional difference between the 2 enzymes. The phenotype gets exacerbated when both enzymes are knocked down in concert. To perform the staining, we used an antibody targeting acetylated alpha-tubulin a major component of the ciliary axonemes. It can be appreciated the fact that stained foci are present on the injected side but their coloration is much fainter, their diameter smaller. The confocal analysis has revealed the reason for this observation. In fact, we have demonstrated that in Suv4-20h double knocked down mosaic epidermis multiciliated cells are correctly specified. This can be deduced by the fact that they have managed to intercalate on the outer layer of the epidermis and they have multiplied their basal body. The multiplication of the basal body is prerogative of multiciliated cells only. Furthermore, at least some of the basal bodies manage to dock apically, a prerequisite to enucleate axoneme formation. Furthermore, in the cells reached by the morpholino injection we do not see an apical actin meshwork, which fails to form. Last, these cells present a strongly reduced number of ciliary axonemes, few axonemal structures protrude from the multiciliated cells and their length is suboptimal. This observation explains why acetylated alpha-tubulin staining is reduced on the morpholino-injected side. Overall this analysis has demonstrated that Suv4-20h enzymes control the formation of axonemal structures and of the

actin cap meshwork in multiciliated cells. These cells are correctly specified, they undergo a proper differentiation reflected by the multiplication of the centriols, but fail to produce proper axonemal structures suggesting that just the very last step of multiciliogenesis differentiation is impaired.

4.9 RNA-Seq analysis of Suv4-20h depleted animal caps

The embryo is a complex organism, composed of many tissues that interact with each other in order to guarantee proper embryonic development. As we said in the result section for the H4K20me3 ChIP-Seq (Figure 13), the noisiness of the result we got could reflect a differential occupancy of K20me3 in different cell types and embryonic compartment. We thought that the Suv4-20h control of multiciliogenesis must be elicited via regulation of gene transcription. In order to validate this hypothesis avoiding looking at the complexity of the entire embryo we took again advantage of the ACs organoids. Our transcriptomic analysis confirmed our previous observations for the *oct25* genes locus and reinforces the idea that K20me3 plays a repressive function. Consequently, we have found out that the majority of mis-regulated genes in the morphant condition were actually down regulated. Moreover, the most enriched gene ontology categories in the down regulated gene cohort have to deal with the formation of ciliary structures, cytoskeletal structures and, most generally, in the production of cell projections.

The situation at NF10.5 is very different. Few genes are mis-regulated, reflecting probably the proximity of this stage to the first

activation of zygotic transcription. Probably at NF10.5 the majority of the embryonic transcriptome is still composed of maternal transcripts (Lee, Bonneau et al. 2014). Most interestingly the majority of the mis-regulated genes is up-regulated and these genes are implicated in the regulation of the spindle apparatus (Stelzer, Rosen et al. 2016), as if depletion of Suv4-20h induces or protract the expression of G2-M genes.

We ruled out that the down-regulation in block of genes at NF15 might be due to the up regulation of a negative regulator of ciliogenesis, possibly a transcription factor, which is under the transcriptional control of K20me3. When we looked for common DNA motifs in the regulatory regions of the affected ciliogenic genes, without surprise, we found them to be enriched for the binding consensus sequence of several members of the RFX transcription factors family. None of these factors is mis-regulated in the RNA-seq tracks upon depletion of the Suv4-20h enzymes. This clearly gave us a hint that the regulation of the multiciliogenic programs by the Suv4-20h enzymes could be direct and pass through the control of transcription.

Given the fact that the majority of the K20me3 peaks localize intergenically (mainly centromere ad telemeres) and our analysis plus the literature tell it is enriched at evolutionarily young transposons (van Kruijsbergen, Hontelez et al. 2017) we thought it favors their repression. Thus, we looked at repetitive DNA elements expression but we found out that very few elements get mis-regulated, with even numbers of them being up and down regulated (data not shown). We concluded from this that K20me3 doesn´t play a central role in transposons repression.

4.10 H4K20me1 is a repressor

From the literature (Jorgensen, Schotta et al. 2013, van Nuland and Gozani 2016) we know that K20me1 is established during the S phase of the cell cycle, when new histones get incorporated into chromatin. Pr-Set7 the enzyme responsible for the methylation travels along the replication machinery bound to PCNA. In this way the newly deposited histones will carry the K20me1 modifications while old histones will be predominantly di and, to a lesser extent, tri-methylated. Incorporation of new histones happens on all the genome so that K20me1 is not restricted to specific sub-genomic loci. K20me1 stays unaltered during the G2- and M-phases of the cell cycle. Interestingly, this modification is considered to be repressive and it has been demonstrated it directly regulates genes involved in cytoskeletal dynamics (Asensio-Juan, Gallego et al. 2012). When cells go in G1 or exit the cell cycle, H4K20 monomethyl gets further methylated. Suv4-20h1 in mouse spreads all over the genome (Schotta, Sengupta et al. 2008), where it turns K20me1 into K20me2 . K20me2 is the most abundant histone modification found in vertebrates, more than 80% of H4 is di-methylated (Jorgensen, Schotta et al. 2013).K20me2 is then neither a repressor nor an activator of transcription decorating virtually every genetic locus. K20me3 on the other hand results to behave as a repressor of transcription as we have shown previously for a subset of genes. Its localization is restricted to specific heterochromatin regions overlapping with the sub-nuclear localization of Suv4-20h2. It is not surprising then that the *oct25* gene locus is found in the sub-

telomeric region of chromosome 8. In this specific context, we speculated that it controls and represses gene expression in a position effect variegation-dependent manner. Summing up all these considerations we have seen that upon removal of the Suv4-20h enzymes we repress in block cytoskeletal genes. This led us to speculate that the repression is due to the persistence on these genetic loci of H4K20me1. In fact, by converting K20me1 into 20me2, Suv4-20h1 would remove a repressive histone modification turning it into a neutral signal. If this hypothesis were true there would be an easy way to verify them. In fact, there would be another mechanism removing H4K20me1 from the genome, opposite to converting it into K20me2, to rescue the repression of cytoskeletal genes and in general multiciliogenesis in Suv4-20h depleted epidermis. This can be achieved via up-regulation of the H4K20me1 specific de-methylating enzyme Phf8.

Phf8 is a promiscuous de-methylase in vivo and in vitro. It acts on several substrates, such as H3K9me1/2 and H3K27me2, although it shows a preference for K20me1 (Liu, Tanasa et al. 2010). It associates with active gene promoters where it is recruited by interaction with H3K4me3. Remarkably, PHF8 protein knockdown in zebrafish embryos (Qi, Sarkissian et al. 2010) leads to an accumulation of K20me1 and closely resembles the phenotypes we have described in Suv4-20h depleted embryos including: loss of cell adhesion, impaired neuronal survival, impaired craniofacial development and cell proliferation. For this reason we induced overexpression of Phf8 alone or in a Suv4-20h double morphant background. Remarkably, Phf8 injection alone is sufficient to increase the amount of Acetylated alpha-tubulin staining in a very penetrant

manner (Figure 22B). Finally, we manage to rescue the Suv4-20h dependent ciliogenic phenotype by co-inecting Phf8 mRNA together with the morpholinos, in a dose dependent fashion. This experiment strongly suggests that inhibition of ciliogenesis in Suv4-20h depleted epidermis is due to a persistence of H4K20me1 on genes involved in the formation of cytoskeletal structures. Phf8 can erase this inhibition via removal of the excess of H4K20me1. Finally, this finding intrinsically suggests a functional model that connects cytoskeletal dynamics, gene expression regulation and cell-cycle signaling.

4.11 A unifying model for H4K20 methylation

All together the observations I presented in my dissertation strongly suggest a model for the Suv4-20h-mediated impairment of multiciliogenesis; H4K20me1 is a repressive histone modification that is written in concert with the incorporation of new histones during the S-phase of the cell cycle. This modification represses the transcription of cytoskeletal regulatory genes such as RhoA, Rac1, GSK3b (Figure 25A). Accumulation of K20me1 via inhibition of its de-methylating enzyme Phf8 leads to cell cycle delays, loss of cell adhesion and induces disorganization of the actin cytoskeleton (Asensio-Juan, Gallego et al. 2012).

All these phenotypes are observed also in the Suv4-20h double knock down embryos. In fact, Suv4-20h enzymes as well as Phf8 remove K20me1. In this scenario, removal of H4K20me1 repression could be achieved either via its de-methylation by Phf8 or via further methylation to K20me2 by the Suv4-20h enzymes (Figure 25B).

Here we propose a model where Suv4-20h knock-down leads to an accumulation of H4K20me1. An increase similar to what is observed with Phf8 inhibition. The accumulation of this histone residue, as already described in the literature, represses in block cytoskeletal genes leading to a disorganization of specialized cytoskeletal structures such as neuronal projections (Asensio-Juan, Gallego et al. 2012). It is possible that in wild type cells, Suv4-20h1 or Phf8 are directly recruited to these cytoskeletal genes, to relieve repression, once cells enter the G1/G0 phase. We do not expect a full overlap between the target genes of the these two epigenetic remodelers. But we propose that in different cell types the two enzymes could be targeted to different subset of cytoskeletal genes so that e.g. in neurons they would de-repress genes involved in the formation of neuronal projections while in multiciliated cells they would de-repress structural ciliary proteins.

Why would de H4K20me1 repression specifically affect cytoskeletal gene? We propose that the expression of structural proteins involved in the formation of such entities would surely interfere with the assembly of the mitotic spindle. That is why it must be limited to the post-mitotic (G1-/G0) phase of the cell cycle. For example, each cilium assembles itself on a centriole, the same structure required for the organization of the mitotic spindle. In line with our hypothesis, cytoskeletal proteins involved in the formation of the spindle should escape K20me1 mediated repression. Coherently with this latter hypothesis we find that NF10.5 ACs depleted for the Suv4-20h

enzymes up-regulate a subset of cytoskeletal genes involved in the regulation of the spindle.

In mouse mutations in the Suv4-20h enzymes lead to mitotic defects and aneuploidy. In this model system, It has been shown that Suv4-20h1 is responsible for the di-methylation of K20 while Suv4-20h2 writes the trimethyl mark. The di-methylation of K20 is spread all over the genome while K20me3 clusters in heterochromatic foci (Figure 25C). Our data have indicate that Suv4-20h1 inhibition is the main responsible for

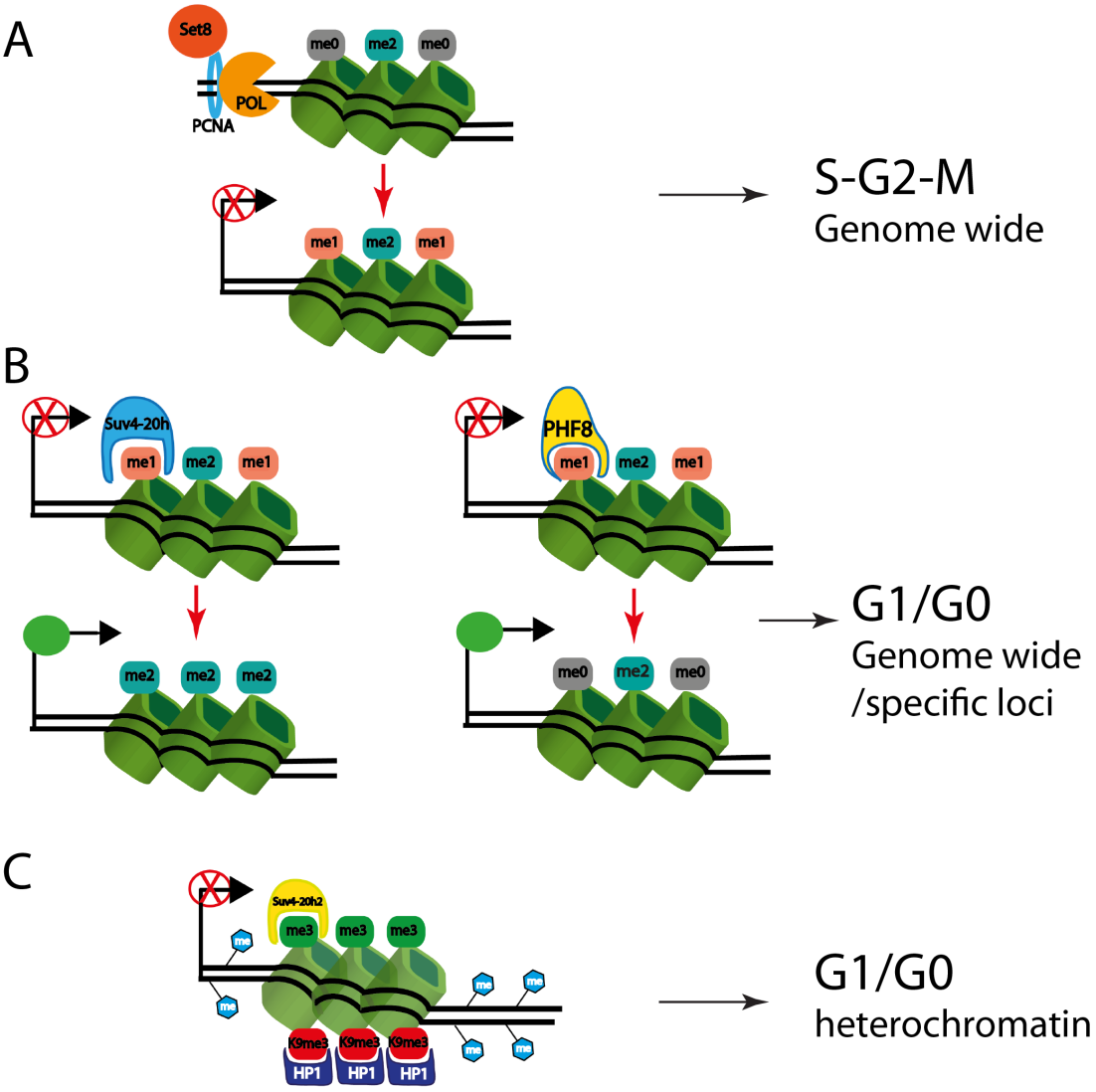


Figure 25: Model A) During the S phase of the cell cycle new histones are synthesized and inserted into the genome, H4K20 is un-methylated (me0) at this stage. Old histones carrying modifications are diluted (me2). Set8 mono-methylates K20 to K20me1 (me1) establishing genome wide a repressive histone modification. B) During the G1/G0 phases of the cell cycle Suv4-20h enzymes (mainly h1) further methylate K20me1 into K20me2 while Phf8 de-methylates K20me1. The removal of K20me1 happens in a genome-wide manner but depending on the cell type the two enzymes can be recruited to specific loci. Both enzymatic activities remove the repressive histone mark H4K20me1 allowing transcription. C) During the G1/G0 phases of the cell cycle Suv4-20h2 is recruited by HP1 to the telomeres, centromeres, sub-telomeric and pericentromeric regions, where it lays down the H4K20me3 repressive histone mark in the context of constitutive heterochromatin (green chromatin).

the defect in multiciliogenesis. This is coherent with our model because K20me1 is the privileged substrate of Suv4-20h1 that would then be, together with Phf8, a major player in its removal.

Thereby we are supporting a scenario where the Suv4-20h enzymes may have evolved as a mechanism to control cytoskeletal dynamics in concert with the cell cycle. During the S and M phases K20 gets mono-methylated on newly synthesized histones, this guarantees that cytoskeletal proteins that may interfere with the spindle formation are repressed. When the cells enter the G1 and G0 phases of the cell cycle the Suv4-20h enzymes get expressed and remove K20me1 by methylating it to K20me2. This would remove the repression on the post-mitotic cytoskeletal genes that would at the end contribute to the formation of post-mitotic cytoskeletal structures, such as ciliary axonemes, apical actin meshworks, or neural projections.

5.0 Bibliography

References

Akkers, R. C., S. J. van Heeringen, U. G. Jacobi, E. M. Janssen-Megens, K. J. Francoijs, H. G. Stunnenberg and G. J. Veenstra (2009). "A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos." Dev Cell **17**(3): 425-434.

Alabert, C., T. K. Barth, N. Reveron-Gomez, S. Sidoli, A. Schmidt, O. N. Jensen, A. Imhof and A. Groth (2015). "Two distinct modes for propagation of histone PTMs across the cell cycle." Genes Dev **29**(6): 585-590.

Allfrey, V. G., R. Faulkner and A. E. Mirsky (1964). "Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis." Proc Natl Acad Sci U S A **51**: 786-794.

Angelov, D., A. Molla, P. Y. Perche, F. Hans, J. Cote, S. Khochbin, P. Bouvet and S. Dimitrov (2003). "The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling." Mol Cell **11**(4): 1033-1041.

Annunziato, A. T. (2005). "Split decision: what happens to nucleosomes during DNA replication?" J Biol Chem **280**(13): 12065-12068.

Asashima, M. and H. Grunz (1983). "Effects of inducers on inner and outer gastrula ectoderm layers of Xenopus laevis." Differentiation **23**(3): 206-212.

Asashima, M., T. Michiue and A. Kurisaki (2008). "Elucidation of the role of activin in organogenesis using a multiple organ induction system with amphibian and mouse undifferentiated cells in vitro." Dev Growth Differ **50 Suppl 1**: S35-45.

Asensio-Juan, E., C. Gallego and M. A. Martinez-Balbas (2012). "The histone demethylase PHF8 is essential for cytoskeleton dynamics." Nucleic Acids Res **40**(19): 9429-9440.

Bannister, A. J., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire and T. Kouzarides (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-124.

Bao, W., K. K. Kojima and O. Kohany (2015). "Repbase Update, a database of repetitive elements in eukaryotic genomes." Mob DNA **6**: 11.

Beck, D. B., H. Oda, S. S. Shen and D. Reinberg (2012). "PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription." Genes Dev **26**(4): 325-337.

Berger, S. L., T. Kouzarides, R. Shiekhattar and A. Shilatifard (2009). "An operational definition of epigenetics." Genes Dev **23**(7): 781-783.

Blitz, I. L., K. D. Paraiso, I. Patrushev, W. T. Y. Chiu, K. W. Y. Cho and M. J. Gilchrist (2017). "A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo." Dev Biol **426**(2): 409-417.

Borchers, A. and T. Pieler (2010). "Programming pluripotent precursor cells derived from Xenopus embryos to generate specific tissues and organs." Genes (Basel) **1**(3): 413-426.

Botuyan, M. V., J. Lee, I. M. Ward, J. E. Kim, J. R. Thompson, J. Chen and G. Mer (2006). "Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair." Cell **127**(7): 1361-1373.

Briggs, J. A., C. Weinreb, D. E. Wagner, S. Megason, L. Peshkin, M. W. Kirschner and A. M. Klein (2018). "The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution." Science **360**(6392).

Bromberg, K. D., T. R. Mitchell, A. K. Upadhyay, C. G. Jakob, M. A. Jhala, K. M. Comess, L. M. Lasko, C. Li, C. T. Tuzon, Y. Dai, F. Li, M. S. Eram, A. Nuber, N. B. Soni, V. Manaves, M. A. Algire, R. F. Sweis, M. Torrent, G. Schotta, C. Sun, M. R. Michaelides, A. R. Shoemaker, C. H. Arrowsmith, P. J. Brown, V. Santhakumar, A. Martin, J. C. Rice, G. G. Chiang, M. Vedadi, D. Barsyte-Lovejoy and W. N. Pappano (2017). "The SUV4-20 inhibitor A-196 verifies a role for epigenetics in genomic integrity." Nat Chem Biol **13**(3): 317-324.

Brooks, E. R. and J. B. Wallingford (2014). "Multiciliated cells." Curr Biol **24**(19): R973-982.

Bulut-Karslioglu, A., I. A. De La Rosa-Velazquez, F. Ramirez, M. Barenboim, M. Onishi-Seebacher, J. Arand, C. Galan, G. E. Winter, B. Engist, B. Gerle, R. J. O'Sullivan, J. H. Martens, J. Walter, T. Manke, M. Lachner and T. Jenuwein (2014).

"Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells." Mol Cell **55**(2): 277-290.

Choksi, S. P., G. Lauter, P. Swoboda and S. Roy (2014). "Switching on cilia: transcriptional networks regulating ciliogenesis." Development **141**(7): 1427-1441.

Chu, J. S., D. L. Baillie and N. Chen (2010). "Convergent evolution of RFX transcription factors and ciliary genes predated the origin of metazoans." BMC Evol Biol **10**: 130.

Cibois, M., G. Luxardi, B. Chevalier, V. Thome, O. Mercey, L. E. Zaragosi, P. Barbry, A. Pasini, B. Marcet and L. Kodjabachian (2015). "BMP signalling controls the construction of vertebrate mucociliary epithelia." Development **142**(13): 2352-2363.

Cole, B. B., R. W. Smith, K. M. Jenkins, B. B. Graham, P. R. Reynolds and S. D. Reynolds (2010). "Tracheal Basal cells: a facultative progenitor cell pool." Am J Pathol **177**(1): 362-376.

Collart, C., N. D. Owens, L. Bhaw-Rosun, B. Cooper, E. De Domenico, I. Patrushev, A. K. Sesay, J. N. Smith, J. C. Smith and M. J. Gilchrist (2014). "High-resolution analysis of gene activity during the Xenopus mid-blastula transition." Development **141**(9): 1927-1939.

Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young and R. Jaenisch (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state." Proc Natl Acad Sci U S A **107**(50): 21931-21936.

Davidson, D. J., F. M. Kilanowski, S. H. Randell, D. N. Sheppard and J. R. Dorin (2000). "A primary culture model of differentiated murine tracheal epithelium." Am J Physiol Lung Cell Mol Physiol **279**(4): L766-778.

De Robertis, E. M. (2006). "Spemann's organizer and self-regulation in amphibian embryos." Nat Rev Mol Cell Biol **7**(4): 296-302.

De Robertis, E. M. and H. Kuroda (2004). "Dorsal-ventral patterning and neural induction in Xenopus embryos." Annu Rev Cell Dev Biol **20**: 285-308.

Didon, L., R. K. Zwick, I. W. Chao, M. S. Walters, R. Wang, N. R. Hackett and R. G. Crystal (2013). "RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium." Respir Res **14**: 70.

Dubaissi, E., K. Rousseau, R. Lea, X. Soto, S. Nardeosingh, A. Schweickert, E. Amaya, D. J. Thornton and N. Papalopulu (2014). "A secretory cell type develops alongside multiciliated cells, ionocytes and goblet cells, and provides a protective, anti-infective function in the frog embryonic mucociliary epidermis." Development **141**(7): 1514-1525.

Eid, A., D. Rodriguez-Terrones, A. Burton and M. E. Torres-Padilla (2016). "SUV4-20 activity in the preimplantation mouse embryo controls timely replication." Genes Dev **30**(22): 2513-2526.

Elgin, S. C. and G. Reuter (2013). "Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila." Cold Spring Harb Perspect Biol **5**(8): a017780.

Esaki, M., K. Hoshijima, N. Nakamura, K. Munakata, M. Tanaka, K. Ookata, K. Asakawa, K. Kawakami, W. Wang, E. S. Weinberg and S. Hirose (2009). "Mechanism of development of ionocytes rich in vacuolar-type H(+)-ATPase in the skin of zebrafish larvae." Dev Biol **329**(1): 116-129.

Fazzari, M. J. and J. M. Greally (2004). "Epigenomics: beyond CpG islands." Nat Rev Genet **5**(6): 446-455.

Felsenfeld, G. and M. Groudine (2003). "Controlling the double helix." Nature **421**(6921): 448-453.

Filion, G. J., J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker and B. van Steensel (2010). "Systematic protein location mapping reveals five principal chromatin types in Drosophila cells." Cell **143**(2): 212-224.

Fraga, M. F., E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suner, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector, Y. Z. Wu, C. Plass and M. Esteller (2005). "Epigenetic differences arise during the lifetime of monozygotic twins." Proc Natl Acad Sci U S A **102**(30): 10604-10609.

Gao, X., C. M. Vockley, F. Pauli, K. M. Newberry, Y. Xue, S. H. Randell, T. E. Reddy and B. L. Hogan (2013). "Evidence for multiple roles for grainyhead-like 2 in the establishment and maintenance of human mucociliary airway epithelium.[corrected]." Proc Natl Acad Sci U S A **110**(23): 9356-9361.

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd and H. O. Smith (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases." Nat Methods **6**(5): 343-345.

Grainger, R. M. (2012). "Xenopus tropicalis as a model organism for genetics and genomics: past, present, and future." Methods Mol Biol **917**: 3-15.

Hahn, M., S. Dambacher, S. Dulev, A. Y. Kuznetsova, S. Eck, S. Worz, D. Sadic, M. Schulte, J. P. Mallm, A. Maiser, P. Debs, H. von Melchner, H. Leonhardt, L. Schermelleh, K. Rohr, K. Rippe, Z. Storchova and G. Schotta (2013). "Suv4-20h2 mediates chromatin compaction and is important for cohesin recruitment to heterochromatin." Genes Dev **27**(8): 859-872.

Hamburger, H. (1955). "Embryogenesis: progressive differentiation - amphibians." Analysis of Development: pp. 230–296.

Hansen, K. H., A. P. Bracken, D. Pasini, N. Dietrich, S. S. Gehani, A. Monrad, J. Rappsilber, M. Lerdrup and K. Helin (2008). "A model for transmission of the H3K27me3 epigenetic mark." Nat Cell Biol **10**(11): 1291-1300.

Heijmans, B. T., E. W. Tobi, L. H. Lumey and P. E. Slagboom (2009). "The epigenome: archive of the prenatal environment." Epigenetics **4**(8): 526-531.

Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson and D. S. Rokhsar (2010). "The genome of the Western clawed frog Xenopus tropicalis." Science **328**(5978): 633-636.

Hemmati-Brivanlou, A. and D. A. Melton (1994). "Inhibition of activin receptor signaling promotes neuralization in Xenopus." Cell **77**(2): 273-281.

Hoogenboom, W. S., D. Klein Douwel and P. Knipscheer (2017). "Xenopus egg extract: A powerful tool to study genome maintenance mechanisms." Dev Biol **428**(2): 300-309.

Hsam, O. (2015). The epigenetic regulation of ciliogenesis by xenopus laevis Suv4-20h histone methyltransferases. Thesis.

Huang, S., B. Yang, B. J. Chen, N. Bliim, U. Ueberham, T. Arendt and M. Janitz (2017). "The emerging role of circular RNAs in transcriptome regulation." Genomics **109**(5-6): 401-407.

Iossifov, I., M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y. H. Lee, G. Narzisi, A. Leotta, J. Kendall, E. Grabowska, B. Ma, S. Marks, L. Rodgers, A. Stepansky, J. Troge, P. Andrews, M. Bekritsky, K. Pradhan, E. Ghiban, M. Kramer, J. Parla, R. Demeter, L. L. Fulton, R. S. Fulton, V. J. Magrini, K. Ye, J. C. Darnell, R. B. Darnell, E. R. Mardis, R. K. Wilson, M. C. Schatz, W. R. McCombie and M. Wigler (2012). "De novo gene disruptions in children on the autistic spectrum." Neuron **74**(2): 285-299.

Isokane, M., T. Walter, R. Mahen, B. Nijmeijer, J. K. Heriche, K. Miura, S. Maffini, M. P. Ivanov, T. S. Kitajima, J. M. Peters and J. Ellenberg (2016). "ARHGEF17 is an essential spindle assembly checkpoint factor that targets Mps1 to kinetochores." J Cell Biol **212**(6): 647-659.

Iwao, Y., Y. Uchida, S. Ueno, N. Yoshizaki and Y. Masui (2005). "Midblastula transition (MBT) of the cell cycles in the yolk and pigment granule-free translucent blastomeres obtained from centrifuged Xenopus embryos." Dev Growth Differ **47**(5): 283-294.

Jaenisch, R. and A. Bird (2003). "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals." Nat Genet **33 Suppl**: 245-254.

Jain, D., S. Baldi, A. Zabel, T. Straub and P. B. Becker (2015). "Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments." Nucleic Acids Res **43**(14): 6959-6968.

Jiang, Y., Y. Han, S. Petrovski, K. Owzar, D. B. Goldstein and A. S. Allen (2015). "Incorporating Functional Information in Tests of Excess De Novo Mutational Load." Am J Hum Genet **97**(2): 272-283.

Jones, E. A. and H. R. Woodland (1987). "The development of animal cap cells in Xenopus: a measure of the start of animal cap competence to form mesoderm." Development **101**(3): 557-563.

Jorgensen, S., G. Schotta and C. S. Sorensen (2013). "Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity." Nucleic Acids Res **41**(5): 2797-2806.

Karimi, K., J. D. Fortriede, V. S. Lotay, K. A. Burns, D. Z. Wang, M. E. Fisher, T. J. Pells, C. James-Zorn, Y. Wang, V. G. Ponferrada, S. Chu, P. Chaturvedi, A. M. Zorn and P. D. Vize (2018). "Xenbase: a genomic, epigenomic and transcriptomic model organism database." Nucleic Acids Res **46**(D1): D861-D868.

Koivisto, A. M., S. Ala-Mello, S. Lemmela, H. A. Komu, J. Rautio and I. Jarvela (2007). "Screening of mutations in the PHF8 gene and identification of a novel mutation in a Finnish family with XLMR and cleft lip/cleft palate." Clin Genet **72**(2): 145-149.

Kudo, T., M. Ikeda, M. Nishikawa, Z. Yang, K. Ohno, K. Nakagawa and Y. Hata (2012). "The RASSF3 candidate tumor suppressor induces apoptosis and G1-S cell-cycle arrest via p53." Cancer Res **72**(11): 2901-2911.

Kumar, L., Shamsuzzama, R. Haque, T. Baghel and A. Nazir (2017). "Circular RNAs: the Emerging Class of Non-coding RNAs and Their Potential Role in Human Neurodegenerative Diseases." Mol Neurobiol **54**(9): 7224-7234.

Kwon, H. J., J. E. Park, H. Song and C. Y. Jang (2016). "DDA3 and Mdp3 modulate Kif2a recruitment onto the mitotic spindle to control minus-end spindle dynamics." J Cell Sci **129**(14): 2719-2725.

Lachner, M., D. O'Carroll, S. Rea, K. Mechtler and T. Jenuwein (2001). "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." Nature **410**(6824): 116-120.

Lee, K. H., Y. Johmura, L. R. Yu, J. E. Park, Y. Gao, J. K. Bang, M. Zhou, T. D. Veenstra, B. Yeon Kim and K. S. Lee (2012). "Identification of a novel Wnt5a-CK1varepsilon-Dvl2-Plk1-mediated primary cilia disassembly pathway." EMBO J **31**(14): 3104-3117.

Lee, M. T., A. R. Bonneau and A. J. Giraldez (2014). "Zygotic genome activation during the maternal-to-zygotic transition." Annu Rev Cell Dev Biol **30**: 581-613.

Legnini, I., G. Di Timoteo, F. Rossi, M. Morlando, F. Briganti, O. Sthandier, A. Fatica, T. Santini, A. Andronache, M. Wade, P. Laneve, N. Rajewsky and I. Bozzoni (2017). "Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis." Mol Cell **66**(1): 22-37 e29.

Liu, W., B. Tanasa, O. V. Tyurina, T. Y. Zhou, R. Gassmann, W. T. Liu, K. A. Ohgi, C. Benner, I. Garcia-Bassets, A. K. Aggarwal, A. Desai, P. C. Dorrestein, C. K. Glass and M. G. Rosenfeld (2010). "PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression." Nature **466**(7305): 508-512.

Loughlin, R., J. D. Wilbur, F. J. McNally, F. J. Nedelec and R. Heald (2011). "Katanin contributes to interspecies spindle length scaling in Xenopus." Cell **147**(6): 1397-1407.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.

MacAlpine, D. M. and G. Almouzni (2013). "Chromatin and DNA replication." Cold Spring Harb Perspect Biol **5**(8): a010207.

Mahjoub, M. R. (2013). "The importance of a single primary cilium." Organogenesis **9**(2): 61-69.

Maienschein, J. (2017). Epigenesis and Preformationism.

Maller, J. L. (2012). "Pioneering the Xenopus oocyte and egg extract system." J Biol Chem **287**(26): 21640-21653.

Margueron, R., N. Justin, K. Ohno, M. L. Sharpe, J. Son, W. J. Drury, 3rd, P. Voigt, S. R. Martin, W. R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg and S. J. Gamblin (2009). "Role of the polycomb protein EED in the propagation of repressive histone marks." Nature **461**(7265): 762-767.

Mi, H., S. Poudel, A. Muruganujan, J. T. Casagrande and P. D. Thomas (2016). "PANTHER version 10: expanded protein families and functions, and analysis tools." Nucleic Acids Res **44**(D1): D336-342.

Milla, C. E. (2016). "The evolving spectrum of ciliopathies and respiratory disease." Curr Opin Pediatr **28**(3): 339-347.

Mitchell, D. R. (2007). "The evolution of eukaryotic cilia and flagella as motile and sensory organelles." Adv Exp Med Biol **607**: 130-140.

Mitchell, D. R. (2017). "Evolution of Cilia." Cold Spring Harb Perspect Biol **9**(1).

Morgan, H. D., H. G. Sutherland, D. I. Martin and E. Whitelaw (1999). "Epigenetic inheritance at the agouti locus in the mouse." Nat Genet **23**(3): 314-318.

Nicetto, D. (2012). On the way to differentiation. PhD, Ludwig Maximilian Universität Munich.

Nicetto, D., M. Hahn, J. Jung, T. D. Schneider, T. Straub, R. David, G. Schotta and R. A. Rupp (2013). "Suv4-20h histone methyltransferases promote neuroectodermal differentiation by silencing the pluripotency-associated Oct-25 gene." PLoS Genet **9**(1): e1003188.

Nieuwkoop, P. D. (1967). Normal table of Xenopus Laevis (Daudin) a systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis. Amsterdam, North-Holland.

Nishioka, K., J. C. Rice, K. Sarma, H. Erdjument-Bromage, J. Werner, Y. Wang, S. Chuikov, P. Valenzuela, P. Tempst, R. Steward, J. T. Lis, C. D. Allis and D. Reinberg (2002). "PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin." Mol Cell **9**(6): 1201-1213.

Oda, H., I. Okamoto, N. Murphy, J. Chu, S. M. Price, M. M. Shen, M. E. Torres-Padilla, E. Heard and D. Reinberg (2009). "Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development." Mol Cell Biol **29**(8): 2278-2295.

Okabayashi, K. and M. Asashima (2003). "Tissue generation from amphibian animal caps." Curr Opin Genet Dev **13**(5): 502-507.

Owens, N. D., I. L. Blitz, M. A. Lane, I. Patrushev, J. D. Overton, M. J. Gilchrist, K. W. Cho and M. K. Khokha (2016). "Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development." Cell Rep **14**(3): 632-647.

Pan, J. H., T. L. Adair-Kirk, A. C. Patel, T. Huang, N. S. Yozamp, J. Xu, E. P. Reddy, D. E. Byers, R. A. Pierce, M. J. Holtzman and S. L. Brody (2014). "Myb permits

multilineage airway epithelial cell differentiation." Stem Cells **32**(12): 3245-3256.

Pennekamp, P., T. Menchen, B. Dworniczak and H. Hamada (2015). "Situs inversus and ciliary abnormalities: 20 years later, what is the connection?" Cilia **4**(1): 1.

Personnic, N., G. Lakisic, E. Gouin, A. Rousseau, A. Gautreau, P. Cossart and H. Bierne (2014). "A role for Ral GTPase-activating protein subunit beta in mitotic regulation." FEBS J **281**(13): 2977-2989.

Probst, A. V., E. Dunleavy and G. Almouzni (2009). "Epigenetic inheritance during the cell cycle." Nat Rev Mol Cell Biol **10**(3): 192-206.

Pugacheva, E. N., S. A. Jablonski, T. R. Hartman, E. P. Henske and E. A. Golemis (2007). "HEF1-dependent Aurora A activation induces disassembly of the primary cilium." Cell **129**(7): 1351-1363.

Qi, H. H., M. Sarkissian, G. Q. Hu, Z. Wang, A. Bhattacharjee, D. B. Gordon, M. Gonzales, F. Lan, P. P. Ongusaha, M. Huarte, N. K. Yaghi, H. Lim, B. A. Garcia, L. Brizuela, K. Zhao, T. M. Roberts and Y. Shi (2010). "Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development." Nature **466**(7305): 503-507.

Quigley, I. K. and C. Kintner (2017). "Rfx2 Stabilizes Foxj1 Binding at Chromatin Loops to Enable Multiciliated Cell Gene Expression." PLoS Genet **13**(1): e1006538.

Rackham, O. J., J. Firas, H. Fang, M. E. Oates, M. L. Holmes, A. S. Knaupp, F. Consortium, H. Suzuki, C. M. Nefzger, C. O. Daub, J. W. Shin, E. Petretto, A. R. Forrest, Y. Hayashizaki, J. M. Polo and J. Gough (2016). "A predictive computational framework for direct reprogramming between human cell types." Nat Genet **48**(3): 331-335.

Regnier, V., P. Vagnarelli, T. Fukagawa, T. Zerjal, E. Burns, D. Trouche, W. Earnshaw and W. Brown (2005). "CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1." Mol Cell Biol **25**(10): 3967-3981.

Reid, C. D., K. Karra, J. Chang, R. Piskol, Q. Li, J. B. Li, J. M. Cherry and J. C. Baker (2017). "XenMine: A genomic interaction tool for the Xenopus community." Dev Biol **426**(2): 155-164.

Robertson, K. D. (2005). "DNA methylation and human disease." Nat Rev Genet **6**(8): 597-610.

Sai, L., L. Li, C. Hu, B. Qu, Q. Guo, Q. Jia, Y. Zhang, C. Bo, X. Li, H. Shao, J. C. Ng and C. Peng (2018). "Identification of circular RNAs and their alterations involved in developing male Xenopus laevis chronically exposed to atrazine." Chemosphere **200**: 295-301.

Saksouk, N., E. Simboeck and J. Dejardin (2015). "Constitutive heterochromatin formation and transcription in mammals." Epigenetics Chromatin **8**: 3.

Sasai, Y., B. Lu, S. Piccolo and E. M. De Robertis (1996). "Endoderm induction by the organizer-secreted factors chordin and noggin in Xenopus animal caps." EMBO J **15**(17): 4547-4555.

Satir, P., L. B. Pedersen and S. T. Christensen (2010). "The primary cilium at a glance." J Cell Sci **123**(Pt 4): 499-503.

Schneider, T. D., J. M. Arteaga-Salas, E. Mentele, R. David, D. Nicetto, A. Imhof and R. A. Rupp (2011). "Stage-specific histone modification profiles reveal global transitions in the Xenopus embryonic epigenome." PLoS One **6**(7): e22548.

Schotta, G., M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg and T. Jenuwein (2004). "A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin." Genes Dev **18**(11): 1251-1262.

Schotta, G., R. Sengupta, S. Kubicek, S. Malin, M. Kauer, E. Callen, A. Celeste, M. Pagani, S. Opravil, I. A. De La Rosa-Velazquez, A. Espejo, M. T. Bedford, A. Nussenzweig, M. Busslinger and T. Jenuwein (2008). "A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse." Genes Dev **22**(15): 2048-2061.

Schuettengruber, B., H. M. Bourbon, L. Di Croce and G. Cavalli (2017). "Genome Regulation by Polycomb and Trithorax: 70 Years and Counting." Cell **171**(1): 34-57.

Session, A. M., Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, S. J. van Heeringen, I. Quigley, S. Heinz, H. Ogino,

H. Ochi, U. Hellsten, J. B. Lyons, O. Simakov, N. Putnam, J. Stites, Y. Kuroki, T. Tanaka, T. Michiue, M. Watanabe, O. Bogdanovic, R. Lister, G. Georgiou, S. S. Paranjpe, I. van Kruijsbergen, S. Shu, J. Carlson, T. Kinoshita, Y. Ohta, S. Mawaribuchi, J. Jenkins, J. Grimwood, J. Schmutz, T. Mitros, S. V. Mozaffari, Y. Suzuki, Y. Haramoto, T. S. Yamamoto, C. Takagi, R. Heald, K. Miller, C. Haudenschild, J. Kitzman, T. Nakayama, Y. Izutsu, J. Robert, J. Fortriede, K. Burns, V. Lotay, K. Karimi, Y. Yasuoka, D. S. Dichmann, M. F. Flajnik, D. W. Houston, J. Shendure, L. DuPasquier, P. D. Vize, A. M. Zorn, M. Ito, E. M. Marcotte, J. B. Wallingford, Y. Ito, M. Asashima, N. Ueno, Y. Matsuda, G. J. Veenstra, A. Fujiyama, R. M. Harland, M. Taira and D. S. Rokhsar (2016). "Genome evolution in the allotetraploid frog Xenopus laevis." Nature **538**(7625): 336-343.

Shnitsar, I., M. Bashkurov, G. R. Masson, A. A. Ogunjimi, S. Mosessian, E. A. Cabeza, C. L. Hirsch, D. Trcka, G. Gish, J. Jiao, H. Wu, R. Winklbauer, R. L. Williams, L. Pelletier, J. L. Wrana and M. Barrios-Rodiles (2015). "PTEN regulates cilia through Dishevelled." Nat Commun **6**: 8388.

Sive, H. L., R. M. Grainger and R. M. Harland (2007). "Animal Cap Isolation from Xenopus laevis." CSH Protoc **2007**: pdb prot4744.

Stelzer, G., N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran and D. Lancet (2016). "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." Curr Protoc Bioinformatics **54**: 1 30 31-31 30 33.

Stubbs, J. L., L. Davidson, R. Keller and C. Kintner (2006). "Radial intercalation of ciliated cells during Xenopus skin development." Development **133**(13): 2507-2515.

Stubbs, J. L., E. K. Vladar, J. D. Axelrod and C. Kintner (2012). "Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation." Nat Cell Biol **14**(2): 140-147.

Swoboda, P., H. T. Adler and J. H. Thomas (2000). "The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans." Mol Cell **5**(3): 411-421.

Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen and C. von Mering (2017). "The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible." Nucleic Acids Res **45**(D1): D362-D368.

Talhouarne, G. J. and J. G. Gall (2014). "Lariat intronic RNAs in the cytoplasm of Xenopus tropicalis oocytes." RNA **20**(9): 1476-1487.

Tan, F. E., E. K. Vladar, L. Ma, L. C. Fuentealba, R. Hoh, F. H. Espinoza, J. D. Axelrod, A. Alvarez-Buylla, T. Stearns, C. Kintner and M. A. Krasnow (2013). "Myb promotes centriole amplification and later steps of the multiciliogenesis program." Development **140**(20): 4277-4286.

Thomas, J., L. Morle, F. Soulavie, A. Laurencon, S. Sagnol and B. Durand (2010). "Transcriptional control of genes involved in ciliogenesis: a first step in making cilia." Biol Cell **102**(9): 499-513.

Todeschini, A. L., A. Georges and R. A. Veitia (2014). "Transcription factors: specific DNA binding and specific gene regulation." Trends Genet **30**(6): 211-219.

Towbin, B. D., C. Gonzalez-Aguilera, R. Sack, D. Gaidatzis, V. Kalck, P. Meister, P. Askjaer and S. M. Gasser (2012). "Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery." Cell **150**(5): 934-947.

Ueki, T., T. Nishidate, J. H. Park, M. L. Lin, A. Shimo, K. Hirata, Y. Nakamura and T. Katagiri (2008). "Involvement of elevated expression of multiple cell-cycle regulator, DTL/RAMP (denticleless/RA-regulated nuclear matrix associated protein), in the growth of breast cancer cells." Oncogene **27**(43): 5672-5683.

van Kruijsbergen, I., S. Hontelez, D. M. Elurbe, S. J. van Heeringen, M. A. Huynen and G. J. C. Veenstra (2017). "Heterochromatic histone modifications at transposons in Xenopus tropicalis embryos." Dev Biol **426**(2): 460-471.

van Nuland, R. and O. Gozani (2016). "Histone H4 Lysine 20 (H4K20) Methylation, Expanding the Signaling Potential of the Proteome One Methyl Moiety at a Time." Mol Cell Proteomics **15**(3): 755-764.

Vastenhouw, N. L., Y. Zhang, I. G. Woods, F. Imam, A. Regev, X. S. Liu, J. Rinn and A. F. Schier (2010). "Chromatin signature of embryonic pluripotency is established during genome activation." Nature **464**(7290): 922-926.

Vidarsson, H., R. Westergren, M. Heglind, S. R. Blomqvist, S. Breton and S. Enerback (2009). "The forkhead transcription factor Foxi1 is a master regulator of vacuolar H-ATPase proton pump subunits in the inner ear, kidney and epididymis." PLoS One **4**(2): e4471.

Vonica, A. and B. M. Gumbiner (2007). "The Xenopus Nieuwkoop center and Spemann-Mangold organizer share molecular components and a requirement for maternal Wnt activity." Dev Biol **312**(1): 90-102.

Walentek, P., C. Hagenlocher, T. Beyer, C. Muller, K. Feistel, A. Schweickert, R. M. Harland and M. Blum (2015). "ATP4 and ciliation in the neuroectoderm and endoderm of Xenopus embryos and tadpoles." Data Brief **4**: 22-31.

Wang, L., P. Joshi, E. L. Miller, L. Higgins, M. Slattery and J. A. Simon (2018). "A Role for Monomethylation of Histone H3-K27 in Gene Activity in Drosophila." Genetics **208**(3): 1023-1036.

Warner, S. M., T. L. Hackett, F. Shaheen, T. S. Hallstrand, A. Kicic, S. M. Stick and D. A. Knight (2013). "Transcription factor p63 regulates key genes and wound repair in human airway epithelial basal cells." Am J Respir Cell Mol Biol **49**(6): 978-988.

Werner, M. E. and B. J. Mitchell (2012). "Understanding ciliated epithelia: the power of Xenopus." Genesis **50**(3): 176-185.

Werner, M. E. and B. J. Mitchell (2013). "Using Xenopus Skin to Study Cilia Development and Function." Methods in enzymology **525**: 191-217.

Westholm, J. O., P. Miura, S. Olson, S. Shenker, B. Joseph, P. Sanfilippo, S. E. Celniker, B. R. Graveley and E. C. Lai (2014). "Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation." Cell Rep **9**(5): 1966-1980.

Wilson, P. A. and A. Hemmati-Brivanlou (1995). "Induction of epidermis and inhibition of neural fate by Bmp-4." Nature **376**(6538): 331-333.

Woodcock, C. L. and R. P. Ghosh (2010). "Chromatin higher-order structure and dynamics." Cold Spring Harb Perspect Biol **2**(5): a000596.

Yang, H., J. J. Pesavento, T. W. Starnes, D. E. Cryderman, L. L. Wallrath, N. L. Kelleher and C. A. Mizzen (2008). "Preferential dimethylation of histone H4 lysine 20 by Suv4-20." J Biol Chem **283**(18): 12085-12092.

Young, M. D., T. A. Willson, M. J. Wakefield, E. Trounson, D. J. Hilton, M. E. Blewitt, A. Oshlack and I. J. Majewski (2011). "ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity." Nucleic Acids Res **39**(17): 7415-7427.

Yu, X., C. P. Ng, H. Habacher and S. Roy (2008). "Foxj1 transcription factors are master regulators of the motile ciliogenic program." Nat Genet **40**(12): 1445-1453.

Zenk, F., E. Loeser, R. Schiavo, F. Kilpert, O. Bogdanovic and N. Iovino (2017). "Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition." Science **357**(6347): 212-216.

Zhang, R., J. Erler and J. Langowski (2017). "Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction." Biophys J **112**(3): 450-459.

Zimmerman, L. B., J. M. De Jesus-Escobar and R. M. Harland (1996). "The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4." Cell **86**(4): 599-606.

Zink, L. M. and S. B. Hake (2016). "Histone variants: nuclear function and disease." Curr Opin Genet Dev **37**: 82-89.

Zlotorynski, E. (2018). "Splicing: Going in circles." Nat Rev Genet **19**(2): 64-65.

Acknowledgements

First, I would like to thank my mentor Prof. Ralph Rupp for giving me the possibility of doing my PhD in his lab. Under his lead I grew up enormously, both scientifically and personally. I am grateful for the opportunities and the support he has given me.

I want to thank also the members of my TAC committee Prof. Peter Becker, Prof. Gunnar Schotta and Prof. Andre Braendli for the support and the many valuable suggestions.

I want to thank particularly our TAs, Barbara and Edith, for all the help they have given me and for all the laughs we had together. Thanks to all the other members of the Rupp lab: Daniil, Gabi, Christian, Julian, Adrian and Claudia for being my companions in these unforgettable years.

A big thanks to Sylvain, Andrea, Alessandra, Sophia, Viola, Nadia, Natalia, Thomas, Nicholas, Toby, Lisa, Lisa, Sara, Mana, Gustavo, Iris, Elisa, Ashish, Petra, Irina, Helena, Magdalini, Zeyang, Catherine, Filippo, Bo, to the “BMC coffee team” community and to all the members of the Becker department of present and past years. Working with you has been a privilege.

I thank a lot Elizabeth Schroeder-Reiter the coordinator of the IRTG for being a reference point, always positive and supportive.

Acknowledgment

Many people have accompanied me through my PhD. In particular, I want to thank Joséphа, for being the best thing that has happened to me in a long time. I am grateful to Gordon for all the adventures we lived together and the carefree time. To Andrea and Andrea for all the intellectual insights they share on Rostock and langlauf skifahren. To Elena, Giorgia, Nicola Silvia and Silvia for all the wonderful moments.

Moreover, I want to express my gratitude to my family for having supported me since when I can remember…