universität wien

MASTERARBEIT

Titel der Masterarbeit

"Allele-specific chromatin state in embryonic and extra-embryonic tissues"

verfasst von

Philipp Bammer, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 066 877
Studienrichtung lt. Studienblatt:	Masterstudium Genetik und Entwicklungsbiologie
Betreut von:	ao. Univ.-Prof. Dr. Christian Seiser

This work was done in the laboratory of Denise Barlow at CeMM
Research Center for Molecular Medicine
of the Austrian Academy of Sciences

Lazarettgasse 14, AKH BT 25.3, 1090 Vienna, Austria

I was supervised by Quanah Hudson, a senior post-doc in the lab.

TABLE OF CONTENTS

1. ZUSAMMENFASSUNG

Genomische Prägung beschreibt ein epigenetisches Phänomen welches zu Allel-spezifischer Genexpression abhängig vom elterlichen Ursprung führt. Geprägte Gene sind oftmals geclustert und assoziiert mit einer langen nicht kodierenden RNS (lnkRNS), welche Allel-spezifische Stilllegung eines ganzen Clusters bedingt. Die Expression dieser lnkRNS ist durch unterschiedliche Methylierung einer CpG reichen DNS Sequenz auf ein Allel begrenzt. Viele geprägte Gene zeigen ein deutliches Gewebe- und Entwicklungsstadium-spezifisches geprägtes Expressionsmuster, wobei ein Teil nur in extra-embryonalen Linien geprägt exprimiert wird. Diese Arbeit ist Teil eines größeren Projekts mit dem Ziel, geprägte Expression im Zusammenhang mit unterschiedlicher Verteilung von Histon Modifikationen in pluripotenten (embryonale Stammzellen (ES Zellen), embryonalen (fetale Leber und MEFs) und extra-embryonalen (viszerales Endoderm des Dottersacks (VE)) Zelltypen zu charakterisieren wofür RNS und Chromatin-Immunpräzipitation Sequenzierung (RNS-seq und ChIP-seq) von reziproken F1 FVB/N und CAST/EiJ Maus Kreuzungen durchgeführt werden. Zusätzlich habe ich erfolgreich vorherig von Feeder-Zellen abhängige Stammzellen an 2i Medium Kultur Bedingungen adaptiert und RNS von diesen Zellen isoliert. In diesem Projekt vergleiche ich die unterschiedliche Verteilung von Histon Modifikationen zwischen embryonalen und extra-embryonalen Linien (EXEL) und ihren Zusammenhang mit Unterschieden in geprägter Expression. Dazu habe ich Antikörper getestet (H3K4me3 und H3K27ac, H3K9me2, H3K9m3 und H3K27me3) und ChIP-seq gegen aktive Histon Modifikationen in MEFs, fetaler Leber und VE durchgeführt. Diese Daten habe ich mit Hilfe eines in unserem Labor entwickelten bioinformatischen Programms untersucht und dabei hauptsächlich bereits bekannte geprägte Gene mit unterschiedlicher Anreicherung von H3K4me3 und H3K27ac an der Promoter Region des aktiven Allels identifiziert. Ich konnte zeigen, dass die 3 Zelltypen (MEFs, fetale Leber und VE) definierte Muster von geprägter Genexpression aufweisen und dadurch bestätigen dass geprägte Expression stark gewebespezifisch ist.

1. ABSTRACT

Genomic imprinting describes the epigenetic phenomenon that leads to parental-of-origin dependent gene expression. Imprinted genes are often clustered and associated with a long non-coding (lnc) RNA that causes parental specific silencing of the entire cluster. The expression of these lncRNAs is restricted to one allele by parental specific methylation of a CpG-rich DNA sequence called the imprint control element (ICE). Many imprinted genes exhibit a distinct tissue specific and developmental stage-specific imprinted expression pattern, with a subset of genes showing extra-embryonic lineage (EXEL) specific imprinted expression. This project is part of a larger work aiming to characterize tissue-specific imprinted expression in pluripotent (embryonic stem (ES) cells), embryonic (fetal liver and MEFs) and extra-embryonic (visceral yolk sac endoderm (VE)) cell types, using RNA and chromatin immunoprecipitation sequencing (RNA-seq and ChIP-seq) of reciprocal F1 FVB/N and CAST/EiJ mouse crosses. Contributing to this, I successfully adapted feeder-dependent ES cells to feeder-independent 2i medium culture conditions, and isolated RNA from these cells for RNA-seq. In this project I compared differential enrichment of histone modifications between embryonic and extra-embryonic tissues and their relationship to differences in imprinted expression. I validated antibodies against histone marks (H3K4me3 and H3K27ac, H3K9me2, H3K9m3 and H3K27me3) and performed ChIP-seq to identify active marks in MEFs, fetal liver and VE. Using an imprinting pipeline developed in the lab, I screened for allele-specific enrichment at promoters. My data showed that genes that display imprinted gene expression also have enrichment of H3K4me3 and H3K27ac at the promoter of the active allele relative to the silent allele. Thus I could show that ChIP-seq can be used to validate imprinted expression detected by RNA-seq. I also showed that three different tissue types (MEFs, fetal liver and VE) have distinct patterns of imprinted expression, supporting claims that imprinted expression is highly tissue-specific.

2. INTRODUCTION

2.1 Genomic imprinting in mammals

Mammals are diploid and therefore possess 2 sets of each autosome and the vast majority of genes are equally expressed or repressed from both alleles. However, to date 150 genes (around 1% of the mouse genome) were reported to show parental allele-specific or imprinted expression (Williamson *et al.* 2014). Genomic imprinting describes the epigenetic mechanism that leads to parent-of-origin dependent expression and repression of certain genes and provides an explanation for some inheritance processes that do not follow Mendelian rules. It fulfills the criteria to be a classical epigenetic phenomenon by being mitotically heritable and leading to reversible alterations in gene expression without changing the underlying DNA sequence (Berger *et al.* 2009). Some imprinted genes are exclusively expressed in a monoallelic fashion from either the maternal or paternal inherited allele, but others show biased expression with one parental allele being more expressed than the other. Important to note at this point is that there is no agreement in the field at what level of bias should be called imprinted expression (Barlow 2011). Genomic imprinting does not only occur in mammals, but also evolved independently in insects and flowering plants (Ferguson-Smith 2011).

The importance of both sexes in mammalian reproduction was shown in the 1980s when mouse embryos with only maternal (parthenogenetic) or paternal (androgenetic) alleles were shown to be non-viable, demonstrating that both the maternal and paternal genomes are required and implying that genomic imprinting blocks parthenogenesis (McGrath & Solter 1983; Surani & Barton 1983; McGrath & Solter 1984; Surani *et al.* 1984). Despite the relatively small number of genes expressed in a monoallelic fashion discovered so far, many of them were characterized to be of crucial importance during embryonic and placental development and their effect was shown to be very dose dependent (Bartolomei & Ferguson-Smith 2011). Well-known examples are the

maternally expressed genes *Igf2r* and *H19* and the paternally expressed gene *Igf2* (Barlow *et al.* 1991; Bartolomei *et al.* 1991; DeChiara *et al.* 1991; Barlow 1995). Moreover, many were identified to be regulated in a spatial (tissue-specific) and temporal (certain developmental stages) specific manner, which may relate to their importance for normal development and growth (Santoro & Barlow 2011; Prickett & Oakey 2012).

Concerning the biological meaning and evolutionary origin of genomic imprinting, there exist a number of explanations, although the topic is still very controversial. In general, from an evolutionary point of view, genomic imprinting is a risky strategy, since at imprinted loci deleterious mutations on the active allele cannot be compensated by an intact gene on the silent allele (Prickett & Oakey 2012). Many imprinted genes were shown to be active in extra-embryonic tissues and in the brain, supporting the 'parental conflict hypothesis' (Moore & Haig 1991). This hypothesis proposes that there are different interests concerning resource allocation of the offspring between the mother and the father. While the father aims to maximize the nutrients assigned to his offspring, the mother has to counteract by carefully weighting between investment for the current and future litters (Moore & Haig 1991). Genes in the brain can act on behavior and genes in the placenta can act on nutrient allocation, both key factors during embryogenesis and early life to grow and develop optimally. For example, deletion of genes in the mouse GNAS cluster were linked to inability of offspring to suckle, leading to death (Cattanach 1986) and also *Peg3* and *Dlk1* have been found to influence suckling behavior (Curley *et al.* 2004; da Rocha *et al.* 2009). However, this explanation is challenged by the imprinting pattern of *Dopa decarboxylase* (*Ddc*), showing tissue-specific imprinted expression solely in heart and is therefore not related to resource allocation in any way (Menheniott *et al.* 2008). Miri and Varmuza (2009) proposed that some genes are only subjected to genomic imprinting because they are 'innocent bystanders' located near to an imprint control element (ICE) that leads them to become silenced by the same mechanism that targets the functionally important neighboring imprinted gene(s). They suggested that this explains imprinted expression of the *Slc22a2* and *Slc22a3* genes in the *Igf2r* cluster, which only

show imprinted expression in extra-embryonic tissues and whose null alleles show a mild phenotype under normal laboratory conditions (Miri & Varmuza 2009).

An alternative explanation is the 'ovarian time bomb hypothesis', that suggests that genomic imprinting arose to minimize the risk of formation of ovarian tumors caused by activation of an unfertilized oocyte (Varmuza & Mann 1994; Weisstein *et al.* 2002). This would explain why genes involved in growth are disproportionately subjected to genomic imprinting and also aims to explain why certain genes are maternally or paternally inactive. This hypothesis is in agreement with the imprinting status of the *Igf2* and the *Igf2r* genes that are maternally and paternally silent, respectively, indicating the general trend of growth enhancing genes being expressed from the paternal allele and growth inhibiting genes from the maternal allele (Weisstein *et al.* 2002).

2.2 Imprinted genes are often clustered

The majority of imprinted genes in mammals occur in clusters (Reik & Walter 2001), together with at least one long noncoding RNA (lncRNA, >200nt) and are associated with differential DNA methylation of an ICE, controlling the expression of the lncRNA. This clustering suggested that imprinting is not controlled individually for each gene, but the mechanism is more wide ranging and *cis*-acting (Barlow 2011). Indeed, by deletions of the ICE in 7 different imprinting clusters it was shown that this is the case. Aside from that, also solo imprinted genes were identified, which were (not yet) associated with any clustering (Hagiwara *et al.* 1997). This could be caused by a lack of knowledge about the imprinted expression status of neighboring genes or they could be genuine solo imprinted genes (Barlow 2011).

To date there are 2 different mechanisms described how allele-specific expression is achieved with the ICE being the master regulator in both of them. One relies on parental specific expression or repression of a *cis*-acting

lncRNA, which is responsible for silencing controlled by the methylation state of the ICE that lies within the promoter for this lncRNA (Weaver & Bartolomei 2013). Discovery of this mechanism provided for the first time an explanation for findings made earlier, that DNA methylation on one allele was correlated with expression of the protein-coding gene (Barlow 2011). LncRNAs are abundant in the mammalian genome (Derrien *et al.* 2011) and recent studies suggest a role in recruitment of chromatin modifiers to regulate gene expression (Guttman & Rinn 2012).

The second mechanism is via the zinc finger transcription factor CTCF (CCCTC-binding factor), binding the unmethylated ICE and acting as an insulator by blocking access to enhancers. The best-studied example for such a mechanism is the *Igf2* cluster, where *Igf2* and *H19* are neighboring genes, sharing enhancers. Binding of the CTCF insulator to the unmethylated ICE on the maternal chromosome blocks interactions with *Igf2* leading to preferential interaction of the enhancer with *H19*, which in turn leads to the expression of the *H19* ncRNA and not *Igf2.* On the paternal allele CTCF does not bind the methylated ICE allowing *Igf2* to access the enhancers and be transcribed, while *H19* cannot access the enhancers and is not transcribed (Bell & Felsenfeld 2000). Notably, *H19* is not involved in imprinted silencing, but a *trans*-silencing effect was reported due to the miRNA within the *H19* gene (Gabory *et al.* 2009).

2.3 From fertilization to birth- mouse embryogenesis

Genomic imprinting was shown to be of crucial importance during mouse development, but the mechanisms by which it is controlled are different in embryonic and extra-embryonic lineages (EXEL) that arise very early during development. Mouse embryogenesis starts with the fertilization of an oocyte arrested in the metaphase of the second meiotic division by the sperm. This triggers the arrested meiosis to be resumed and leads to formation of the female pronucleus that fuses with the already existing male pronucleus (Rossant & Tam 2002). The first cell division is finished by embryonic day (E)

1.5, followed by two more divisions, resulting in an 8-cell embryo at E2.5 (Fig. 1). At this stage, morphological changes take place (compaction) and the previously loosely attached cells become more compacted by maximizing cell contact to each other, leading to a smoother and spherical shape (Johnson 2009). At E3.0, the embryo is comprised of 16 cells, resembling the shape of a blackberry and therefore called morula (Fig. 1). This is followed by the process of cavitation at the 32-cell stage characterized by the formation of fluid-filled vesicles that later fuse and consequently form the expanding blastocoel cavity (Rossant & Tam 2002; Johnson 2009)

At E3.5 (early blastocyst) two cell types are distinguishable, the inner cell mass (ICM) clustered at one end, surrounded by an outer layer and the trophectoderm (EXEL). The ICM is shielded at this point from the blastocoelic fluid by trophoblastic processes and differentiates into the epiblast, a pluripotent cell population (Fleming *et al.* 1984). Some of the epiblast cells secrete fibroblast growth factor 4 (FGF4), leading to differentiation of nearby cells to primitive endoderm (EXEL), formed as a single cell layer (Tanaka *et al.* 1998). At late blastocyst stage, the primitive endoderm layer is clearly separated, isolating the epiblast from the blastocyst cavity (Fig. 1) and these cells are now restricted in their developmental potential and committed to certain lineages (Johnson 2009).

At late preimplantation stage (E4.5) the embryo is comprised of three distinct cell populations: the epiblast and two EXEL: the trophectoderm and the primitive endoderm (Rossant & Tam 2002). The embryo itself is derived from epiblast cells, except for a small part of the embryonic hindgut, which is formed from a derivative of the primitive endoderm, the visceral endoderm (Kwon *et al.* 2008). The primitive endoderm gives rise to the visceral and parietal endoderm, both extra-embryonic tissues. The trophectoderm differentiates exclusively into EXEL, namely the extra-embryonic ectoderm, ectoplacental cone and the primary and secondary giant cells (Rossant & Tam 2002). At late E5.0 the blastocyst hatches by bursting the zona pellucida that prevented it from attaching during its travel along the oviduct. This allows

it to be implanted into the maternal uterus, a receptive region of the uterine wall (Wang & Dey 2006).

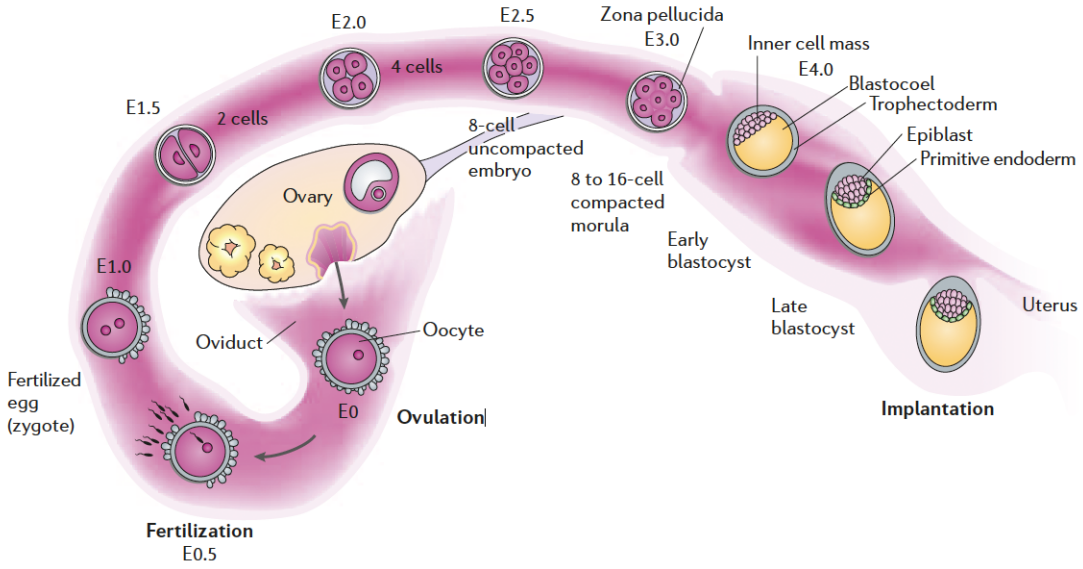

Fig. 1 Early mouse development from fertilization to implantation. The migration of the zygote through the oviduct towards the uterus is depicted showing the key steps during early development (image from Wang and Dey 2006).

Implantation is followed by gastrulation starting around E6.5 including reorganization of the epiblast and generation of the three germ layers ectoderm, mesoderm and endoderm (Downs & Davies 1993). It starts with formation of a node on the posterior side of the epiblast, a specialized organizer structure governing cell migration, fate choice and axis formation (Rossant & Tam 2002). Beta-catenin signaling initializes formation of the primitive streak, establishing bilateral symmetry and pattering the anterior-posterior axis (Mohamed *et al.* 2004).

At E8.0 organogenesis starts, the stage where the three germ layers form most of the organs and transcriptome studies revealed a major increase in transcript diversity marking this stage (Mitiku & Baker 2007). Most organs are not distinguishable at this time point, e.g. the heart, which does not start beating until after E8.5 and most organs function much later. The first organ to be formed during organogenesis is the placenta, which originates not from any of the three germ layers but from EXEL (Rossant & Cross 2001). Within the next three days the embryo is transformed from a cylindrical cup shape

into a more recognizable embryo-like shape with the formation of a heart, head, spinal cord and limbs (Nagy *et al.* 2003). By E12.5, the embryo is surrounded by three EXEL bilaminar membranes, namely the amnion, the visceral yolk sac (VYS) and the parietal yolk sac (PYS) (Hudson *et al.* 2010). Except for the amnion, these membranes interface with the placenta and the PYS is in contact with the uterine wall (Fig. 2). The VYS plays a major role in nutrient and oxygen exchange.

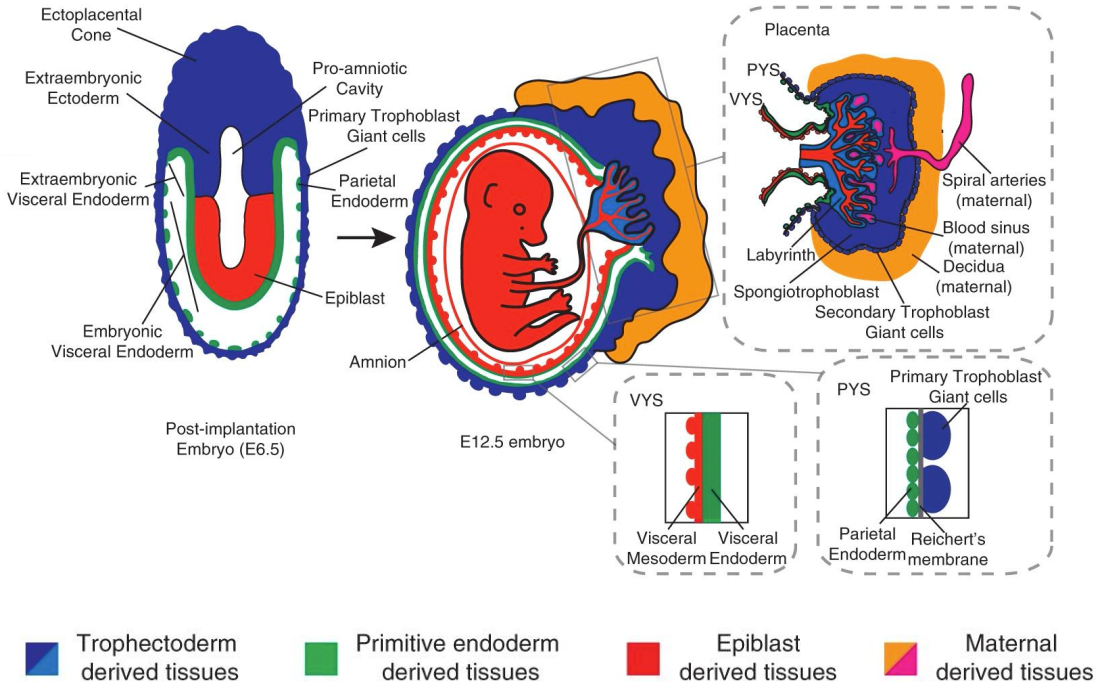

Fig. 2 Development of extra-embryonic tissues showing a post-implantation embryo (E6.5) and an E12.5 embryo (image from Hudson et al. 2010).

Birth occurs between E19.0-E21.0 and mice may live up to 2 years. The generation time is relatively short; it only takes ten weeks from being born to giving birth, which is one of the main advantages to use it as a model organism. Others are that there exist many inbred strains that enable controlled experiments, a high quality genome sequence is available and many well-characterized mutant animals were generated.

2.4 Extra-embryonic mouse tissues differ in their epigenetic environment

During mouse embryogenesis, extra-embryonic tissues safeguard the embryo by providing a protective, well-regulated and favorable environment for development and growth. Extra-embryonic tissues are discarded upon birth, meaning that the lifespan of extra-embryonic tissue is short relative to the two-year lifespan of a mouse. Moreover, embryonic lineages make up more complex and specialized tissues and differentiate to a much higher number of cell types during development. Based on the restricted lifespan and also taking into account the different developmental origin and function, one can speculate that there may be different requirements in terms of global gene regulation and overall chromatin state in extra-embryonic compared to embryonic tissue. Even though extra-embryonic tissues are lost upon birth, it is of great interest to understand the biology of these lineages, especially since they are not only needed to support the forming embryo, but are also involved in signaling and patterning the developing embryo (Rossant & Tam 2002).

During development the EXEL, the trophectoderm and the primitive endoderm, are the first cell lineages to differentiate starting at the blastocyst stage (E3.5). In contrast, in the epiblast, that later gives rise to embryonic lineages, no new lineages are distinguished until after the onset of gastrulation at E6.5 when the embryonic mesoderm and endoderm are formed. This temporal difference during development should be taken into account when features in gene regulation in embryonic and extra-embryonic tissues are investigated.

In general, the epigenetic hallmarks of extra-embryonic tissues resemble an early state of development throughout embryogenesis. After fertilization, DNA methylation marks, with the exception of those on ICEs, are globally removed and are restored only later in post-implantation embryos, where DNA methylation marks increase again (Smallwood & Kelsey 2012). Extra-

embryonic tissues are in general hypomethylated, retaining this early feature (Chapman *et al.* 1984; Rossant *et al.* 1986; Popp *et al.* 2010), and it has been suggested that histone modifications may play a greater role in gene regulation in these tissues. In particular, the H3K27me3 and H3K9me3 repressive histone modifications have been shown to be more associated with imprinted silencing in placenta than in embryo (Lewis *et al.* 2004; Umlauf *et al.* 2004; Hudson *et al.* 2010; Kulinski *et al.* 2013). The importance of repressive histone marks in maintaining imprinted silencing in extra-embryonic tissues has been demonstrated by defects in imprinted silencing for some genes in placenta when components of repressive histone modifying enzymes such as EHMT2 (G9a), PRC1 and PRC2 are knocked out (Umlauf *et al.* 2004; Nagano *et al.* 2008; Terranova *et al.* 2008; Wagschal *et al.* 2008). For example, genes in the *Kcnq1* cluster showing imprinted expression only in EXEL are biallelically expressed in placenta of knockout mice lacking EHMT2 and the same holds true for *Slc22a2* in the *Igf2r* cluster (Nagano *et al.* 2008; Wagschal *et al.* 2008).

Another epigenetic difference between embryonic lineages and EXEL relates to regulation of X-inactivation, which is imprinted in the early embryo by silencing the paternally derived X chromosome, which is then later reactivated and subjected to random X-inactivation in post-implantation embryos (Wutz 2011). However, in some EXEL the paternally derived X-chromosome remains preferentially silenced throughout development, again resembling the epigenetic regulation of the early embryo (Takagi & Sasaki 1975). In terms of dosage sensitivity to gene expression, EXEL were shown to tolerate polyploidy to a larger extent compared to embryonic tissues, where it leads to abortion of embryogenesis (Tarkowski *et al.* 1977). In fact, trophoblast giant cells of the PYS and the placenta are normally polyploid. Experiments generating chimeric embryos consisting of tetraploid and diploid cells showed that extra-embryonic tissues are mainly made of tetraploid cells as opposed to the embryo, which is almost exclusively diploid (Tarkowski *et al.* 1977). These findings indicate that dosage regulation may not be so crucial in extra-embryonic tissues, although it is not clear that the dose would not also be regulated in polyploidy cells.

2.5 Genomic imprinting in extra-embryonic tissues

Many imprinted genes display a distinct tissue-specific and developmental stage-specific imprinted expression pattern, with a subset of genes showing imprinted expression only in extra-embryonic tissues, such as the placenta and the VYS. This is why imprinted expression can be sub-divided into genes showing multi-lineage (ML) imprinted expression and genes showing imprinted expression only in EXEL (Hudson *et al.* 2011). Genome wide analyses examining a full range of mouse tissues are missing, but in imprinted clusters there is the general trend that genes showing EXEL imprinted expression tend to be further away ('outer' genes) from the ICE compared to genes that show ubiquitously imprinted expression ('inner' genes) (Lewis *et al.* 2004; Terranova *et al.* 2008; Hudson *et al.* 2010). This holds true for both mechanisms controlled by an insulator mechanism, such as within the *Igf2* cluster where the *Ins2* gene shows imprinted expression only in VYS (Giddings *et al.* 1994), but also in clusters controlled by lncRNA expression, for example, the *Igf2r* cluster where the genes *Slc22a2* and *Slc22a3* are subjected to imprinted expression only in EXEL tissues (Sleutels *et al.* 2002; Hudson *et al.* 2011).

The most parsimonious explanation for these tissue-specific observations is that the epigenetic initiator (lncRNA or insulator complex) extends its range by an unknown mechanism and is capable of silencing more genes (Fig 3a, c) (Berger *et al.* 2009; Hudson *et al.* 2010). This could be explained by increased sensitivity to the initiator in EXEL tissues by some unknown mechanism (Hudson *et al.* 2010). One hypothesis to explain the extended silencing domain is that DNA hypomethylation in EXEL leads to more open chromatin, allowing the epigenetic initiator to act over a greater distance (Kulinski *et al.* 2013), in line with reports that imprinted lncRNAs form a larger RNA cloud in placenta than in the embryo (Morison *et al.* 2001). An alternative explanation is that in EXEL tissue-specific enhancers are blocked by an epigenetic initiator such as a lncRNA leading to imprinted silencing of genes activated by these enhancers only in EXEL (Pauler *et al.* 2012).

Most, but not all, genes that show EXEL-specific imprinting are expressed from the maternal allele. For example, 16/18 EXEL genes listed in Hudson et al. (2010) show maternal imprinted expression (Hudson *et al.* 2010). In contrast, in brain this trend is not found, with three paternally and five maternally expressed imprinted genes whose imprinted expression is specific to brain (Prickett & Oakey 2012). Although, this list is far from being complete and new imprinted genes are frequently published, these trends indicate that some feature of EXEL tissues may favor maternal over paternal imprinted expression.

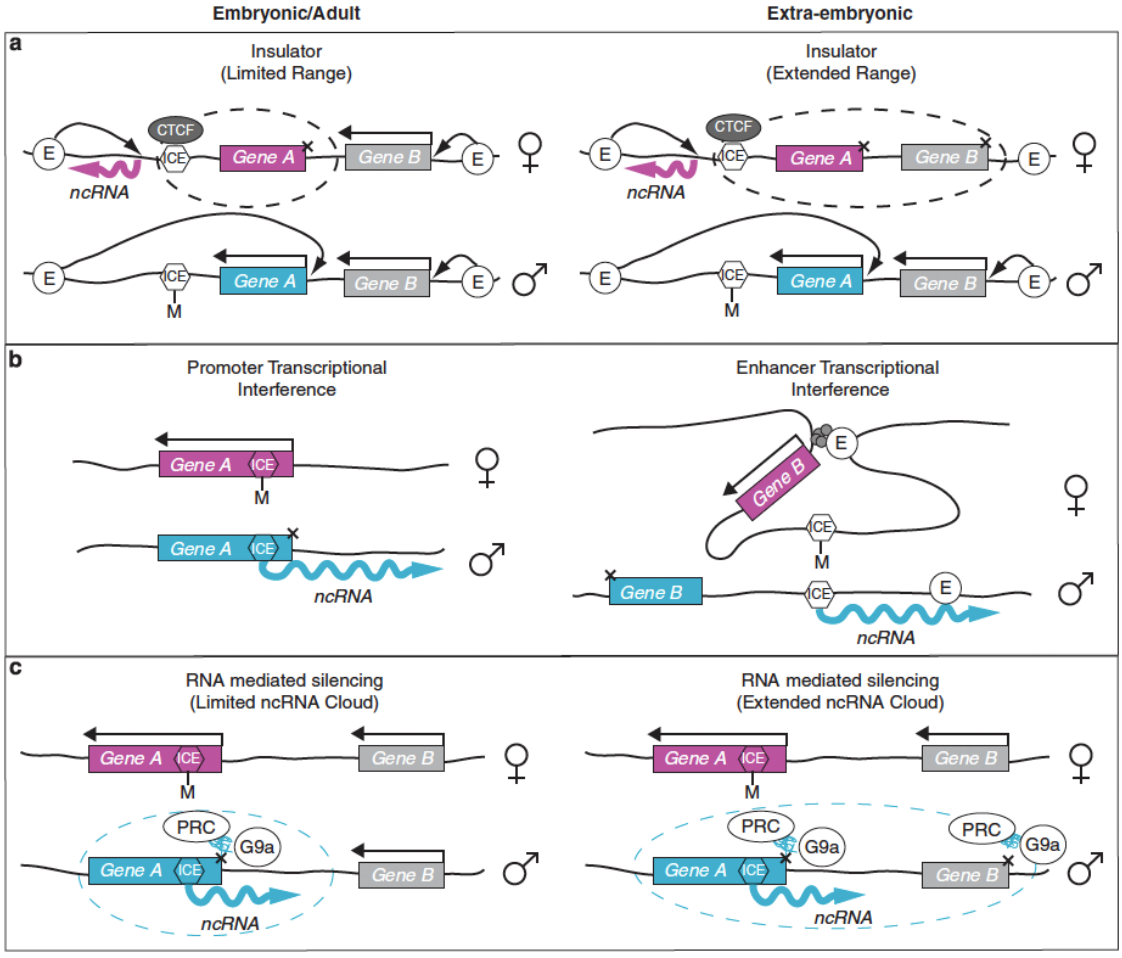

Fig. 3 Mechanisms aiming to explain differences in imprinted gene expression between embryonic and extra-embryonic tissues (a) (c) The initiator extends its range and is capable of silencing more genes In EXEL (b) A tissue specific enhancer is disrupted by expression of the ncRNA in EXEL preventing formation of a loop to bring promoter and enhancer in close proximity (image from Hudson et al. 2010).

2.6 The advantage of using visceral yolk sac endoderm over placenta

In order to investigate regulation of imprinted expression in extra-embryonic tissues it is important to have an appropriate model tissue. Earlier studies have mostly been conducted in placenta as a representative of extra-embryonic tissue, mostly because it is easily accessible and can be harvested at relatively large amounts (Umlauf *et al.* 2004; Green *et al.* 2007). Nevertheless, using placenta has several disadvantages when investigating cell lineage-specific imprinting. First of all, the placenta is a complex tissue originating from multiple embryonic lineages. Since imprinted expression is very tissue-specific, this mixture can lead to masking of imprinted gene expression if a gene shows imprinted expression in one cell type and biallelic expression in another (Kulinski *et al.* 2013). Another problem is that some genes that show imprinted expression are only expressed in only one lineage, which alleviates the problem of imprinted expression masking, but makes interpretation of DNA methylation and histone modification data difficult where the whole placenta is examined. Moreover, placenta is not a pure extra-embryonic tissue, since it is contaminated with maternal derived tissue, such as blood cells, blood vessels and decidua and it is not feasible to remove all the maternal tissue by dissection. Therefore, special precautions need to be taken to validate studies of imprinted expression and regulation in placenta (Proudhon & Bourc'his 2010; Hudson *et al.* 2011; Okae *et al.* 2012).

An alternative model to study imprinted gene expression in extra-embryonic tissues to placenta is the VYS. It is a simple tissue, comprised of only two layers, namely the visceral yolk sac endoderm (VE) and the visceral yolk sac mesoderm (VM), which includes the blood islands (Fig. 4). This tissue has the added advantage that if necessary the VE layer, which shows EXEL imprinted expression whereas the VM does not, can be efficiently isolated to give a highly homogenous population of cells using a technique developed in the lab (Hudson *et al.* 2011). In the present work, VE was used as the model EXEL tissue to compare to embryonic liver and mouse embryonic fibroblasts (MEFs)

to screen for tissue specific imprinted gene expression and associated differences in histone modifications (Fig. 4).

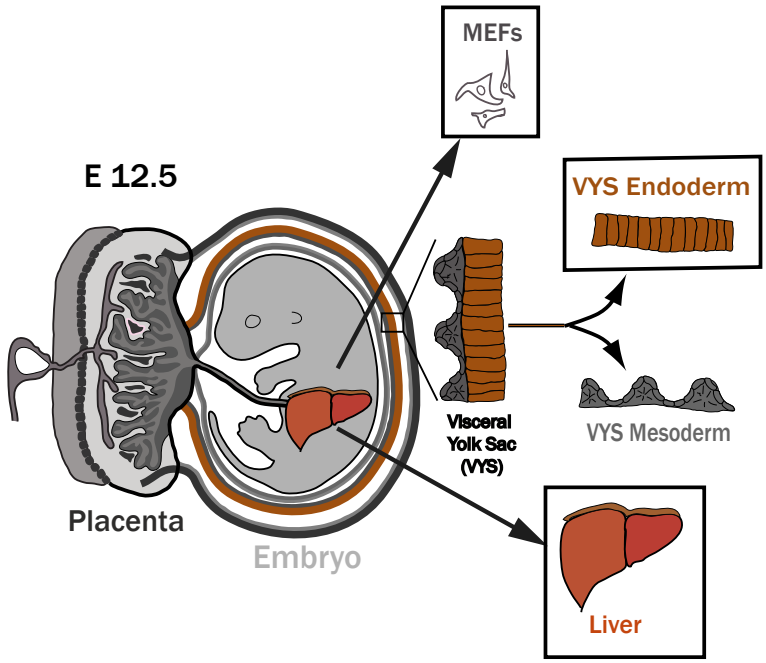

Fig. 4 In the present study E12.5 mouse embryos were dissected to study both, embryonic and extra-embryonic tissue-specific imprinted gene regulation. Both layers of VYS were mechanically separated and the VE layer was collected. In addition, embryonic liver and MEFs were used to screen for tissue specific imprinted gene expression (image adapted from Hudson et al. 2011).

2.7 Regulation of gene expression by histone modifications

To achieve the necessary 10000- 20000-fold compaction of chromatin to fit it into the nucleus, DNA is organized into nucleosomes and also tightly packed by higher order folding (Woodcock & Ghosh 2010). In the eukaryotic cell, 147bp of DNA (Luger *et al.* 1997; Richmond & Davey 2003) are tightly wrapped around a histone octamer comprised of two copies each of four different core histones, namely H2A, H2B, H3 and H4. However, the compact packing must still allow access for the transcriptional machinery, which is controlled by mechanisms including covalent modifications of histone N-terminal tails. This can increase or decrease accessibility of the chromatin depending on the modification. The amount of newly discovered histone modifications is rapidly growing and to date more than 100 distinct histone modifications have been identified, ranging from prominent ones like

acetylation, methylation and phosphorylation to less well studied ones such as citrullination and crotonylation (Zentner & Henikoff 2013).

These post-translational modifications can have an impact on gene transcription by controlling accessibility to the DNA. In general, genes more loosely bound to nucleosomes are more frequently transcribed, an effect that comes with acetylation by neutralizing the positive charge of lysine and as a result relaxing DNA-histone binding. The effect of methylation is not so straightforward; it can either be activating (H3K4, H3K36, H3K79) or repressive (H3K9, H3K27, H4K20). Furthermore there is growing evidence that histone methylation has additional functions and is for example involved in stabilizing the nucleosomes (H3K36me3) (Zentner & Henikoff 2013).

Genome wide chromatin immunoprecipitation (ChIP) coupled with microarray analysis (ChIP-chip) or high throughput sequencing (ChIP-seq) has significantly improved our understanding of the effect of distinct histone modifications by linking them to their biological role and helped to identify *cis*-regulatory elements such as promoters and enhancers. Promoters are associated with focal peaks of H3K4me3 while H3K27ac together with H3K4me1 mark active enhancers (Creyghton *et al.* 2010). On the other hand, the modifications H3K27me3, H3K9me2 and H3K9me3 have been associated with the repression of gene expression, including in imprinted gene silencing as mentioned previously. Therefore, in the present study I am using ChIP-seq to compare the genomic distribution of these active and repressive histone modifications in embryonic and extra-embryonic tissues, particularly in relation to imprinted genes, in order to gain some insight into the regulation of imprinted expression in these different lineages.

2.8 Tissue-specific and developmental stage-specific imprinted gene expression in the *Igf2r cluster*

The *Igf2r/Airn* imprinted region is located on mouse chromosome 17 and is extensively used as an epigenetic discovery model since it shows all common

features of imprinted gene clusters (see 2.2). The *Igf2r* gene was the first gene shown to have imprinted expression (Barlow *et al.* 1991), providing an explanation for the parental origin effect described for T-hairpin (T^{hp}) mice (Johnson 1974). These mice have a hairpin tail phenotype caused by haploinsufficiency of *Bry*, which is included within the approximately 5Mb deletion on chromosome 17 that also includes the *Igf2r* cluster (Johnson 1974; Barlow *et al.* 1991). Offspring that maternally inherit the T^{hp} allele die around E15.0, whereas paternally inherited mice are born and show no abnormalities apart from the tail phenotype. This is caused by the fact that the *Igf2r* gene is expressed monoallelically from the maternal chromosome and silenced on the paternal chromosome in almost all tissues and developmental stages. The lncRNA *Airn* (**a**ntisense to ***I****gf2r* **R**NA **n**on-coding) is expressed in a reciprocal pattern to protein coding genes in the cluster and acts *in cis* by repressing genes on the same parental chromosome (Wutz *et al.* 1997; Lyle *et al.* 2000; Sleutels *et al.* 2002). *Airn* is expressed from the paternal allele, transcribed in antisense over the *Igf2r* promoter silencing it by transcriptional interference (Fig. 7) (Latos *et al.* 2012). *Airn* is an unusual RNA polymerase II transcript being 118kb long, mostly inefficiently spliced (less than 5% are spliced to isoforms of around 1kb), unstable and not exported to the cytoplasm (Seidl *et al.* 2006). Besides of silencing the *Igf2r* gene, *Airn* was also shown to silence two additional genes, *Slc22a2* and *Slc22a3* by a long-range *cis*-silencing mechanism only in EXEL tissues such as placenta and VYS (Fig. 7) (Sleutels *et al.* 2002; Hudson *et al.* 2011).

Currently there exist two models to explain this silencing effect, one from the published data from Nagano et al. (2008), that indicates that *Airn* interacts with the histone methyltransferase EHMT2 and targets it to the *Slc22a3* promoter, which leads to deposition of repressive histone marks on lysine 9 (Nagano *et al.* 2008). However, this model cannot explain tissue specificity because both, *Airn* expression and also histone modifying enzymes are not tissue-specific. An alternative model was proposed by Pauler et al. (2012), suggesting the existence of a tissue-specific enhancer that is disrupted on the paternal allele by expression of *Airn* (Pauler *et al.* 2012). Under this model, on the maternal chromosome a loop is formed, allowing the enhancer to interact

with the promoters of both genes *Slc22a2* and *Slc22a3*, leading to expression. On the paternal chromosome, the enhancer is disrupted by transcriptional interference by *Airn* and as a consequence formation of the loop is inhibited, preventing upregulation of the genes (Fig. 5).

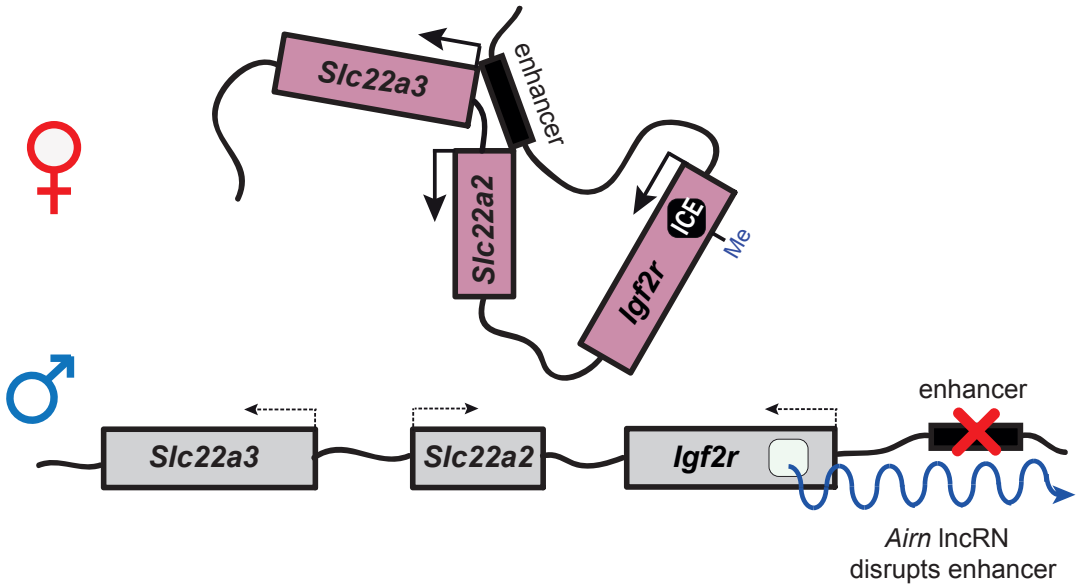

Fig. 5 Enhancer-interference model with a loop formed on the maternal chromosome allowing the enhancer to interact with the promoter of both genes *Slc22a2* and *Slc22a3* and inhibition of loop formation by *Airn* on the paternal chromosome preventing upregulation of *Slc22a2* and *Slc22a3*. Adapted from Pauler et al. (2012).

To explain the formation of the loop, and why it is not preformed before the onset of *Airn* expression, Pauler et al. (2012) propose the existence of 'transcriptional activators' that bind at the post-implantation stage. They are needed for loop formation and are not present earlier during development. On the paternal allele, *Airn* transcription prevents this binding, leading to no upregulation of *Slc22a2* and *Slc22a3* expression (Fig. 6) (Pauler *et al.* 2012). As a second layer of silencing to lock in the imprinting state, EHMT2 is recruited (Nagano *et al.* 2008) and in turn PRC1 and PRC2 are attracted, depositing H2AK119u1 and H3K27me3, respectively. This leads to formation of a repressive compartment bringing *Airn*, EHMT2 and the promoters of *Slc22a2* and *Slc22a3* in close physical proximity (Fig. 6) (Terranova *et al.* 2008; Pauler *et al.* 2012). Support for this model comes from chromatin conformation capture (3C) and ChIP-seq data, which shows maternal specific loops between EXEL gene promoters and elements within the *Airn* gene body

correlating with maternal specific enrichment of enhancer marks only in EXEL tissues (Quanah Hudson, unpublished data).

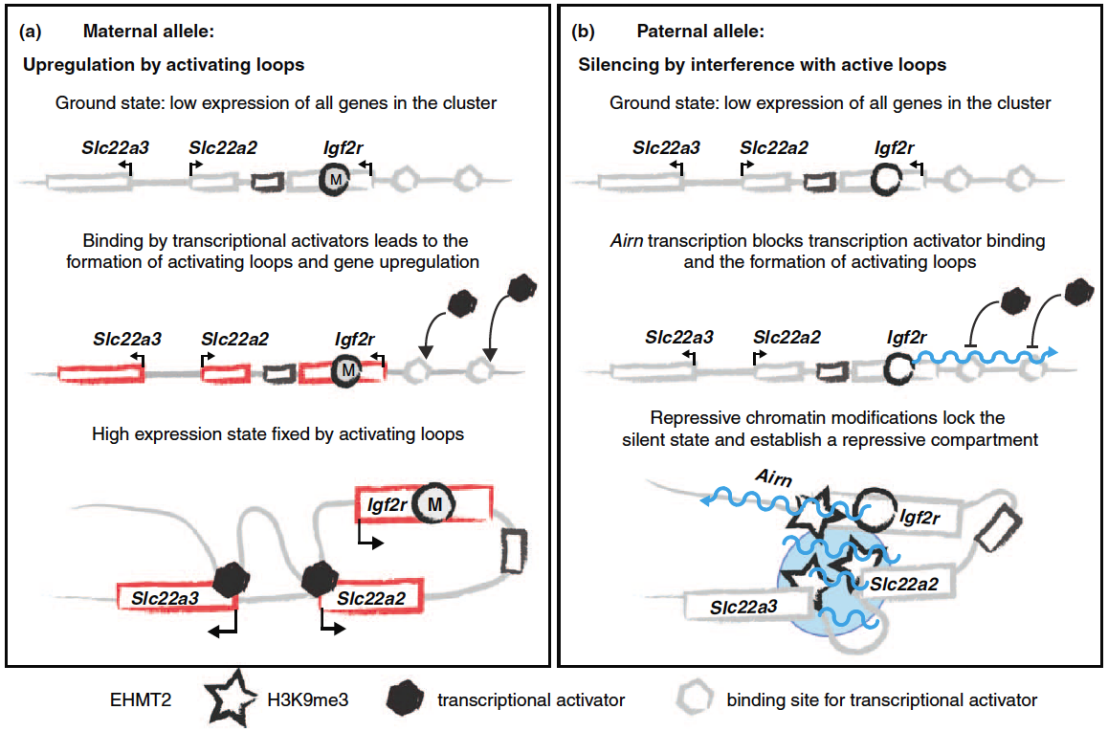

Fig. 6 Establishment of parental specific gene expression by chromosomal looping. Transcriptional activators bind to the maternal chromosome and induce the formation of a activating loop, whereas on the paternal loop, transcription of *Airn* inhibits binding of these transcriptional activators. As a second step the repressive chromatin-modifying enzyme EHMT2 deposits H3K9me2/me3 (image from Pauler et al. 2012).

The *Igf2r* cluster does not only show a distinct tissue-specific, but also a developmental stage-specific imprinted expression pattern. In undifferentiated embryonic stem (ES) cells, none of the genes in the cluster is expressed in an imprinted way and *Airn* is not expressed allowing *Igf2r* to be expressed biallelically (Fig. 7). In preimplantation embryos basal levels of *Igf2r*, *Slc22a2* and *Slc22a3* transcription are detectable and expression is not upregulated before post-implantation (Szabo & Mann 1995; Lerchner & Barlow 1997; Hudson *et al.* 2011). The maternal imprint on the promoter of *Airn* prevents it from being transcribed, but *Airn* is also not expressed from the paternal allele, most likely because of missing transcription factors at this stage (Latos *et al.* 2009). Interestingly, *Slc22a3* loses its imprinted expression and switches to biallelic expression around E15.5 in both placenta and VYS (Zwart *et al.* 2001;

Hudson *et al.* 2011), indicating *Slc22a3* may be less sensitive to the silencing effect of *Airn* at later development stages (Nagano *et al.* 2008).

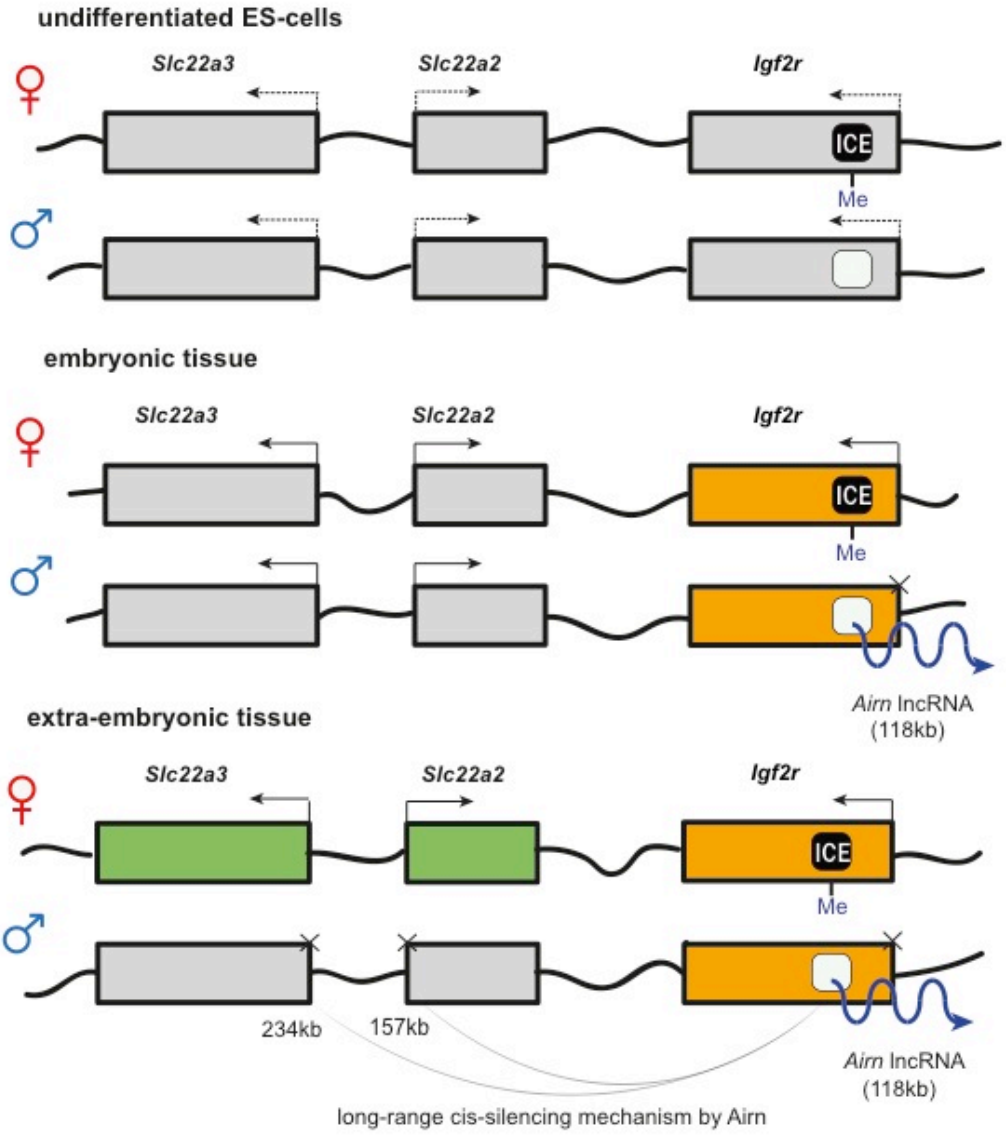

Fig. 7 Expression state in the *Igf2r* cluster in (A) undifferentiated ES cells, (B) embryonic and (C) extra-embryonic tissues. In undifferentiated ES cells imprinted gene expression is not yet established and all genes are expressed biallelically at basal levels. In embryonic tissue *Igf2r* is silenced by transcriptional interference by the lncRNA *Airn* on the paternal allele. In extra-embryonic tissues in addition to *Igf2r* silencing, *Airn* initiates long-range *cis*-silencing of *Slc22a2* and *Slc22a3*. The differential DNA methylation mark on the ICE is present at all developmental stages.

2.9 Aim of this project

The present work is part of a larger project led by Quanah Hudson with the aim to characterize imprinted gene expression and regulation in different cell

lineages, namely pluripotent (ES cells), embryonic (E12.5 fetal liver and MEFs) and extra-embryonic (E12.5 VE) cell types. To address this question, both RNA-seq (mainly carried out by Daniel Andergassen, a former diploma student in the Barlow lab) and ChIP-seq (my project) are being used. To distinguish the parental alleles we are using reciprocal crosses of the FVB/N and CAST/EiJ mouse strains that have over 20 million high quality single nucleotide (SNPs) between them that have been annotated by the Sanger Institute (Keane *et al.* 2011). These SNPs are incorporated into a bioinformatics pipeline conceived by Daniel Andergassen and developed to completion by Christoph Dotter (current masters student in the Barlow lab) that detects imprinted expression and parental allele-specific enrichment of histone modifications from RNA-seq and ChIP-seq data.

My first aim was to collect biological material for the project. Conventional ES-cell culture is feeder dependent, and contamination from these feeders would complicate analysis of imprinted expression in ES cells. To solve this problem I adapted previously feeder dependent FVB/CAST ES cells to feeder independent and serum free culture conditions using 2i medium, and I was able to collect RNA from these cells for RNA-seq. In order to examine differential enrichment of active and repressive histone modification in embryonic liver and extra-embryonic VE, which was the main aim of my project, it was necessary to collect large amounts of material (about 20-40 embryos necessary per antibody, 2 x biological replicates needed) from reciprocal crosses of FVB/N and CAST/EiJ mice. This took longer than expected due to the poor breeding success of these crosses, so I was not able to complete all of the ChIP-seq experiments that I initially planned.

The second goal of my project was to validate ChIP-antibodies for the histone modifications (H3K4me3, H3K27ac, H3K9me2, H3K9me3, H3K27me3) using MEF cell lines (MEFF and MEFB1 cells). Furthermore, it was necessary to optimize the amount of chromatin required for a successful ChIP and also the amount of antibody necessary. Each ChIP was validated with qPCR assays using known positive and negative controls and paying particular attention to the well-characterized *Igf2r* cluster. Antibodies found to be performing well,

were then further tested on wild type mouse tissues before conducting ChIP on material from the FVB x CAST reciprocal crosses for ChIP-seq.

The final part of my project was to perform ChIP-seq for the active H3K4me3 and H3K27ac modifications on embryonic (MEFs and fetal liver) and extra-embryonic tissues (VE) from FVB x CAST reciprocal crosses. I then used the bioinformatics pipeline mentioned earlier to examine differential enrichment of these marks on promoters. I found that, using this data set, parental allele-specific enrichment of H3K4me3, and to a lesser extent H3K27ac enrichment, correlated with imprinted gene expression indicating that these marks can be used as a proxy of imprinted expression.

Future work, outside the scope of this thesis, will use the repressive histone modification antibodies that I have validated, and the FVB/CAST embryonic material that I have collected, to conduct ChIP-seq and determine if these modifications are associated in a parental specific manner with imprinted gene silencing.

3. RESULTS

3.1 Optimization of feeder independent and serum free ES cell culture

The main aim of my thesis was to examine if there are differences in differential enrichment of histone modifications at imprinted loci between embryonic and extra-embryonic tissues. However, this is part of larger study in the laboratory that compares pluripotent, embryonic and EXEL for differences in imprinted expression and differential enrichment of histone modifications. In this section I describe my work developing tissue culture conditions for FVB/CAST and CAST/FVB F1 mouse ES cells that represent the pluripotent lineage, as part of my contribution to this larger project (ES cells were derived for us by Christian Theußl and Daniel Wenzel (from the laboratory of Josef Penninger at IMBA, Vienna) based on a published protocol (Bryja *et al.* 2006).

In order to examine imprinted expression in undifferentiated ES cells it is necessary to both have a robust method of maintaining the ES cells in an undifferentiated state, and to overcome the problem of feeder cell contamination, which may complicate analysis. To address this issue I adapted ES cells derived from reciprocal F1 crosses of FVB/NJ and CAST/EiJ that were previously grown in serum-supplemented medium on a feeder layer, to serum and feeder-free culture conditions using 2i medium. Furthermore, I determined the passaging procedure for these cells to maintain appropriate cell density and ensure optimal growth.

2i medium is a serum-free medium where nutrients and growth factors are provided together with a combination of glycogen synthase kinase 3 beta (GSK3B) and Mek1/2 inhibitors to keep the cells in an undifferentiated, pluripotent state (Ying *et al.* 2008). In addition to dispensing with the problem of feeder contamination, growing ES cells in 2i media provides several advantages over culturing in serum on feeders. By interfering with pathways

involved in differentiation, 2i medium is selective against differentiating cells and therefore leads to a more homogeneous and undifferentiated cell population compared to cells grown in serum-supplemented medium. Furthermore, as a defined media, lot-to-lot variations between different 2i batches should be kept to a minimum compared to variation between batches of serum, allowing more reproducible results.

3.1.1 Adaptation of feeder dependent mouse ES cells to 2i media

According to the adaptation protocol of ESGRO-2i medium (Millipore), cells should be first grown in serum-supplemented medium after thawing and at the first splitting seeded in 2i-medium. However, this adaptation was observed not to be necessary (Tomasz Kulinski, personal communication) and therefore I tested both, sticking to the adaptation protocol and seeding them directly into 2i medium. In general, cells seeded directly after thawing into 2i medium attached to a much lesser extent and many dying cells were observed. In contrast, cells plated onto a feeder layer in ES cell media with serum attached very well. This is as expected, since cells attach better to a feeder layer than directly to a gelatinized dish (Fig. 8).

Cells thawed onto ES cell media with serum or directly into 2i media showed both growth, but the ES cell colony morphology looked better for ES cells thawed in serum containing media, with colonies showing a distinctive round shape with a bright edge (Fig. 8). 2i medium is a selective medium against differentiated cells and it seems there is a strong selection process going on when cells are plated directly into 2i medium without any adaptation. However, after the first passage when cells were re-seeded into 2i media, both cells thawed into ES cell media with serum and those thawed directly into 2i, grew well and had good ES cell colony morphology.

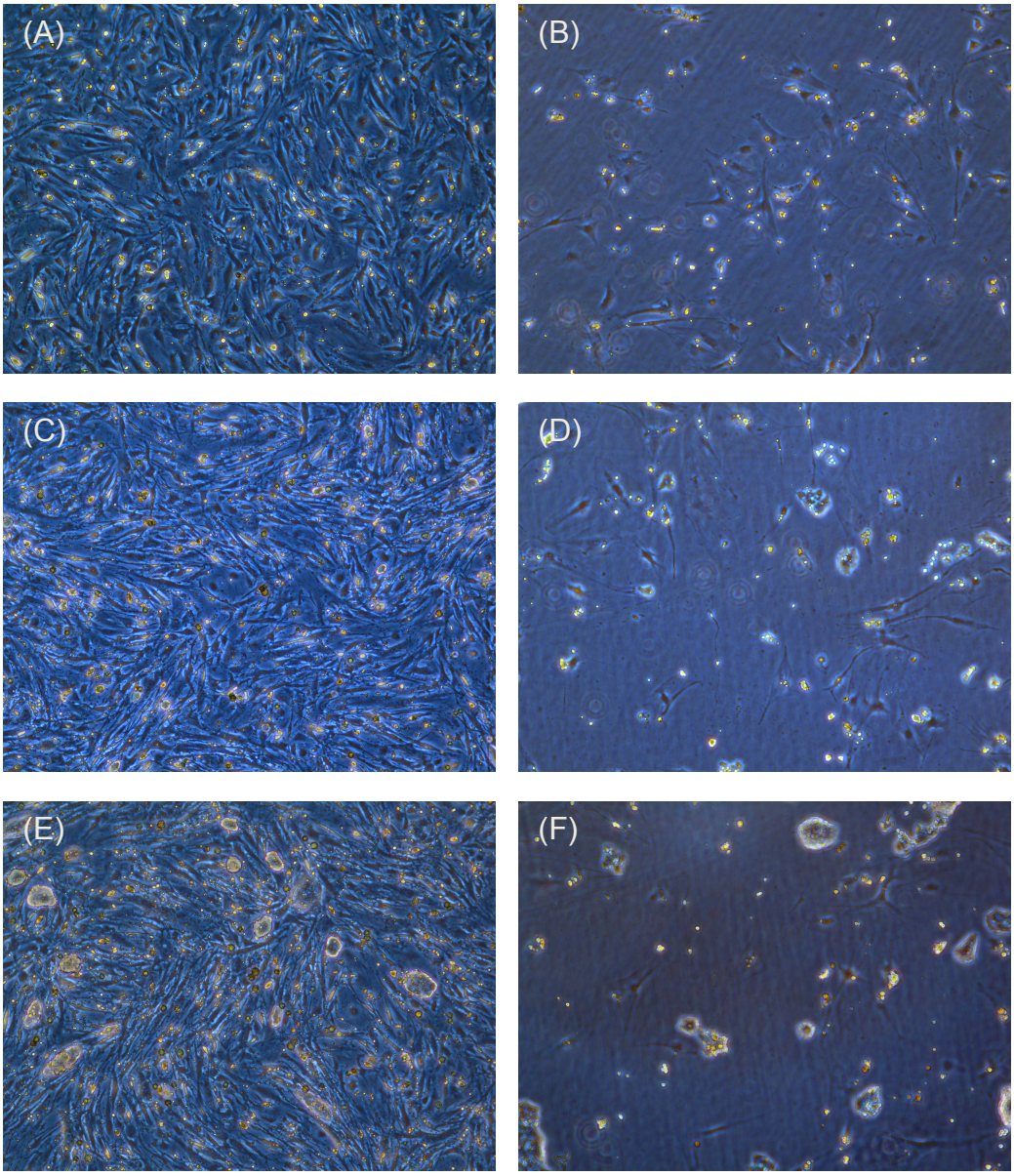

Fig. 8 Thawed ES cells approximately 24h (A and B), 48h (C and D) and 72h (E and F) after plating. (A), (C), (E) ES cells plated on a 6cm gelatinized dished containing a feeder-layer in ES cell media with serum (B), (D), (F) ES cells plated on a 6cm gelatinized dish in 2i medium (some feeder cells remain present as the thawed cells were grown on a feeder layer before freezing). Considerably more cells attached when cells were seeded on a feeder layer and the cells appeared to have better ES cell morphology when grown in ES cell medium with serum (representative images shown).

3.1.2 Splitting cells

Determining the appropriate splitting ratio for ES cells is required to allow optimal growth conditions for undifferentiated ES cells. On the one hand, colonies of ES cells seeded too densely start growing into each other, which induces cells to differentiate. On the other hand, it is necessary to have a

certain minimum cell number on a dish, because too sparsely seeded ES cells stop growing. After determining this optimal splitting ratio my aim was to determine how long to grow the cells to obtain a dense dish in order to maximize the yield of RNA or chromatin from one dish.

Splitting was done in general after two days using both trypsin and accutase, which is recommended for 2i medium culture conditions (adaptation protocol ESGRO-2i MEDIUM, Millipore). Normally, the serum contained in the ES cell medium stops trypsin activity, but since 2i medium culture is serum free, the enzyme was stopped by pelleting the cells and removing all medium. This worked to a certain extent, but some problems with cell attachment were observed, as cells started detaching at the edges of the dishes one day after seeding them. This problem was no longer present when accutase was used to disassociate the cells. Furthermore, by using accutase the amount of cells required for seeded to obtain a dense dish could be reduced by at least half, simply because of the fact that more cells reattached. $5x10^5$ cells seeded on a 6cm culture dish showed the best results (Fig. 9), and while lower seeding densities down to $1x10^5$ cells also grew fine, the density was quite low (not shown). In general, feeder cells were depleted after 2 passages.

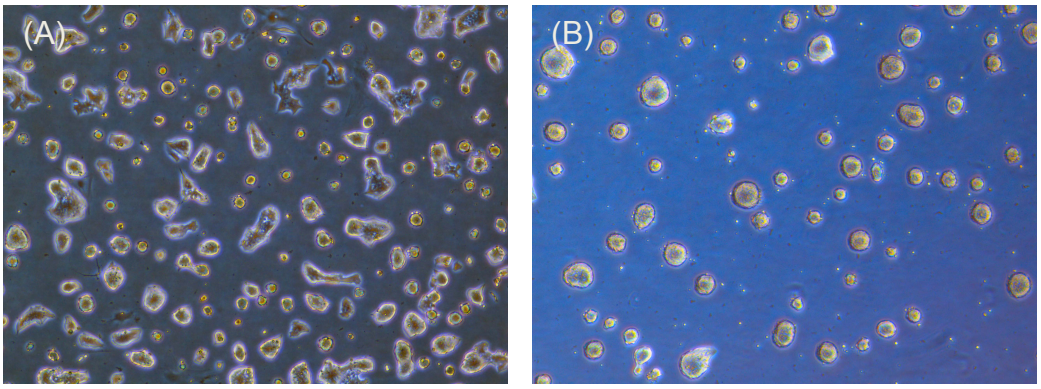

Fig. 9 ES cells thawed directly in 2i-medium about 48h (A) and 96h (B) after first passage ($5x10^5$ cells (A) and $2.5x10^5$ cells (B) seeded). ES cell colony morphology appears good and seeding $5x10^5$ cells yielded a dense dish.

In summary, the adaptation step is dispensable but it is advisable to seed an entire vial of feeder-dependent ES cells intended for a 10cm dish onto a 6cm dish when using 2i medium directly after thawing. This helps to overcome the low cell number at the first passage, simply because the starting cell number is greater. Using an entire vial on a 6cm dish can lead to a relatively dense feeder layer due to the feeder cells left before freezing, which can be depleted by feeder depletion methods, or diluted out by subsequent passaging. The first passage appears to be critical without adaptation, but the cells that survive this first selection grow very well and subsequently appear morphologically normal.

I used these culturing approaches to adapt 2 different FVB/CAST ES cell lines (A2 and B2) and 2 different CAST/FVB (D2 and E2) cell lines to 2i media. After 4 passages I then harvested one 6cm dish from each of these cell lines and performed a standard RNA extraction using Trizol. The RNA from these cells will be subjected to RNA-seq and analyzed with the imprinted pipeline (developed by Daniel Andergassen and refined by Christoph Dotter, described below) to detect genes showing imprinted expression in undifferentiated ES cells. The ES cell culturing protocol that I optimized and the histone modification antibodies that I validated (see subsequent chapter) can then be used to also conduct ChIP-seq studies to detect differential enrichment of histone modifications at imprinted loci in these cells using the imprinted pipeline, as described for embryonic and extra-embryonic cells below.

3.2 Separation of visceral yolk sac endoderm from the mesoderm layer

In the present work VE was used as a representative for an EXEL tissue, which made it necessary to separate the two layers of the VYS. I did this separation together with Rita Casari following an efficient protocol developed in the laboratory (Hudson *et al.* 2011). To check the quality of this separation, RNA was isolated from the entire VYS and VE of E12.5 FVB +/+ and cDNA

was synthesized. *Flk1* is a marker gene for mesoderm and was used to screen for mesoderm contamination after separation of the two layers of the VYS (VE and VM). The qPCR assay yielded about 12.5% signal for *Flk1* in VE relative to the entire VYS (set to 100%) indicating a low level of mesoderm contamination. When testing for the endoderm marker *Afp*, separated VE shows an about 2.8x higher signal than in the entire VYS as is expected for an enriched endoderm sample (Fig. 10). The housekeeping gene cyclophilin was used to normalize all samples. These results were obtained early during the project and it can be assumed that the efficiency increased with more experience.

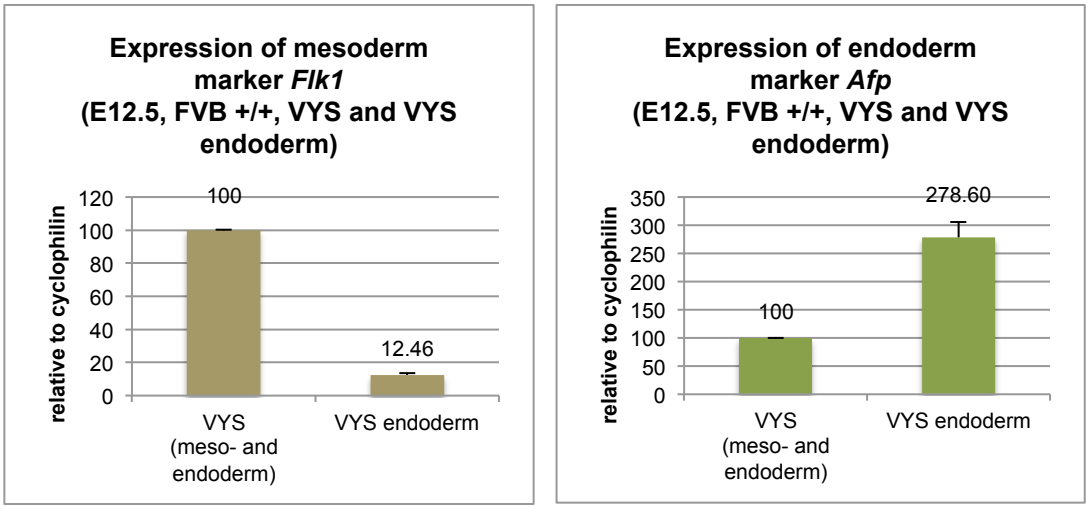

Fig. 10 Result of qPCR analysis to test for mesoderm contamination after separating the two layers of VYS. *Flk1* (mesoderm marker) showed about 12.5% of the signal of VYS in VE and *Afp* (endoderm marker) showed an about 2.8x higher signal in VE compared to VYS. Signals for VYS were set to 100 and all samples were normalized relative to cyclophilin.

3.3 Validation of antibodies against active marks H3K4me3 and H3K27ac

The main part of my project was to perform chromatin immunoprecipitation (ChIP) and therefore the next step was to validate antibodies and to optimize ChIP in MEFs, fetal liver and VE. Note that experiments were done following a native ChIP protocol adapted from Regha et al. (2007).

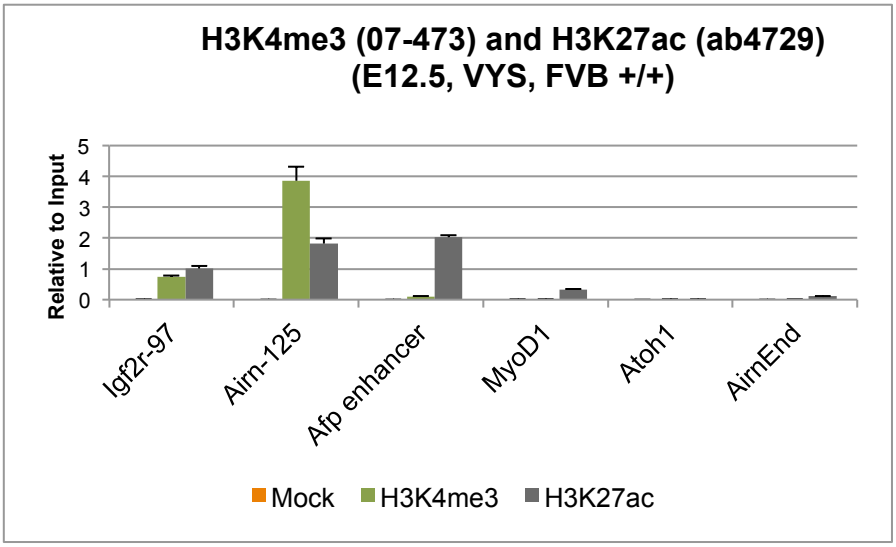

Fig. 11 Validation of H3K4me3 (07-473) and H3K27ac (ab4729) ChIP-antibodies on E12.5 VYS (FVB +/+) material by qPCR using assays for active promoters (*Igf2r* and *Airn*), active (*Afp* and *Myod1*) and inactive (*Atoh1*) enhancers and an intragenic region (AirnEnd). All controls show the expected pattern, indicating that the antibodies detect the correct epitope.

Fig. 11 shows chromatin immunoprecipitation (ChIP) qPCR validation of antibodies for the active histone modifications H3K4me3 (07-473 Millipore, active promoter mark) and H3K27ac (ab4729 Abcam, active promoters and enhancers) using E12.5 FVB VYS. The *Igf2r* promoter (Igf2r-97) and *Airn* promoter (Airn-125) known to be active in VYS were used as positive controls (see Fig. 12 for binding positions). Both showed good enrichment for H3K4me3 and H3K27ac, with the *Airn* promoter showing an about 5x stronger signal for H3K4me3 compared to the *Igf2r* promoter indicating that the *Airn* promoter is a very strong promoter (Fig. 11). The alpha-fetoprotein (*Afp)* is a marker gene for the VYS and therefore its enhancer is expected to be active. Indeed, it shows good enrichment for H3K27ac but not for the promoter mark H3K4me3 (Fig. 11). The *Afp* gene is highly transcribed in VYS and fetal liver and its enhancer is reported to be highly active (Tilghman 1985), which is confirmed by the ChIP qPCR data for VYS shown here. Also the second positive control for enhancers, the *MyoD1* enhancer shows enrichment only for H3K27ac, and not for H3K4me3, with the signal about 3x less than for the *Afp* enhancer.

As a negative control, the enhancer of the *Atoh1* gene, known for its role in neurogenesis, was used (Ben-Arie *et al.* 1996). As expected, ChIP qPCR

showed a very low signal not very different from the mock control, for both active histone marks. The primer pair AirnEnd binds within the gene body of the *Airn* gene, a region where high levels of H3K4me3 or H3K27ac would not be expected (Fig. 12). This was confirmed by ChIP qPCR that showed a very low H3K4me3 signal and a weak H3K27ac signal that was much lower than in the positive controls. The mock signal was very weak at all tested regions, demonstrating the background to be very low. In conclusion, ChIP qPCR assays showed that both the H3K4me3 (07-473) and H3K27ac (ab4729) antibodies are specific and give a good signal, which indicated that these antibodies can be used for ChIP-seq experiments.

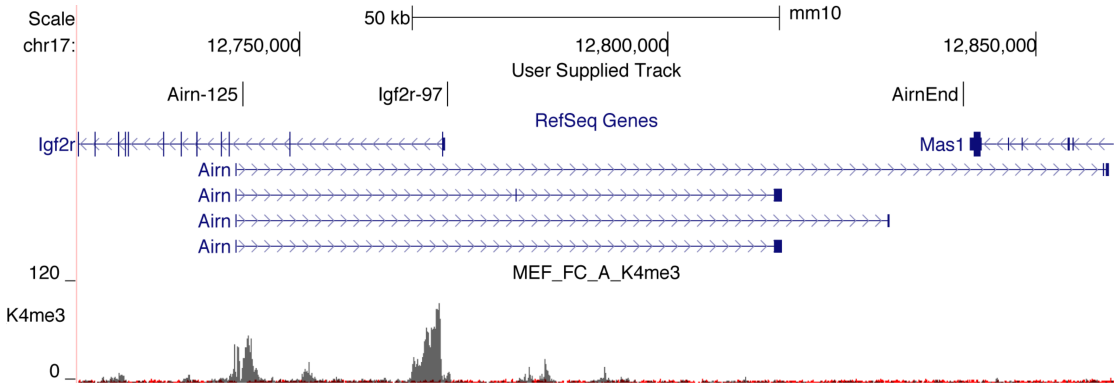

Fig. 12 Overview of the binding positions of qPCR primers used to validate H3K4me3 (07-473) and H3K27ac (ab4729) ChIP-antibodies in the *Igf2r*/*Airn* region. Airn-125 and Igf2r-97 bind at the promoters of *Airn* and *Igf2r* and AirnEnd is within the *Airn* gene body. The sequencing track shows H3K4me3 ChIP-seq from MEFs with peaks over the *Airn* and *Igf2r* promoters.

3.4 Using MEFs to validate ChIP-antibodies against repressive marks

The MEFB1 (+/T^{hp}) and MEFF (T^{hp}/+) immortalized cell lines (Regha *et al.* 2007) were used to test antibodies for ChIP before using them on embryonic tissue. MEFs have the advantage that they can be grown relatively easily and cost effectively, and cells can be quickly obtained in relatively large amounts, which are advantages for extensive testing of antibodies. Moreover, MEFB1 and MEFF cells are hemizygous for a 5 Mb deletion containing the *Igf2r* imprinted cluster and allow specific interrogation of the maternal and paternal

allele of the *Igf2r* cluster, which has known parental allele-specific histone modifications and thus can be used to validate antibodies (Regha *et al.* 2007). For example, the *Airn* promoter has H3K9me3 signal enriched on the silent maternal allele, while the *Igf2r* promoter shows H3K9me3 enrichment on the silent paternal allele. To test that the MEFB1 and MEFF cell lines were as expected, I first characterized the allele-specific DNA methylation pattern at the *Igf2r and Airn* promoters. Southern blots were performed cutting with EcoRI only (control) and EcoRI together with a methylation sensitive restriction enzyme, Not1 (GC↓GGCCGC) for the *Igf2r* promoter and Mlu1 (A↓CGCGT) for the *Airn* promoter. Methylation sensitive restriction enzymes are unable to cleave methylated-cytosine residues, resulting in the methylated allele showing the same fragment length as those samples digested only with EcoRI.

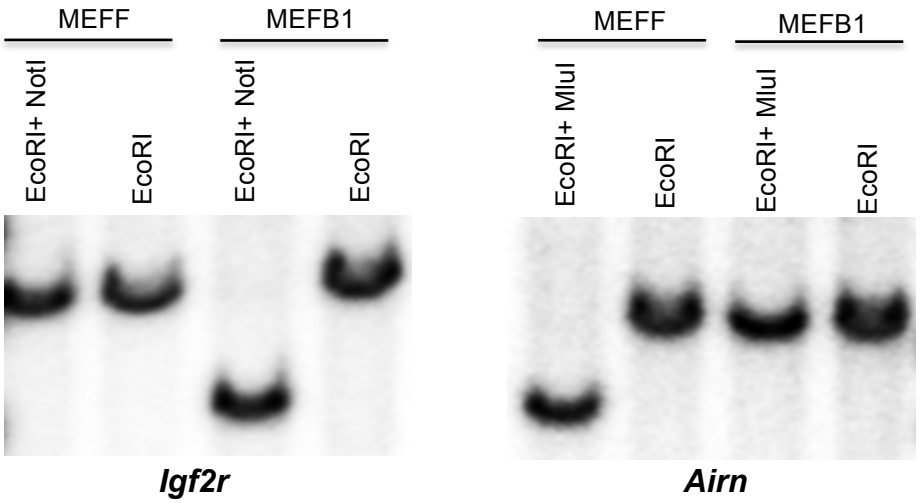

Fig. 13 Result of methylation assay by Southern blot shows the expected result of DNA methylation at promoters of inactive alleles. MEFF cells are DNA methylated at the *Igf2r* locus and MEFB1 at the *Airn* locus. The probes En-4kb (*Igf2r*) and Msi (*Airn*) were used.

The methylation assay showed the expected pattern of DNA methylation on the promoters of silent alleles. MEFF cells representing the paternal state are DNA methylated on the *Igf2r* promoter and not on the *Airn* promoter while MEFB1 cells representing the maternal state show the reverse pattern (Fig. 13). This is in agreement with previous reports (Pauler *et al.* 2005).

3.4.1 Histone 3 lysine 9 di-methylation

Two antibodies directed against H3K9me2 were tested using 30µg of chromatin obtained from MEFF and MEFB1 cells as a starting material for ChIP. The previously published positive controls Mage-a2, IAP (LTR) and MusD (LTR) were used to check the specificity of the antibodies (Dong *et al.* 2008). *Intracisternal A-particle* (*IAP*) and *MusD* are retrotransposons and *Mage-a2* is a class II endogenous retroviruses on the X-chromosome (Mager & Freeman 2000; Ribet *et al.* 2007) that were shown to be enriched for H3K9me2 (Dong *et al.* 2008). *Gapdh* as an expressed housekeeping gene should not show the repressive mark H3K9me2 and was therefore used as a negative control.

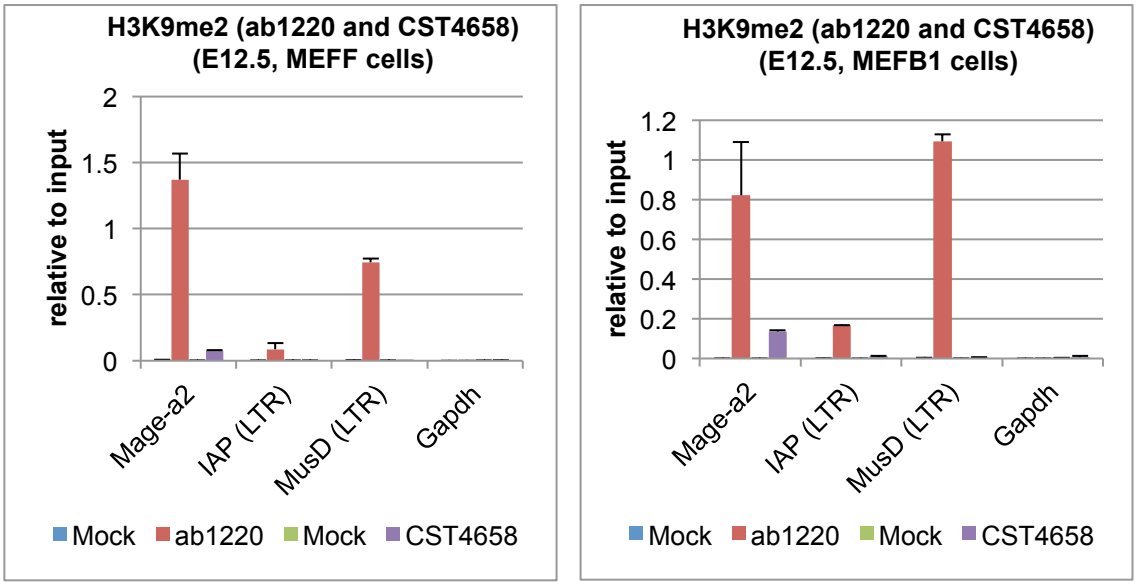

Fig. 14 Result of validation of two antibodies against H3K9me2 (ab1220 and CST4658) with the positive controls Mage-a2, IAP (LTR) and MusD (LTR) and the negative control Gapdh. ab1220 performs well showing the expected pattern, whereas CST4658 has a low signal and can therefore not be used in ChIP-seq experiments.

CST4658 (Cell Signaling Technology) showed some enrichment for Mage-a2, but IAP and MusD showed weak signals similar to the negative control Gapdh. In contrast, ab1220 (Abcam) showed good enrichment for all three positive controls and very low signal for Gapdh. Both antibodies showed a similar pattern in both cell lines (MEFF and MEFB1) and mock signals were very low at all tested positions (Fig. 14). In conclusion, ab1220 showed the

expected results for positive and negative controls and had a robust level of enrichment on positive controls indicating that it can be used for ChIP-seq. In contrast, CST4658 showed a low level of enrichment on positive controls indicating that it would not work for ChIP-seq. The next step was to validate ab1220 in embryonic tissue using E12.5 liver from FVB +/+.

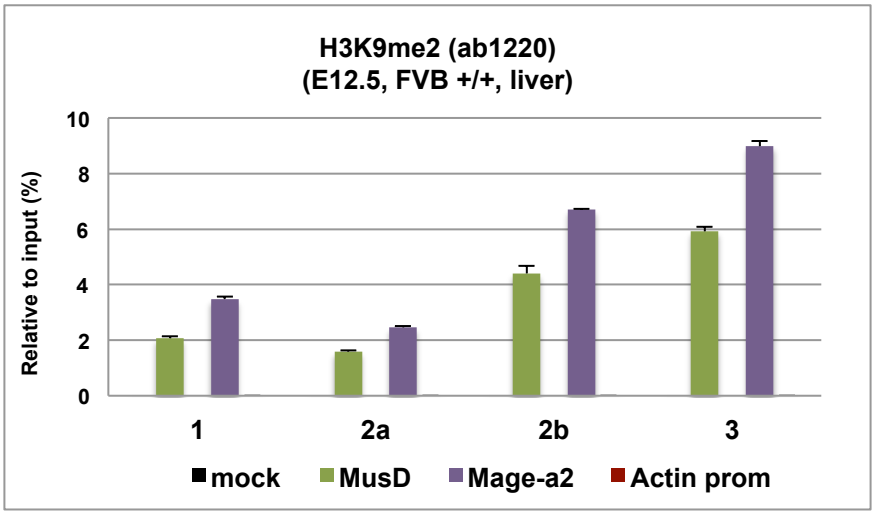

Fig. 15 Result of validation of ChIP on embryonic liver (FVB +/+) using the ab1220 antibody against H3K9me2 with 3 biological replicates (2a and 2b are technical replicates). Positive controls MusD and Mage-a2 show good enrichment and the negative control Actin prom shows a very low signal, indicating the antibody is specific.

Fig. 15 shows the qPCR results of a ChIP using ab1220 against H3K9me2 in E12.5 liver (FVB +/+) for 3 biological replicates. Samples 2a and 2b are the same biological replicate, but were split after dialysis into two technical replicates since the amount of chromatin was more than two times higher than in the other two biological replicates. Both positive controls (MusD and Mage-a2) showed good enrichment for H3K9me2 whereas the negative control (*Actin* promoter) was very low and comparable to the mock antibody. However, it was surprising that the two technical replicates 2a and 2b showed such different signal relative to input, with 2b more than 2 times higher for MusD and almost 3 times higher for Mage-a2 than 2a. In summary, for embryonic liver, in addition to MEFs, the ab1220 H3K9me2 antibody gives the expected results for positive and negative controls and a robust signal for positive controls indicating that the antibody can be used for ChIP-seq for both cell culture cells and tissues.

3.4.2 Histone 3 lysine 9 tri-methylation

The silent maternal *Airn* promoter is enriched with H3K9me3 (MEFB1 cells), while the active paternal promoter is not (MEFF cells). Conversely, the silent paternal *Igf2r* promoter is enriched for H3K9me3 (MEFF cells), while the active maternal promoter is not (MEFB1 cells) (Regha *et al.* 2007). Therefore, I used assays for the *Airn* and *Igf2r* promoters in these cells as positive and negative controls for testing H3K9me3 antibodies for ChIP. Three different antibodies were tested: 4861 (laboratory of Thomas Jenuwein), MAb-153-050 (Diagenode) and SN-146-100 (Diagenode).

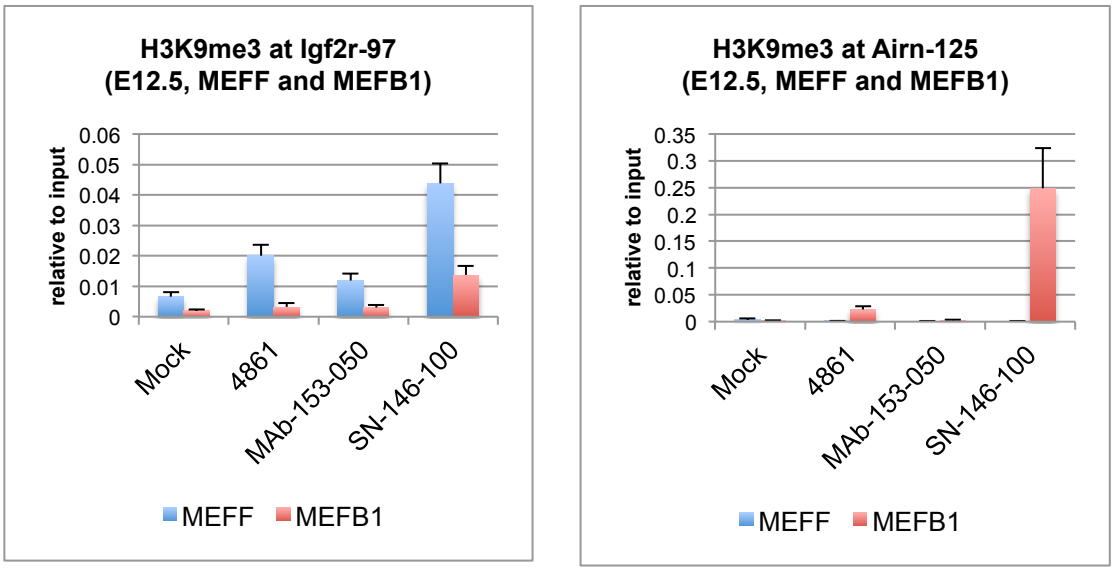

Fig. 16 Result of validation of three different ChIP-antibodies against H3K9me3 (4861, MAb-153-050 and SN-146-100) in MEFF and MEFB1 cells at the *Igf2r* and *Airn* promoters. The outcome shows that 4861 and SN-146-100 show the expected results with a robust level of enrichment, whereas MAb-153-050 shows very low enrichment indicating that it is not useful for ChIP-seq.

All three antibodies gave the expected pattern, with SN-146-100 showing the best results in terms of absolute enrichment on the silent alleles (positive controls), but also the highest signal for the active alleles (negative control) (Fig. 16). The optimal ChIP-antibody should show robust enrichment, but should also be highly specific, which is why the signal for both the positive and negative controls should be taken into account. When calculating the relative ratios between the silent and active *Igf2r* promoter in both cell lines, 4861 showed about 6x more enrichment on the paternal allele (MEFF) than

maternal allele (MEFB1), making it the most specific at this loci, as SN-146-100 showed 3x enrichment and MAb-153-050 4x more enrichment. At the *Airn* promoter, SN-146-100 showed the best enrichment on the silent maternal allele (MEFB1) compared to the active paternal allele (MEFF) with 283x enrichment, followed by 4861 (16x) and finally MAb-153-050 (2.5x) (Fig. 16). In conclusion, MAb-153-050 gave poor enrichment, but both SN-146-100 and 4861 with robust enrichment indicating that they should work for ChIP-seq. Since we already had the 4861 in relatively large amounts in the laboratory, different amounts of this antibody were tested to optimize the experiment and to get a higher signal relative to input.

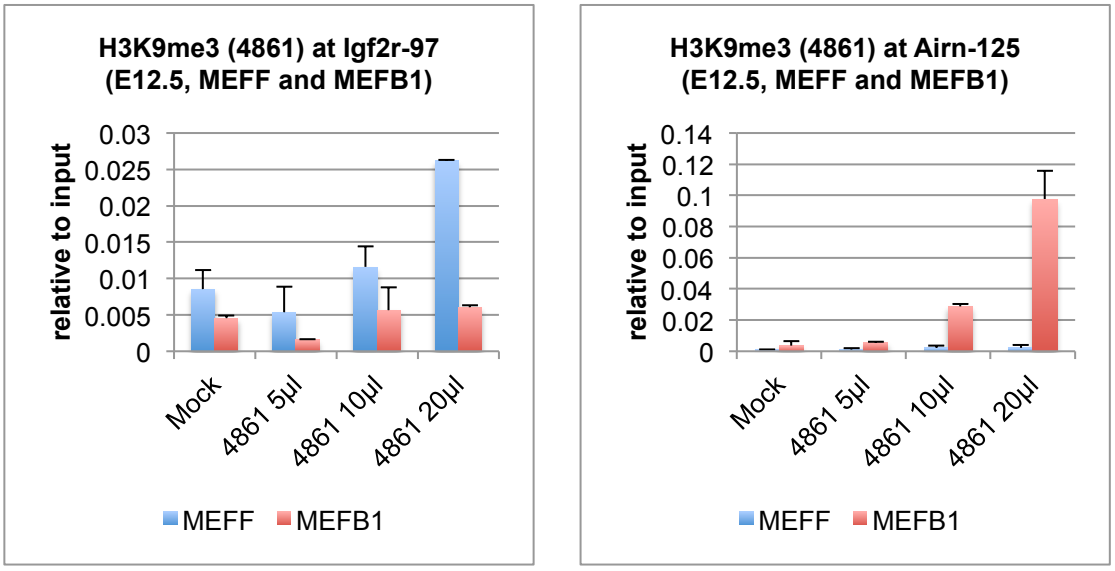

Fig. 17 Result of validation of the 4861 H3K9me3 antibody by qPCR using increasing amounts (5, 10 and 20µl) of antibody. 20µl shows the best results not only with the highest signal relative to input on the silent alleles (positive controls), but also the highest specificity (lowest silent to active allele ratios).

As expected, increasing amounts of 4861 pulled down more fragments, resulting in a higher signal compared to input. Again, the relative ratio between MEFF and MEFB1 cells at the *Igf2r* promoter and *Airn* promoter were calculated with the result that a higher amount of antibody not only pulled down more, but was also more specific. Especially at the *Airn* locus, signals were 5x (5µl), 14x (10µl) and 45x (20µl) enriched, showing that 20µl gave the best results. Surprisingly, the results for 5µl were not much different from mock levels, at the *Igf2r* promoter it was even lower, indicating this concentration of antibody in the ChIP was too low to detect low level

enrichment like that seen at the silent *Igf2r* promoter (Fig. 17). Next the antibody 4861 was tested in tissues using E12.5 liver of FVB +/+.

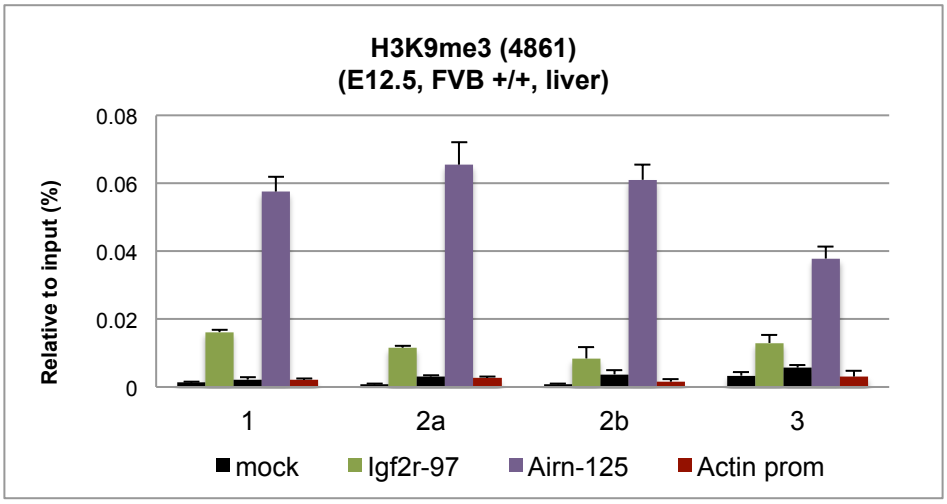

Fig. 18 Validation of the 4861 H3K9me3 antibody in E12.5 liver (FVB +/+) with 3 biological replicates (2a and 2b are technical replicates), resulting in the expected pattern of robust enrichment at the *Igf2r* promoter and the *Airn* promoter and low signals at the *Actin* promoter. Overall signals are lower in embryonic liver than in cells, although more chromatin was used.

Fig. 18 shows qPCR validation of ChIP using 20µl of 4861 targeted against H3K9me3 in E12.5 liver. The antibody yielded robust signals for the positive controls *Igf2r* promoter and *Airn* promoter and low signals for the negative control *Actin* promoter. The overall signal relative to input was unexpected low, even though 20µl of antibody was used and the ChIP starting material was between 46µg (1) and 86µg (2a and 2b) compared to 30µg used for MEF cells. The *Airn* promoter showed a relatively high signal compared to the *Igf2r* promoter similar to what was observed in MEFs and confirming previous reports (Regha *et al.* 2007; Latos *et al.* 2012).

3.4.3 Histone 3 lysine 27 tri-methylation

To test the antibodies against H3K27me3 with qPCR, published positive and negative controls were used (Pauler *et al.* 2009). H3K27me3 was shown to form broad local enrichments (BLOCs) rather than focal peaks and the primers used lie either outside (BV1 and BV5) or within a BLOC (BV14, BV21, BV25 and BV32) that covers part of the *Igf2r* imprinted cluster (Fig. 19). BV25 was reported to show a very weak signal, although it lies within the BLOCs

region, which is why it was used as a negative control in the validation assay (Pauler *et al.* 2009).

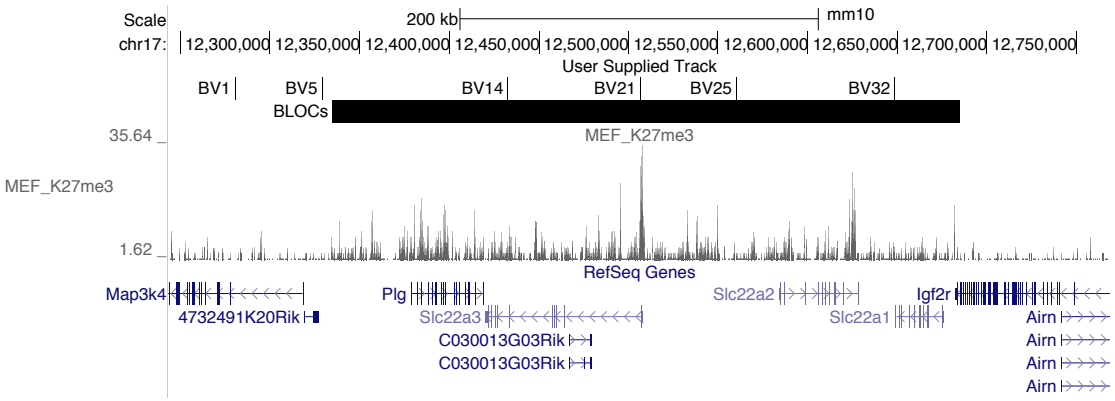

Fig. 19 Overview of the region used to validate ChIP-antibodies against H3K27me3. The BLOC region reported by Pauler et al. (2009) is indicated and also the primer positions used in this project (BV1, BV5, BV14, BV21, BV25 and BV32) (Pauler *et al.* 2009).

Two antibodies directed against the repressive histone modification H3K27me3 were tested (ab6002 (Abcam) and 6523 (laboratory of Thomas Jenuwein)) and both were found to perform well (Fig. 20).

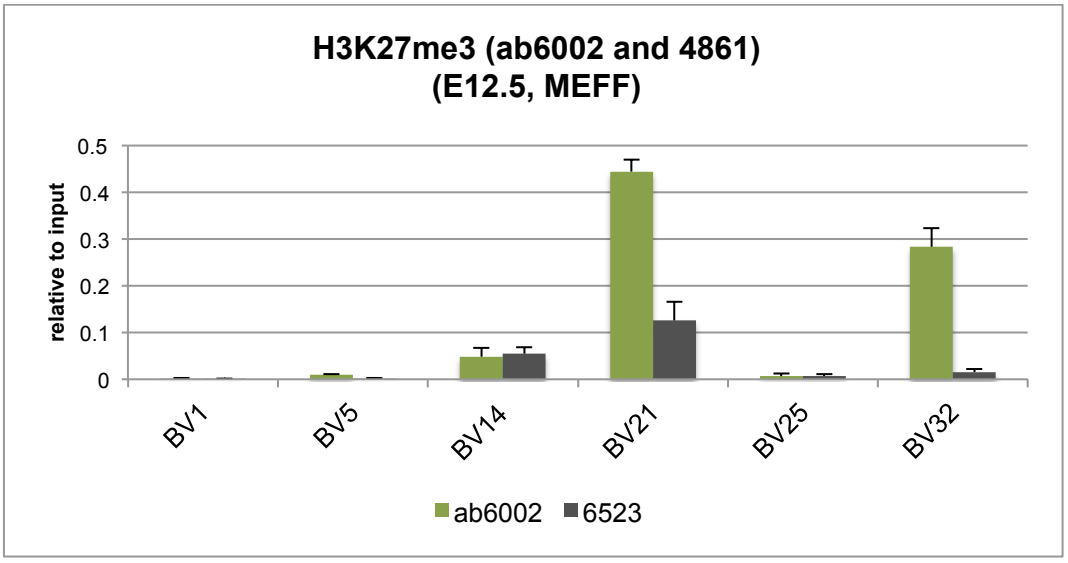

Fig. 20 Validation of the ab6002 and 6523 H3K27me3 antibodies with ChIP-qPCRs of positive (BV14, BV21, BV32) and negative controls (BV1, BV5, BV25). Both antibodies show the expected pattern.

In general, both antibodies against H3K27me3 show the expected pattern with low signals for the primers outside the BLOCs region (BV1 and BV5) and

for BV25. At BV32 ab6002 gave a surprisingly high signal, which is not expected from visual inspection of the sequencing track (Fig. 19) or the qPCR results from Pauler et al. (Pauler *et al.* 2009). The primer BV21 is at the promoter of the *Slc22a3* gene and yielded a strong signal from both antibodies (Fig. 20). This is also reflected by the sequencing track, showing a high peak at this position. The mock was very low for all tested primers (data not shown), demonstrating low background. Similar to the result for the 4861 (H3K9me3) antibody, the signal relative to input for 6523 was quite low when using only 5µl. (Fig. 20). Therefore, different amounts of antibody were tested to improve the signal, using 5, 10 or 20µl of antibody on MEFF cells (30µg starting material each). All concentrations of antibody showed the expected pattern, but the amount of pull-down relative to input was quite different, with 20µl giving the best signal relative to input (Fig. 21). Although, enrichment over negative controls also increased with increasing amounts of antibody, the overall ratio between negative and positive controls was better with 20µl antibody compared to 5 and 10µl. For all concentrations, the signal of the positive control BV14 was surprisingly high when compared to what was seen by Pauler et al. (2009), almost reaching levels of BV21, especially when 10 and 20µl of antibody was used. In conclusion, both antibodies appear suitable to be used in ChIP-seq, and while 5µl of ab6002 gave sufficient signal, 20µl of 6523 was necessary to obtain comparable results in terms of signal strength relative to input.

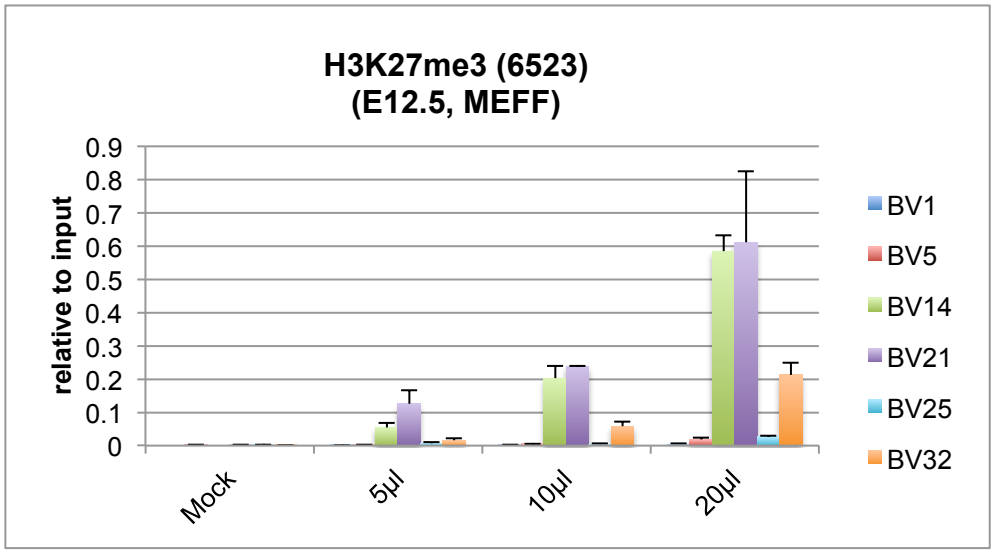

Fig. 21 Result of ChIP qPCR validation with increasing amounts of the 6523 antibody (5, 10 and 20µl) directed against H3K27me3. All concentrations showed the expected pattern with 20µl yielded the highest enrichment over positive controls, and the best positive to negative control ratio.

3.5 Using the imprinting pipeline to analyze ChIP-seq data

Following validation of histone modification antibodies for ChIP detailed in the previous chapter, I conducted ChIP-seq on the active histone modifications H3K4me3 and H3K27ac in tissues obtained from E12.5 embryos (MEFs, liver and VE) from reciprocal crosses from the FVB/N and CAST/EiJ mouse strains. Following ChIP, I prepared libraries for sequencing according to the TruSeq ChIP Sample Prep Kit Guide with minor changes indicated in the methods section. The samples were then sent for sequencing 50bp single end (SE) to the Biomedical Sequencing Facility (BSF) at CeMM. The amount of reads were in the range of 40- 50 million reads for each sample, except for H3K27ac in VE, for which we only obtained 22- 26 million reads.

I then used a bioinformatics pipeline developed in our laboratory by Daniel Andergassen and refined by Christoph Dotter to detect imprinted expression and differential enrichment of histone modifications, known as the imprinted pipeline (a methods paper by Daniel Andergassen et al., describing the pipeline is in preparation). Previous incarnations of this pipeline were used to detect imprinted expression from RNA-seq, which was done by Daniel Andergassen. At this point I would also like to acknowledge Daniel

Andergassen for setting up the pipeline and Christoph Dotter for providing great bioinformatics support and helping me to get familiar with the imprinting pipeline. In my analysis I focused on detecting parental allele-specific differential enrichment of the active histone modifications H3K4me3 and H3K27ac at promoters, to determine if this correlates with imprinted expression, and as a proof of principle that this experimental setup and the bioinformatics pipeline used can detect differential enrichment of histone modifications.

The following sections contain tables listing genes detected by the imprinted pipeline in MEFs, liver and VE from ChIP-seq data for H3K4me3 and H3K27ac. The tables show the location of the loci identified, the gene name, the imprinting score (IS), the ratio of maternal and paternal enrichment and the SNP coverage. One caveat of this approach is that if a gene or promoter region contains a large number of SNPs, a minor parental bias may have an imprinting score higher than the false discovery rate (FDR) even though it is debatable if this is a biologically meaningful bias. Therefore, I set an arbitrary cutoff requiring a minimum parental allelic ratio of 35:65 or 65:35 and excluded genes below this ratio from further analysis. Additionally, miRNA loci that lie within known imprinted genes were excluded. All identified genes were compared to both, the Otago Imprinted Gene Catalogue (Morison *et al.* 2001; Morison *et al.* 2005) and the MRC Harwell Imprinting Web page (Williamson *et al.* 2014) with the last column in each table stating the status of the identified gene. It can either be already known to show imprinted expression (*) or previously not reported as showing imprinted expression (unknown, u). The colors of the lines indicate the parental enrichment, with blue=paternal and red=maternal. The annotation used to analyze ChIP-seq was +/- 2kb of transcription start sides (TSS) of RefSeq-genes.

3.5.1 Setting the false discovery rate (FDR)

The imprinting pipeline adds up all the maternal and paternal sequencing reads over SNPs within a gene or other annotated region, such as a promoter, and calculates an imprinting score ($\log_{10}$ P-value) based on the

degree of bias in reciprocal comparisons (deviation from the binomial distribution) and SNP coverage (more highly covered genes/promoters show a higher IS for the same ratio) (DeVeale *et al.* 2012). The imprinting score was then given a positive value if the bias was maternal and a negative value if the bias was paternal. The imprinting pipeline used in the present work uses biological replicates, which allows a 4-way comparison between the reciprocal crosses and a mock comparison between the two biological replicates (Fig. 22).

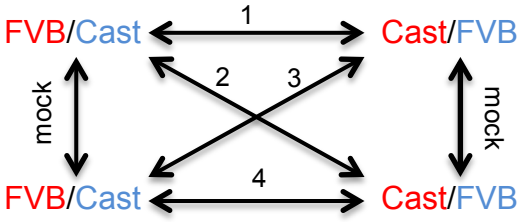

Fig. 22 All comparisons possible when using biological replicates of reciprocal crosses (4-way and mock). Mock comparisons between biological replicates allow the false discovery rate to be empirically set.

The mock comparison allows the FDR to be set empirically instead of being calculated based on mathematical models that cannot account for biological and technical biases (DeVeale *et al.* 2012). The FDR was set to 5% in all analyses performed, meaning that the IS necessary to call a gene is determined as a value where the number of genes detected in the mock comparisons (which should not differ much since they are biological replicates) is 5% of those identified in the F1 comparisons (the reciprocal crosses) (Fig. 23 FDR calculations for H3K4me3; Fig. 24 FDR calculations for H3K27ac). The imprinted pipeline calculates the IS strand-specifically for RNA-seq, but not for ChIP-seq, which is not strand-specific. In an advance over DeVeale et al. (2012) on which the pipeline is based, we make use of the 4 comparisons to calculate a mean score minus the standard deviation, and use this value to determine if a gene passes the FDR threshold thus reducing the number of false positives (DeVeale *et al.* 2012).

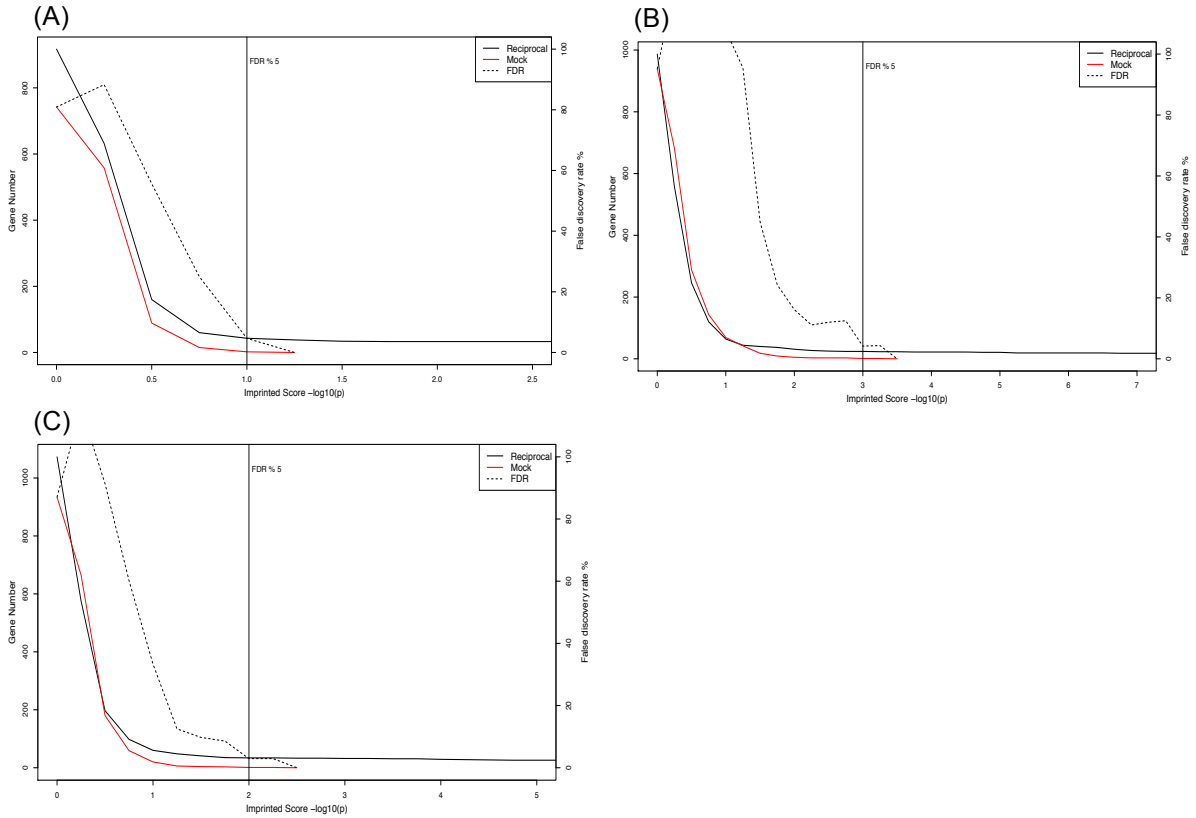

Fig. 23 Plot detail for H3K4me3 ChIP-seq data showing the imprinting score necessary at a FDR of 5% to call a promoter as differentially enriched. (A) MEFs (B) liver (C) VE. Scores are 1, 3 and 2 (MEFs, liver and VE, respectively).

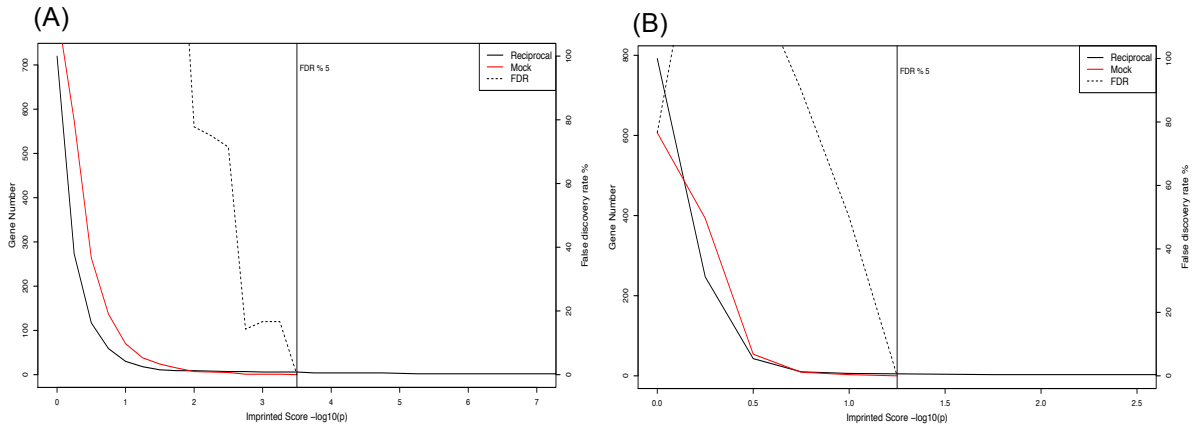

Fig. 24 Plot detail for H3K27ac ChIP-seq showing the imprinting score necessary at a FDR of 5% to call a promoter as differentially enriched. (A) liver (B) VE. Scores are 3.5 and 1.25 (liver and VE, respectively).

3.5.2 Validation of imprinting pipeline results for ChIP-seq

Screening for parental allele-specific differential enrichment of H3K4me3 identified 43 (MEFs), 24 (liver) and 34 (VE) loci. This result needed to be validated manually, first by removing genes that showed a parental bias ratio of less than 35:65 or 65:35. This analysis resulted that in MEFs 10 loci were below the cut-off (*A530072M11Rik*, *Slc7a5*, *Heatr2*, *U2af2*, *Acot8*, *Zswim3*, *Mir1949*, *Pwwp2a*, *Tcf19* and *Cchcr1*), all with a parental ratio not very different from 50:50 and a low IS (less than 1.58), but a high SNP coverage (at least 385). In liver 2 loci (*Pold3* and *Gtpbp4*) and in VE 3 loci (*Ddc*, *2410003L11Rik* and *L1td1*) below the cut off were identified. They showed a higher IS (3.09 and -3.59 for *Pold3* and *Gtpbp4* in liver, and -10.78, -2.28 and 2.75 for *Ddc*, *2410003L11Rik* and *L1td1* in VE) and while both loci in liver were highly covered (883 and 1480 reads, respectively), in VE *Ddc* and *L1td1* were highly covered (1005 and 924), but *2410003L11Rik* showed less reads (292). Except for *Ddc*, all loci excluded were not previously reported to show imprinted expression, and *Ddc* was only reported to show paternal imprinted expression in embryonic and neonatal heart (Menheniott *et al.* 2008).

The two micro RNAs *Mir675* and *Mir483* were identified in all cell types and showed a robust parental ratio. However, both were excluded because *Mir675* is processed from its host transcript *H19* (Cai & Cullen 2007) and *Mir483* is within intron 2 of *Igf2* (Landgraf *et al.* 2007). In addition, the miRNA *Mir335* was excluded because it maps within *Mest* (Royo *et al.* 2006). The ncRNA *R74862* identified in VE was excluded, because its promoter is on the antisense strand to *Cd81* and since ChIP-seq does not allow strand-specific analysis, it is likely that the differential H3K4me3 peak detected belongs rather to the known imprinted gene *Cd81* than to *R74862*. *Gab1* was reported to show expression from the paternal allele in placenta and biallelic expression in yolk sac at E13.5 (Okae *et al.* 2012). Surprisingly, in the present data, it was identified by enrichment of H3K4me3 on the maternal allele in VE, which is against published data, but since this is no reason to exclude it from subsequent data analysis it was retained. In MEFs, 3 loci not previously reported to show imprinted expression were detected (*Mir344d-2, Mir344c*

and *A330076H08Rik*) that do not lie within a gene known to show imprinted expression, which is why they were retained for further analysis. In summary, 30 (MEFs), 20 (liver) and 28 (VE) genes were identified to show parental specific enrichment of H3K4me3 at their promoter, all of which except for 3 (identified in MEFs) are already known to show imprinted gene expression.

3.5.2.1 Differential H3K4me3 in MEFs (n=30)

Screening for differential H3K4me3 in MEFs yielded 30 genes of which 27 were previously reported to show imprinted gene expression. The loci *Mir344d-2*, *Mir344c* and *A330076H08Rik* are not listed in both lists and are therefore considered to be potentially novel (Table 1).

Chr.	Gene name	Start	End	Score	Ratio mat.	Ratio pat.	Coverage	Status
chr2	Mcts2	152685147	152687942	-21,57	0,75	99,25	125	*
chr2	Gnas	174282319	174286319	7,90	100	0	34	*
chr2	Nespos	174293436	174297436	-76,41	0	100	312	*
chr6	Sgce	4745204	4749204	-27,52	1	99	137	*
chr6	Peg10	4745305	4749305	-22,01	0,5	99,5	116	*
chr6	Mest	30731505	30735505	-3,77	3,75	96,25	18	*
chr7	Zim1	6694432	6698432	40,98	91	9	322	*
chr7	Peg3	6728419	6732419	-93,72	0	100	486	*
chr7	Usp29	6728740	6732740	-93,01	0	100	480	*
chr7	Snurf	60003156	60007156	-152,30	0	100	658	*
chr7	Mir344d-2	61685271	61687351	-5,89	0,5	99,5	34	u
chr7	Mir344c	61837310	61839402	-6,49	1,5	98,5	39	u
chr7	A330076H08Rik	61980303	61984303	-31,98	3	97	176	u
chr7	Ndn	62346276	62349927	-4,91	3	97	43	*
chr7	Magel2	62374978	62378978	-7,88	0,5	99,5	51	*
chr7	Mkrn3	62418139	62422139	-28,46	2,5	97,5	157	*
chr7	Peg12	62462510	62466510	-18,00	5	95	106	*
chr7	H19	142576146	142580146	15,62	99,5	0,5	77	*
chr7	Igf2os	142657692	142661692	-7,47	1,5	98,5	42	*
chr7	Igf2	142659035	142663035	-16,16	1	99	84	*
chr7	Kcnq1ot1	143294547	143298547	-104,63	0,25	99,75	448	*
chr10	Plagl1	13088787	13092787	-78,66	0	100	317	*
chr11	Grb10	12035420	12039420	39,74	98,25	1,75	177	*
chr11	Zrsr1	22970004	22974004	-30,48	0,25	99,75	205	*
chr12	Meg3	109538995	109542995	52,80	99	1	292	*
chr15	Peg13	72808324	72812324	-181,96	0	100	914	*
chr15	Slc38a4	97053956	97057956	-60,10	0,75	99,25	322	*
chr17	Airn	12739310	12743310	-56,30	1,25	98,75	269	*
chr17	Igf2r	12767706	12771706	130,31	98,75	1,25	518	*
chr18	Impact	12970251	12974251	-78,75	0,25	99,75	412	*

Table. 1 Result of the imprinting pipeline, listing gene promoters identified to show differential H3K4me3 enrichment in MEFs. 30 loci were identified (27 known imprinted genes). The color indicates the parental enrichment (blue= paternal, red= maternal). The last column states whether the gene was previously shown to be imprinted expressed (*) or if its imprinting status is unknown (u). 3 loci were identified that are neither listed in Otago Imprinted Gene Catalogue nor in the MRC Harwell Imprinting Web page.

3.5.2.2 Differential H3K4me3 in liver (n=20)

Analysis of allele-specific H3K4me3 in liver identified 20 already known imprinted genes. All of them were also found in MEFs (Table 2).

Chr.	Gene name	Start	End	Score	Ratio mat.	Ratio pat.	Coverage	Status
chr2	Mcts2	152685147	152687942	-43,32	7,5	92,5	291	*
chr2	Nespos	174293436	174297436	-88,78	0	100	333	*
chr6	Sgce	4745204	4749204	-13,88	0	100	64	*
chr6	Peg10	4745305	4749305	-11,12	0	100	56	*
chr7	Zim1	6694432	6698432	28,89	95,75	4,25	177	*
chr7	Peg3	6728419	6732419	-11,88	24,5	75,5	314	*
chr7	Usp29	6728740	6732740	-11,18	24,75	75,25	306	*
chr7	Snurf	60003156	60007156	-79,23	0	100	301	*
chr7	H19	142576146	142580146	8,57	95,5	4,5	62	*
chr7	Igf2	142659035	142663035	-9,88	1	99	56	*
chr7	Kcnq1ot1	143294547	143298547	-111,76	0	100	406	*
chr10	Plagl1	13088787	13092787	-8,84	5,25	94,75	78	*
chr11	Grb10	12035420	12039420	5,11	92,5	7,5	37	*
chr11	Zrsr1	22970004	22974004	-35,25	14	86	316	*
chr12	Meg3	109538995	109542995	6,66	88,75	11,25	75	*
chr15	Peg13	72808324	72812324	-100,39	0,75	99,25	521	*
chr15	Slc38a4	97053956	97057956	-12,85	0	100	63	*
chr17	Airn	12739310	12743310	-20,88	0	100	102	*
chr17	Igf2r	12767706	12771706	17,94	94,75	5,25	142	*
chr18	Impact	12970251	12974251	-54,13	5	95	292	*

Table. 2 Gene promoter that were identified to show differential H3K4me3 enrichment in liver. 20 genes were identified, which are all already known to show imprinted gene expression. The color indicates the parental enrichment (blue= paternal, red= maternal). The last column states whether the gene was previously shown to be imprinted expressed (*) or if its imprinting status is unknown (u).

3.5.2.3 Differential H3K4me3 in visceral yolk sac endoderm (n=28)

Using the imprinting pipeline to analyze ChIP-seq data against H3K4me3 in VE revealed 28 genes that show parental specific enrichment at their promoter region (Table 3).

Chr.	Gene name	Start	End	Score	Ratio mat.	Ratio pat.	Coverage	Status
chr2	Sfmbt2	10368450	10372450	-43,48	2,25	97,75	234	*
chr2	Mcts2	152685147	152687942	-18,47	1,75	98,25	121	*
chr2	Nespos	174293436	174297436	-53,81	0	100	198	*
chr3	Jade1	41553733	41557733	-24,49	27,75	72,25	734	*
chr6	Sgce	4745204	4749204	-15,72	1,25	98,75	83	*
chr6	Peg10	4745305	4749305	-10,80	2	98	56	*
chr7	Zim1	6694432	6698432	4,59	85	15	167	*
chr7	Peg3	6728419	6732419	-55,24	3	97	412	*
chr7	Usp29	6728740	6732740	-54,26	3	97	401	*
chr7	Snurf	60003156	60007156	-6,59	2	98	46	*
chr7	H19	142576146	142580146	11,17	93,5	6,5	126	*
chr7	Igf2os	142657692	142661692	-12,58	2,5	97,5	70	*
chr7	Igf2	142659035	142663035	-52,40	1	99	266	*
chr7	Cd81	143050749	143054749	4,31	77	23	112	*
chr7	Kcnq1ot1	143294547	143298547	-76,36	0,25	99,75	332	*
chr7	Slc22a18	143471754	143475754	3,80	71,5	28,5	69	*
chr7	Phlda2	143501547	143504524	3,41	83,75	16,25	76	*
chr8	Gab1	80878479	80882479	9,59	65,25	34,75	397	*
chr10	Plagl1	13088787	13092787	-18,37	0,5	99,5	86	*
chr11	Grb10	12035420	12039420	3,94	88,25	11,75	38	*
chr11	Zrsr1	22970004	22974004	-16,32	5,75	94,25	172	*
chr12	Dlk1	109450822	109454822	-7,08	5,5	94,5	53	*
chr12	Meg3	109538995	109542995	26,57	90,5	9,5	305	*
chr15	Peg13	72808324	72812324	-22,89	1,5	98,5	153	*
chr17	Slc22a3	12505704	12509704	21,22	81,75	18,25	200	*
chr17	Airn	12739310	12743310	-26,37	2	98	164	*
chr17	Igf2r	12767706	12771706	34,74	94,25	5,75	214	*
chr18	Impact	12970251	12974251	-52,69	2	98	350	*

Table. 3 Gene promoters that were identified to show differential H3K4me3 enrichment in VE. 28 genes were identified all of which have been previously reported to show imprinted expression. The color indicates the parental enrichment (blue= paternal, red= maternal). The last column states whether the gene was previously reported to show imprinted expressed (*) or if its imprinting status is unknown (u).

3.5.2.4 More paternally expressed genes show differential H3K4me3

Fig. 25 shows the number of loci with differential H3K4me3 identified in MEFs (30), liver (20) and VE (28) grouped according to their parental enrichment. To date, more maternally expressed imprinted genes have been reported (Williamson *et al.* 2014) but surprisingly in all three cell types examined here, more paternally expressed genes were found (24 in MEFs, 15 in liver and 18 in VE). The most maternally expressed genes were identified in VE (10, compared to 6 in MEFs and 5 in liver).

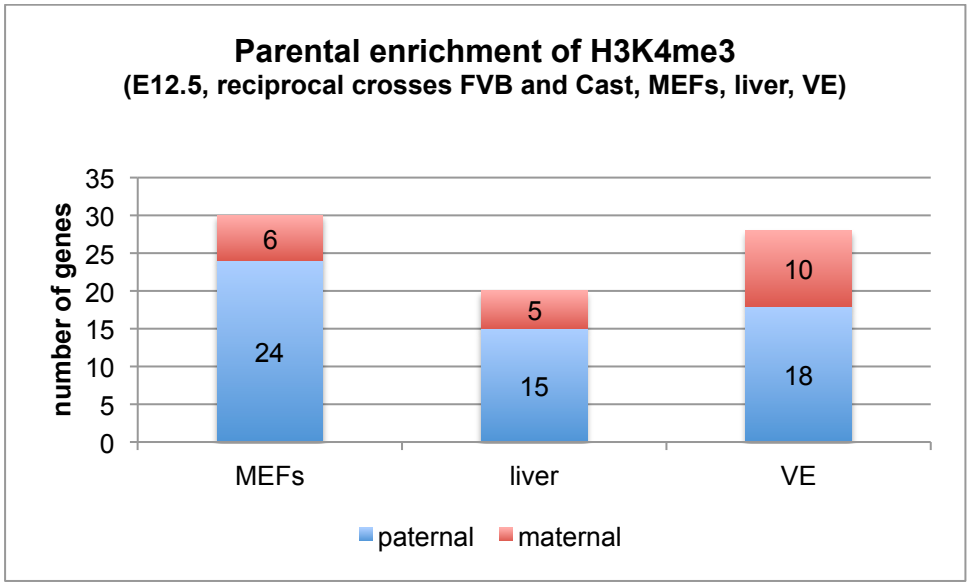

Fig. 25 Number of genes showing parental specific enrichment of H3K4me3 at promoter regions in MEFs, liver and VE. All three cell types show more paternally enriched genes.

3.5.2.5 Comparison of differential H3K4me3 in three cell types

Imprinted gene expression is tissue specific and Fig. 26 shows a comparison of genes identified in the three investigated cell types.

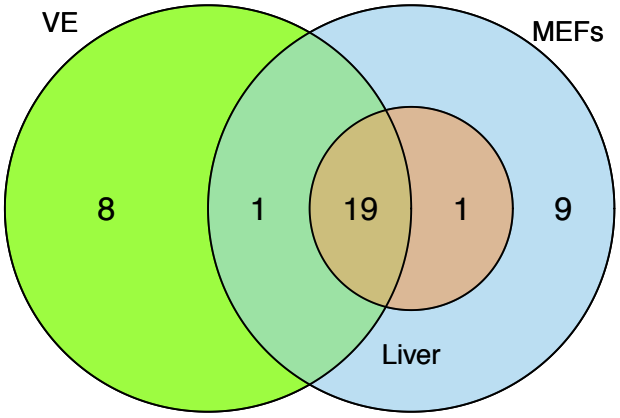

Fig. 26 Venn diagram comparing genes found in MEFs, liver and VE. 8 unique genes were identified in VE and 9 in MEFs. Genes in liver overlap 100% with MEFs.

Analysis of tissue specificity yielded 19 genes that were found in all three cell types, and also showed that all 20 genes identified in liver were also identified in MEFs. One gene identified in liver and MEFs was not found in VE (*Slc38a4*) and one gene found in VE and MEFs was not found in liver (*Igf2os*). VE shows 8 unique genes including 5 EXEL genes (*Gab1, Cd81*, *Dlk1*, *Jade1*,

Phlda2, *Sfmbt2*, *Slc22a18* and *Slc22a3*) and MEFs had 9 unique genes (*A330076H08Rik*, *Gnas*, *Magel2*, *Mest*, *Mir344c*, *Mir344d-2*, *Mkrn3*, *Ndn* and *Peg12*) (Fig. 26).

3.5.2.6 Results of differential H3K27ac in liver (n=6) and VE (n=5)

Following ChIP against H3K27ac in liver and VE the imprinting pipeline was used to screen for differential enrichment over promoter regions of known RefSeq genes (Tables. 4 and 5). Analysis of ChIP against H3K27ac in liver and VE resulted in the identification of 6 and 5 genes, respectively, that showed differential enrichment of this histone modification at their promoter region (+/- 2kb of TSS of RefSeq genes) (Tables. 4 and 5). The amount of genes detected compared to the number of genes identified in the same tissues by differential H3K4me3 is low, suggesting H3K27ac may be less useful to identify imprinted gene expression. *Fam192a* on mouse chromosome 8 found in liver was the only novel gene identified, and was not detected by H3K4me3 ChIP-seq or reported from RNA-seq before. It has a ratio just above the cut off and a weak IS and would need to be validated by other techniques to determine if it is a true imprinted gene. All other genes identified in liver showed very robust parental ratios and were previously reported to show imprinted expression (Table. 4).

Chr.	Gene name	Start	End	Score	Ratio mat.	Ratio pat.	Coverage	Status
chr2	Mcts2	152685147	152687942	-3,67	0	100	23	*
chr2	Nespos	174293436	174297436	-4,92	4,5	95,5	35	*
chr8	Fam192a	94599726	94603726	3,51	66,75	33,25	137	u
chr11	Zrsr1	22970004	22974004	-11,03	1,25	98,75	56	*
chr15	Peg13	72808324	72812324	-5,02	0,75	99,25	58	*
chr18	Impact	12970251	12974251	-11,66	1,75	98,25	78	*

Table. 4 Genes identified to show differential H3K27ac enrichment in liver. 6 genes were identified, of which 5 were already known to show imprinted gene expression. The color indicates the parental enrichment (blue= paternal, red= maternal). The last column states whether the gene was previously shown to be imprinted expressed (*) or if its imprinting status is unknown (u).

In VE there were more variable parental ratios, but all fell within the specified range (Table. 5). All genes that showed differential H3K27ac enrichment were also identified to show differential H3K4me3 enrichment, except for *Fam192a*

in liver (Tables. 2 and 3). Note that the reads obtained from the sequencing was only about half for VE compared to liver and since the imprinting pipeline is based on coverage and ratio, it is likely that re-analyzing of better quality sequencing could identify more genes.

Chr.	Gene name	Start	End	Score	Ratio mat.	Ratio pat.	Coverage	Status
chr2	Mcts2	152685147	152687942	-1,738967365	16,25	83,75	13	*
chr7	Peg3	6728419	6732419	-5,915328357	6,25	93,75	44	*
chr7	Slc22a18	143471754	143475754	1,480529587	74,25	25,75	23	*
chr7	Usp29	6728740	6732740	-7,102123696	3,5	96,5	42	*
chr11	Zrsr1	22970004	22974004	-3,46480749	10,25	89,75	29	*

Table. 5 Genes identified to show differential H3K27ac enrichment in VE. 5 genes were identified that are all already known to show imprinted gene expression. The color indicates the parental enrichment (blue= paternal, red= maternal). The last column states whether the gene was previously shown to be imprinted expressed (*) or if its imprinting status is unknown (u).

4. DISCUSSION

The idea of this project was to determine the imprinting signature of MEFs, liver and VE by performing ChIP-seq against H3K4me3, H3K27ac, H3K9me2, H3K9me3 and H3K27me3. However, due to a number of reasons e.g. troubles with breeding the mice and finding efficient antibodies, I could only do part of this. I managed to find efficient antibodies against all histone modifications mentioned above and performed ChIP-seq for the active marks. This was then analyzed to screen for parental- and tissue-specific enrichment using an imprinting pipeline developed by Daniel Andergassen and refined by Christoph Dotter in the Barlow laboratory. With this approach I could identify 30 (MEFs), 20 (liver) and 28 (VE) genes that showed differential enrichment of H3K4me3, and 6 (liver) and 5 (VE) genes with parental specific enrichment of H3K27ac. Most of these loci are already known, but I could also identify 3 for H3K4me3 in MEFs and 1 for H3K27ac in liver that are not listed in the Otago Imprinted Gene Catalogue (Morison *et al.* 2001; Morison *et al.* 2005) and the MRC Harwell Imprinting Web page (Williamson *et al.* 2014).

4.1 Importance of validation of antibodies

The first goal of this project was to validate ChIP-antibodies for specificity and efficiency. On the one hand, an ideal antibody should enrich the target DNA and should pull down a reasonable amount of DNA fragments, which can be measured relative to input. On the other hand, it should be highly specific leading to very low background signal. This requires extensive candidate testing done in the present work by performing ChIP followed by qPCR using primers of known positive and negative controls. In a large-scale study to determine the quality of antibodies, Egelhofer et al. (2010) tested more than 200 commercially available antibodies against a number of histone modifications (Egelhofer *et al.* 2011). They found about 25% not to perform well and also reported that some only work in certain assays but not in others. There exist a number of different methods to assess the quality of antibodies including western blot, dot blot, ChIP-chip and ChIP-seq (Egelhofer et al.,

2010), but also mass spectrometry (Peach *et al.* 2012). Nevertheless, an antibody shown to work for one application still needs to be validated for other applications with conditions that match the actual experiment as much as possible. In the present work, MEFs were used to initially test antibodies for ChIP and optimize conditions since it is not feasible to obtain embryonic tissue for extensive testing. Where possible, these conditions were tested at least once on an embryonic tissue to confirm that experimental conditions determined using cell culture cells are transferable.

In general, there exist two different kinds of antibodies on the market, namely polyclonal and monoclonal, which both have their advantages and disadvantages. The most striking one is that lot-to-lot variation is common for polyclonal antibodies, meaning each new batch needs to be validated before using in experiments. Monoclonal antibodies should not vary between batches, but often produce a lower level of enrichment than polyclonal antibodies. This may be because monoclonal antibodies only recognize one epitope on an antigen, whereas polyclonal antibodies may recognize multiple epitopes on the antigen. The antibodies ab1220, SN-146-100 and ab6002 validated in the present work are monoclonal (against H3K9me2, H3K9me3 and H3K27me3, respectively) but showed a relatively robust enrichment at positive controls. In contrast, MAb-153-050 (against H3K9me3) is also monoclonal and failed in the present validation. It was specific in that it showed the expected enrichment pattern, but pull-down efficiency was very low.

4.1.1 Finding suitable antibodies for ChIP-seq

The main conclusion from testing a number of antibodies for ChIP is that finding efficient and specific antibodies was much easier for the active marks H3K4me3 and H3K27ac than it was for the repressive marks H3K9me2, H3K9me3 and H3K27me3. In fact, for the active marks the ones already present and in use in the laboratory could be used, whereas finding good working ones for the repressive marks turned out to considerably harder and needed a lot more testing. For the non-commercial polyclonal antibodies 4861

(H3K9me3) and 6523 (H3K27me3), the performance increased with more antibody used, with increased enrichment and an improved ratio between positive and negative controls. The commercial antibodies SN-146-100 (H3K9me3) and ab6002 (H3K27me3) did not require this step of optimization and gave an acceptable signal/noise ratio with 5µl antibody.

Another important point in ChIP is to determine the starting amount of chromatin needed for a successful experiment and, especially since in the present work mostly embryonic tissue was used, limiting it to a minimum amount is also important. For all antibodies confirmed to be suitable for ChIP-seq an input amount of 30µg was sufficient, with more not giving better results (data not shown), despite the original protocol demanding 100 µg of chromatin. This could be because previously up to six different antibodies were used per ChIP, and many PCR assays were conducted that required more product than ChIP-seq (Regha *et al.* 2007). It could be speculated that even less than 30µg could be used for ChIP, especially for the active marks that show very robust signals in qPCR.

In summary, in the present work antibodies for the active histone modifications H3K4me3 and H3K27ac were validated by qPCR and were later successfully used for ChIP-seq. Identification of ChIP-grade antibodies for the repressive modifications H3K9me2, H3K9me3 and H3K27me3 required more testing, but antibodies for each of these marks have been validated for ChIP and are planned to be used for ChIP-seq in the near future.

4.2 Identification of known and potentially novel imprinted genes

4.2.1 Imprinted expression is tissue specific

To date, the Harwell MouseBook Imprinting Catalog lists 150 genes that were reported to show imprinted expression (Williamson et al., 2014). Of these 150, 89 are expressed from the maternal and 61 from the paternal allele. However, from the present ChIP-seq data, only 30 (MEFs, of which 3 are novel and not

on the Harwell list), 20 (liver) and 28 (VE) genes could be identified to show differential H3K4me3 on their promoter.

There are biological and technical reasons why not all imprinted genes are detected as showing imprinted expression in a given tissue. First, imprinted gene expression is very tissue specific and as mentioned earlier, a subset of genes shows EXEL-specific imprinted expression or only shows imprinted expression in a tissue that was not examined in this study, such as the brain (Mancini-DiNardo *et al.* 2003; Peters & Williamson 2007; Wang *et al.* 2008; Hudson *et al.* 2010; Hudson *et al.* 2011; Santoro & Barlow 2011; Wang *et al.* 2011). Furthermore, some genes were identified to exhibit a distinct developmental stage-specific imprinted expression pattern, and therefore do not necessarily show imprinted expressed at the investigated time point (E12.5). Examples are the EXEL-specific imprinted genes *Ins2* in the *Igf2* cluster and *Slc22a3* in the *Igf2r* cluster. *Ins2* gains imprinted expression after E14.5 and Slc22a3 looses imprinted expression after E15.5 (Giddings *et al.* 1994; Deltour *et al.* 1995; Zwart *et al.* 2001; Nagano *et al.* 2008). Second, imprinted expression cannot be detected for some genes due to a strain bias in expression between FVB and CAST that is greater than the expression difference due to imprinted expression. Third, with the approach used in the present work relying on SNPs, differences can only be detected when there is at least one SNP within the investigated region (+/- 2kb of TSS of RefSeq genes for ChIP-seq). Although FVB and Cast can be distinguished by about 20 million SNPs (Keane *et al.* 2011), it is impossible to assign sequencing reads according their parental origin if a gene promoter does not contain a SNP, and therefore the imprinted expression status of such genes cannot be determined. For each SNP position, a score is calculated and all SNPs within a region are summarized, meaning the more SNPs, the more robust imprinted gene expression can be detected. Moreover, the SNP score is a combination of coverage and ratio so it is more difficult to detect imprinted expression from genes with low SNP coverage. On the other hand, genes that are very well covered by SNPs and only show a mild bias in expression rates may be detected, although the biological relevance of such minor biases in expression is debatable (DeVeale *et al.* 2012).

Analysis of RNA-seq data done by Daniel Andergassen using the same cell types and bioinformatics pipeline (but an older version) as in the present work, identified more genes to be differentially expressed than what I detected from H3K4me3 and H3K27ac enrichment (Daniel Andergassen, unpublished data). I confirmed this result by re-analyzing the RNA-seq data with the current version of the pipeline. A recent study conducted in MEFs found 32 genes that were all described before to be imprinted expressed (Tran *et al.* 2014) and in the present work, 30 loci were identified to show allele-specific enrichment for H3K4me3, of which 27 are known imprinted genes. 20 out of 32 genes identified by Tran et al. (2014) were also detected in the present work to show differential H3K4me3 enrichment at their promoter, but 12 could not be identified in the present work. However, 10 genes found with parental specific H3K4me3 enrichment, of which 7 were previously reported to show imprinted gene expression, were not identified by Tran et al. (2014). Notably, they used different F1 mouse crosses that were different by about 15 million SNPs compared to the 20 million SNPs of the crosses used in the present work (Keane *et al.* 2011; Tran *et al.* 2014).

The difference in the number of identified loci between RNA and ChIP-seq using the same pipeline may be caused by technical reasons. For analysis of RNA-seq data the annotated gene body was used. This is generally a larger region than the 4kb window over promoters used for ChIP-seq in this study. As a consequence, the RNA-seq approach may be more powerful, because more SNPs are within the investigated regions and since the SNP scores are added up, the differences can be more easily detected. However, it is not known if all imprinted genes would show differential H3K4me3 at their promoters. If this is not the case then this would be a biological explanation that accounts for the lower number of imprinted genes identified using ChIP-seq.

4.2.2 The importance of a good annotation

A key element of the imprinting pipeline is the annotation used. In the present work a window of +/- 2kb of TSS of RefSeq genes was used, which works well

when looking for differential promoter marks of well-characterized genes. Alternatively, one could call peaks with the help of MACS (Zhang *et al.* 2008) and use the result as an annotation, since H3K4me3 forms focal peaks. Like a custom annotation of the transcriptome for RNA-seq, peak calling would enable a more complete picture not just including RefSeq genes. When screening for enhancer or repressive histone marks peak calling may not work, since for example H3K27me3 forms broader regions rather than focal peaks (Pauler *et al.* 2009) and also H3K27ac was observed from the sequencing tracks (this work) to form larger regions rather than clearly separated peaks, which could lead to non-satisfying results. Although recently the existence of super-enhancers was reported proposing median sizes of about 9kbp (Hnisz *et al.* 2013; Whyte *et al.* 2013), enhancers are typically not as large as the H3K27ac peak is, which is why an alternative approach is needed. A possibility would be a sliding-window approach, meaning to take, for example, a 1kb window and move this in 100bp steps along the genome, screening for significant differences between the parental alleles.

4.2.3 Strand-specific analysis is necessary to resolve some issues

RNA-seq allows strand-specific analysis, but for ChIP-seq data this is not possible. With the present annotation this leads to problems with promoter regions that are in close proximity, even though they are on opposite strands. This problem was observed with the ncRNA *R74862* identified in VE. Although *R74862* and *Cd81* lie on opposite strands, the H3K4me3 peak is detected for both of them and as a result both are reported to show differential enrichment. These cases needed to be resolved manually by visual inspection of the sequencing tracks. In the present case it is very likely that the differential peak is at the promoter of *Cd81*, since it was reported earlier to show imprinted gene expression in early development in EXEL but not in the embryo (Fig. 27) (Caspary *et al.* 1998; Dao *et al.* 1998; Lewis *et al.* 2004; Umlauf *et al.* 2004). The same problem was observed for *Sgce* and *Peg10*, *Peg3* and *Usp29* and *Igf2os* and *Igf2* which are all on opposing strands, but are in close proximity and differential H3K4me3 cannot be assigned to one

gene, but since all these genes are known to show imprinted expression, they were kept for subsequent analysis.

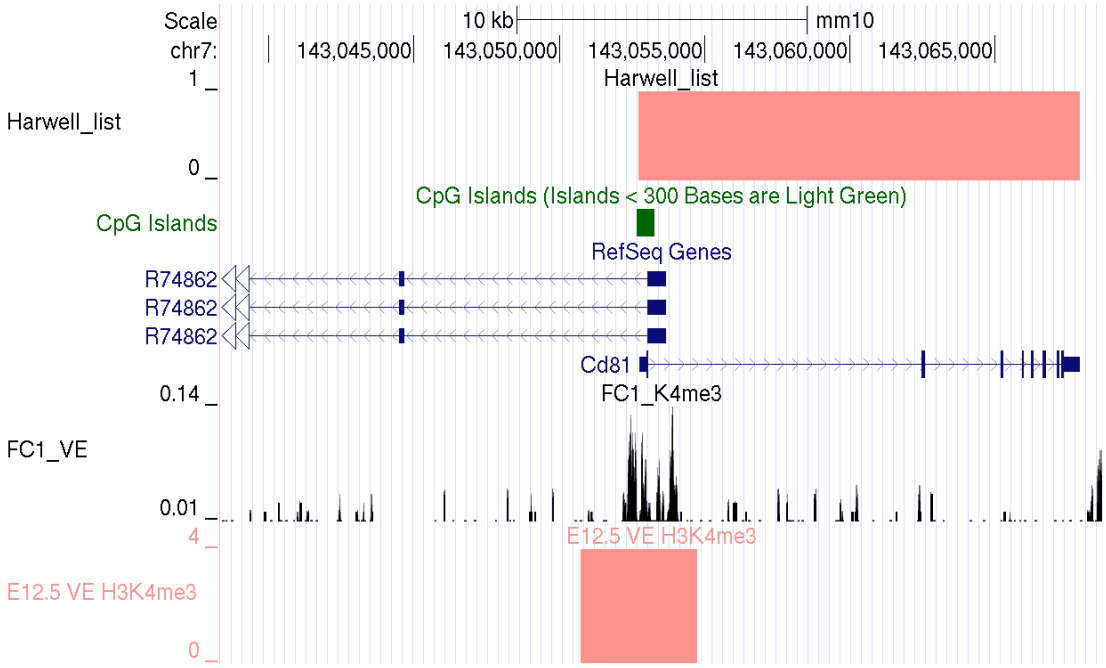

Fig. 27 Screenshot of the promoter region of *Cd81* and *R74862* highlighting the problem of promoters that are in close proximity to each other. Although they are on opposite strands, both are detected because the 4kb window included both promoters and for ChIP-seq no strand-specific analysis is possible.

4.2.4 A subset of genes shows extra-embryonic lineage specific imprinted expression

So far, the global imprinted expression pattern in VE has not been defined, since placenta was used as a representative of an EXEL tissue in RNA-seq studies to screen genome-wide for imprinted gene expression (Wang *et al.* 2011; Okae *et al.* 2012). However, analysis of imprinted expression in placenta is complicated as it contains multiple cell types including maternal contaminating tissues whereas VE as a homogenous population of EXEL cells overcomes these problems and was therefore used in this project (Hudson *et al.* 2011).

In the present work, I used VE as a model for an EXEL imprinted state and identified via ChIP-seq against H3K4me3 eight genes only detected in VE to have differential H3K4me3 enrichment on their promoter (five are known to

only show imprinted expressed in EXEL tissues), of which three show paternal and five maternal enrichment. 2 of these genes are within the known imprinted clusters *Kcnq1* (*Cd81*) and *Igf2r* (*Slc22a3*) and were previously reported to show imprinted expression in placenta (Zwart *et al.* 2001; Lewis *et al.* 2004). Moreover, *Phlda2* and *Slc22a18*, both part of the *Kcnq1* cluster and reported to show ubiquitous imprinted expression (Dao *et al.* 1998; Dunwoodie & Beddington 2002), were only found in VE to be differentially enriched for H3K4me3 (*Phlda2* and *Slc22a18*) and H3K27ac (Slc22a18). *Sfmbt2* was reported to be expressed from the paternal allele only in placenta and VYS (Kuzmin *et al.* 2008) and was in the present work also only detected to show differential enrichment of H3K4me3 in VE. *Jade1* doesn't map to any known imprinted gene cluster and is therefore considered to be a solo imprinted gene on mouse chromosome 3 and was reported to be paternally expressed in E17.5 placenta (Wang *et al.* 2011). In the present work it showed differential enrichment of H3K4me3 on the paternal allele indicative of paternal specific imprinted expression also in E12.5 VE. The paternally expressed gene *Dlk1* is part of the *Dlk1-Dio3* cluster (Kobayashi *et al.* 2000; Hagan *et al.* 2009). *Gab 1*, previously reported to show expression from the paternal allele in placenta was detected with enrichment of H3K4me3 on the maternal allele in VE (Okae *et al.* 2012). This is surprisingly and the reproducibility of this finding would need further proof, for example via Sanger sequencing. Exclusive expression from the paternal allele was also shown in primary trophoblast stem cell culture, but in embryo, yolk sac and adult tissue it was reported to be biallelically expressed (Okae *et al.* 2012).

Screening for differential H3K27ac resulted in general a smaller number of genes in the investigated tissues liver and VE with 6 and 5, respectively. Of these, *Fam192a* (in liver) was previously not reported to show imprinted gene expression but it was also not identified in the present study to have differential H3K4me3 and requires further validation. *Slc22a18* and *Usp29* were only found in VE, but not in liver. For both, imprinted gene expression was reported in placenta, with *Usp29* being paternally and *Slc22a18* maternally expressed (Wang *et al.* 2011; Okae *et al.* 2012). This is in agreement with the parental specific H3K27ac enrichment detected. However,

Usp29 was identified in all three cell types to have H3K4me3 enrichment on the paternal allele, showing that it is not EXEL specific. *Slc22a18* was only found in VE to show maternal enrichment for both histone modifications, but it was reported to show ubiquitous imprinted expression (Dao *et al.* 1998). Xie et al. (2012) reported *Usp29* to be paternally enriched for H3K4me3 and H3K27ac, which is consistent with the findings in the present work, although they used mouse frontal cortex (Xie *et al.* 2012).

It should be noted here as mentioned earlier that the sequencing result for H3K27ac in VE contained a lot of low quality reads, which is why after a quality filter was applied only between 22-26 million reads were left (compared to about 45 million for liver). Since coverage is crucial parameter for detecting a gene as showing imprinted expression by the pipeline, the present results might underestimate the number of genes showing differential H3K27ac.

4.2.5 Potentially novel imprinted genes

The list of genes showing imprinted expression is far from being complete since studies investigating a wide range of different tissues and developmental stages are lacking. However, novel imprinted genes are frequently reported, one study even suggested the existence of more than 1300 loci to show imprinted gene expression in brain (Gregg *et al.* 2010), but re-evaluation indicated that most were false positives (DeVeale *et al.* 2012). Therefore, particular attention needs to be paid to proper analysis to detected imprinted genes. The pipeline used in the present work uses biological replicates to allow the FDR to be empirically set, which minimizes the danger of detecting false positives. Nevertheless, genes that were identified by the imprinting pipeline and are not already known to show parental specific gene expression, were manually validated by first checking allele-specific differences in H3K4me3 and H3K27ac enrichment, and genes with a bias of less than 35:65 or 65:35 expression were excluded. Although there is no agreement in the imprinting field about when to call a gene as showing imprinted gene expression (Barlow 2011), it is unlikely that a ratio of 45:55

has any biological meaning and is rather caused by technical limitations of the approach. Therefore, it is advisable to exclude these findings or even to implement a function in future versions of the imprinting pipeline that allows a ratio cut off to be automatically set. Three genes annotated in RefSeq were identified to show differential H3K4me3 enrichment in MEFs that have not been reported to show imprinted expression (*Mir344d-2, Mir344c and A330076H08Rik*). All three map in close proximity to one another with *Mir344c* lying within and *Mir344d-2* close to 3' end of the lncRNA *A230057D06Rik* in the *Snurf/Snrp* cluster, making it likely that both map to one transcript from which the miRNAs are processed. Both miRNAs are part of the *Mir344* cluster, which is only present in rodents and not in humans and to date is it not clear whether they are associated with imprinted gene expression (Royo & Cavaille 2008). Using mouse frontal cortex, Xie et al. (2012) identified a strong paternal H3K4me3 peak at *Mir344g* and a weaker paternal enrichment at *Mir344b*, *Mir344-2* and *Mir344f* (Xie *et al.* 2012). Visual inspection of total RNA-seq data assembled by Cufflinks (Trapnell *et al.* 2010) showed that there is continuous expression throughout the entire region, supporting the idea that these 3 genes are part of the same transcript (Fig. 28) (Philipp Günzl, unpublished data).

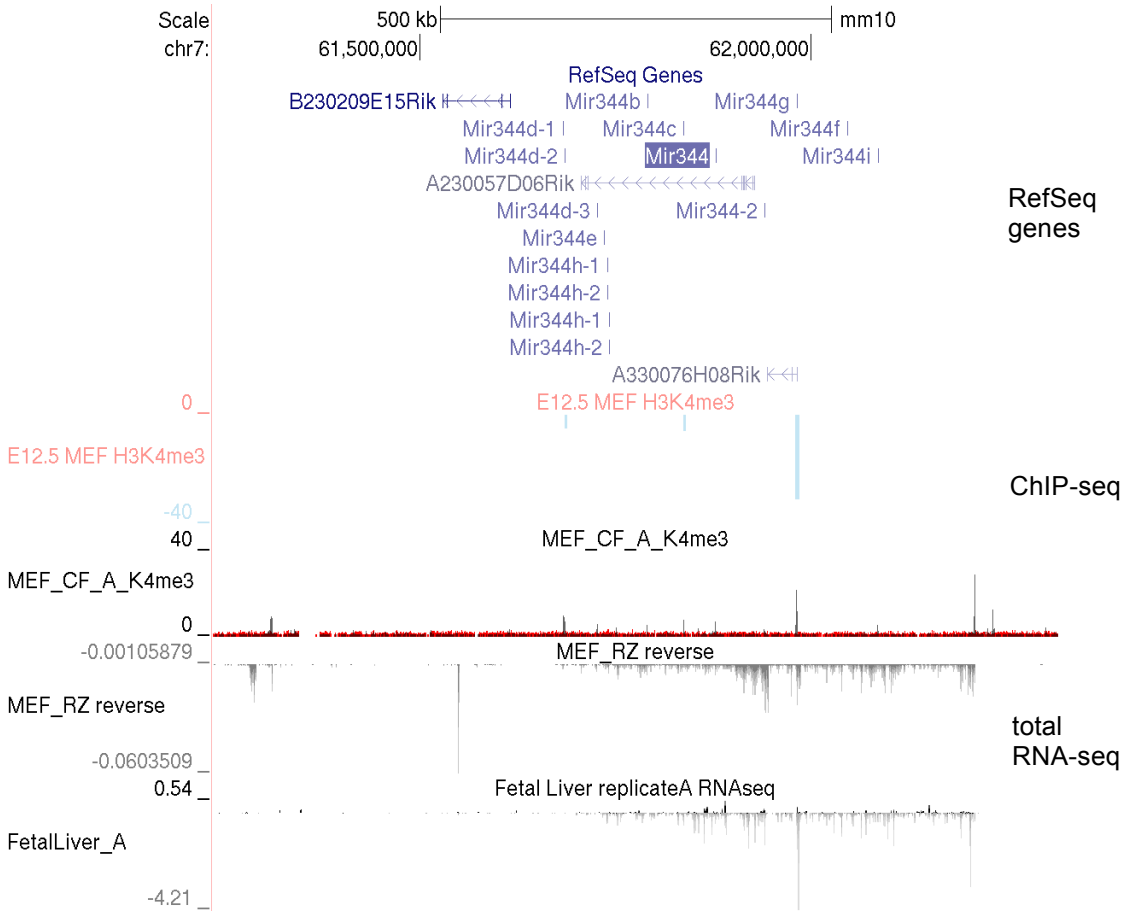

Fig. 28 Screenshot showing the 3 loci on chromosome 7 identified to show differential H3K4me3 enrichment in MEFs (*Mir344d-2*, *Mir344c* and *A330076H08Rik*). ChIP-seq shows focal peaks at TSS of all genes identified and the total RNA-seq track shows transcription throughout the entire region containing all three loci.

4.3 Future perspective

The present work is part of a larger study that is still ongoing. In my project I compared differential enrichment of the active histone marks H3K4me3 and H3K27ac at promoter regions in embryonic (MEFs and fetal liver) and extra-embryonic (VE) tissues by performing ChIP-seq on F1 mouse crosses and found that each tissue had a unique pattern of enrichment on the predicted active allele of imprinted genes. The aim of the larger study is to define the imprinted expression signature of tissues representative of the pluripotent (ES cells), embryonic (MEFs and fetal liver) and extra-embryonic (VE) lineages using RNA-seq and differential enrichment of H3K4me3 and H3K27ac over promoters. This will then be compared to differential enrichment of enhancer

(H3K27ac) and repressive histone marks (H3K9me2, H3K9me3 and H3K27me3) in the same tissues at imprinted loci. The bioinformatics imprinting pipeline used in the present work was developed by Daniel Andergassen and is still being modified and there are still some improvements to come. Therefore, reevaluation of the obtained results should be considered once the final version is available, although the general conclusions of my study are not likely to change.

During my master's project I also collected a sufficient amount of embryonic tissue (VE and fetal liver) and successfully validated antibodies suitable to be used in ChIP-seq for all repressive histone modifications mentioned above. The next steps would be to perform ChIP on the collected cells, validate the quality of the ChIP by using known positive and negative controls in qPCR, and prepare the ChIP-seq libraries for sequencing. For the analysis of these data I would call the peaks of the sequencing tracks and use this as an annotation, since I would also expect to find informative modifications outside of the RefSeq promoter regions used in this thesis.

To investigate genomic imprinting in pluripotent cells, I would prepare libraries and perform RNA-seq using RNA that I isolated from reciprocal crosses of FVB/N and CAST/EiJ F1 ES cells. I would then thaw up aliquots of the back-frozen 2i-adapted ES cells and grow them to near confluency to isolate enough chromatin for ChIP. Little is yet known about imprinted expression and differential enrichment of histone marks in undifferentiated ES cells. A study conducted by Mikkelsen et al. (2007) detected some allele-specific differences at imprinted loci for H3K4me3 and H3K36me3 indicating imprinted expression, and the association of H3K9me3 with the methylated allele of ICEs, but at the time they only could assess 3.5 million SNPs, did not have reciprocal crosses or biological replicates and their ES cells were grown with serum (Mikkelsen *et al.* 2007). The mouse strains used in the present work now have over 20 million SNPs defined (Keane *et al.* 2011), allowing more regions of the genome to be assessed than in Mikkelsen et al. (2007). We also have reciprocal crosses, biological replicates and a bioinformatics pipeline that uses empirical data to calculate the false discover rate allowing

us to be confident in our results. Finally, as I cultured our ES cells in 2i media, they should be a more homogenous population of undifferentiated cells than those cultured in serum by Mikkelsen et al. (2007), allowing us to assess imprinted expression and differential enrichment of histone modifications in a true pluripotent cell type.

It has been reported that the repressive histone modifying complexes EHMT2, PRC1 and PRC2 play more of a role in imprinted silencing in placenta than in the embryo (Umlauf *et al.* 2004; Nagano *et al.* 2008; Terranova *et al.* 2008; Wagschal *et al.* 2008), but genome-wide studies in an extra-embryonic tissue examining repressive modifications deposited by these complexes, such as H3K9me2, H3K9me3 and H3K27me3 have yet to be done. It has also been hypothesized that differences in chromatin state between embryonic and extra-embryonic tissues may explain why the silencing agent, such as an imprinted lncRNA, can extend its silencing range over a greater distance in extra-embryonic tissues (Kulinski *et al.* 2013). Why repressive histone modifying complexes would act specifically in extra-embryonic tissues to cause imprinted silencing has not been explained, given that neither these complexes nor the imprinted lncRNAs that should recruit and target these complexes to imprinted gene promoters are tissue-specific. An alternative model for EXEL imprinted expression has been proposed that suggests lncRNAs act by blocking tissue-specific enhancers (Pauler *et al.* 2012). Therefore, mapping of differential enrichment of H3K27ac (enhancers) and the repressive histone modifications H3K9me2, H3K9me3 and H3K27me3 in pluripotent, embryonic and extra-embryonic tissues should give some insights into chromatin differences between the different lineages that may explain how tissue-specific imprinted expression patterns arise.

5. MATERIALS AND METHODS

5.1 Methods

5.1.1 Mouse strains

The visceral yolk sac endoderm (VE) and fetal liver was isolated from embryonic day (E) 12.5 mouse embryos from reciprocal F1 crosses between the two genetically distinct mouse strains FVB/NJ and CAST/EiJ and pooled by genotype for ChIP-seq experiments. Over 20 million high quality single nucleotide polymorphisms (SNPs) have been annotated between these 2 strains (Keane *et al.* 2011). The same tissues were also collected from FVB +/+. For ChIP about 20-40 embryos were necessary per antibody and for each genotype biological replicates were done. The range of number of embryos is caused by quite a high variation of chromatin yield, making the number needed hard to predict.

5.1.1.1 Tissue preparation

Pregnant inbred mice (FVB/NJ and CAST/EiJ) were sacrificed at 12.5 days after conception and the embryos were dissected (reciprocal F1 crosses of FVB/NJ and CAST/EiJ or FVB +/+). First, the embryos were separated and taken out of the uterus. Second, the placenta, the PYS and the amnion were removed and the VYS was collected. The embryonic liver was collected and frozen in liquid nitrogen. The VYS was then incubated for 2.5 hours at 4°C in 1% Dispase II in DPBS and for an additional 30 minutes or longer in DPBS at 4°C (Hudson *et al.* 2011). The VE layer was separated from the VM layer mechanically using forceps under a microscope. Using a pipette, the VE was collected in a drop of DPBS on Parafilm and frozen as a droplet in liquid nitrogen. All dissections were done on ice and samples were stored on dry ice after freezing in liquid nitrogen before storing them at -80°C until needed. Livers and VE of one litter were pooled separately. Rita Casari deserves acknowledgement at this point for being a huge help with the dissections.

5.1.2 Cell culture

All cells (MEFs and ES cells) were grown at 37°C with 5% CO_2 in an incubator (Forma Series II Water-Jacketed CO_2 Incubator, Thermo Scientific). MEF cell culture was spatially separated from ES cell culture. All media were preheated to 37°C in a water bath prior to usage on cells.

5.1.2.1 MEFF cell culture

Primary mouse embryonic fibroblasts (pMEF) cells were made from E12.5 female embryos derived from reciprocal F1 crosses between FVB/NJ and CAST/EiJ by Daniel Andergassen (former Barlow Diploma student). Briefly, the embryonic body from E12.5 embryos was dissected minus the head, viscera and urogenital system and washed in DPBS (dissections by Quanah Hudson). The bodies were incubated separately in 3ml Trypsin (0.25% Trypsin/EDTA (1X) Gibco) for 20 minutes at 37°C and then an extra 2ml Trypsin was added. Each embryo body was then homogenized by pipetting up and down with a P-1000 tip and 4 ml of MEF media was added (DMEM, 10% FCS, pen/strep, L-glutamine) and the embryo was further homogenized by pipetting up and down with a P-1000 and then a 20G needle. The homogenized cells were then centrifuged (200g, 5min), the supernatant discarded and the cells resuspended in 4ml MEF media and plated on 0.1% gelatin coated 6cm dishes. The next day cells were then split 1:3 onto 10cm dishes and grown for 2 days until near confluence. RNA was then isolated from one 10cm dish for RNA-seq (passage 2). Remaining cells were split 1:3 onto 3 x T-175cm^2 flasks and grown until near confluence before being trypsinized, pooled in a 50ml Falcon tube, centrifuged (200g, 5min), resuspended in DPBS and centrifuged again to pellet the cells. All DPBS was then removed and the cell pellet frozen on dry ice and stored at -80°C until conducting the ChIP (passage 3). The pMEFs subject to RNA-seq and ChIP-seq were FVB x CAST (FC)15 (FC1, dissected 01.02.2013), FC21 (FC2, dissected 01.02.2013), CF9 (CF1, dissected 08.02.2013) and (CF2, dissected 08.02.2013). Note that the maternal allele or mother in a cross is always written on the left.

For testing of antibodies, the immortalized MEFB1 (+/T^{hp}) and MEFF (T^{hp}/+) cell lines were used (Regha *et al.* 2007) that carry a maternal or paternal 5mb deletion on mouse chromosome 17 including the *Igf2r* cluster (Johnson 1974; Barlow *et al.* 1991). The cells were grown using standard conditions with MEF media described above except that pen/strep was excluded.

5.1.2.2 Adaptation of feeder dependent ES cells to 2i medium

For adaptation of feeder dependent ES cells to 2i-medium culture conditions, ES cells of reciprocal crosses of mouse strains FVB/NJ and CAST/EiJ made by Christian Theussl (Transgenic Service, IMP/IMBA) and Daniel Wenzel (Josef Penninger laboratory, IMBA) based on a published protocol (Bryja *et al.* 2006) were used. Feeder dependent cells grown in ES medium were seeded onto a gelatinized 10cm or 6cm dish (when only half of a vial was used) containing a feeder layer prepared the day before and 8ml (10cm dish) or 2ml (6cm dish) of ES cell medium (D-MEM+ Hepes, 15% FCS, MEM non, L-Glutamine, Gentamicin, 2-mercaptoethanol, Na-Pyruvate, LIF). Feeder cells (FVB/Bl6^{Hygro} (p5) or FVB +/+ (p3)) were mitotically inactivated by 5min gamma irradiation. ES cells grown in 2i medium (ESGRO-2i Medium Millipore) were directly seeded on a gelatinized 6cm dish containing 2ml 2i medium. To detach ES cells grown in ES cell medium, the medium was removed, one wash with DPBS was performed and the cells were trypsinized (0.05% Trypsin-EDTA Gibco) for 5min at 37°C using 0.5 ml (6cm dishes) or 1ml (10cm dishes). Trypsin was stopped by adding 5ml (6cm dishes) or 10 ml (10cm dishes) of ES cell medium (with serum), pipetting up and down to obtain a single cell suspension and centrifuging (178xg, 5min). Cells were either harvested by transferring the pellet on dry ice and later to -80°C or resuspended in an appropriate amount of ES cell medium and seeded in a desired splitting ratio on a new dish containing ES cell medium. To detach cells grown in 2i medium, the medium was removed, 0.5ml (6cm dishes) or 1ml (10cm dishes) of Accutase (Millipore) was added and the dish was incubated for 3-5min at 37°C. By adding 5ml of DMEM/F12 (Millipore) and by pipetting up and down, a single cell suspension was obtained and pelleted by centrifuging (161xg, 5min). The pellet was resuspended in 5ml DMEM/F12,

the cells were counted using Casy Counter (Innovatis) and centrifuged again (161xg, 5min). For further culturing, the pellet was resuspended in 2i medium depending on the cell number, aiming to have about one million cells per 0.5-1ml. To harvest the cells, after centrifugation the media was removed, one wash with DPBS was performed and the cell pellet was frozen on dry ice and later stored at -80°C. For RNA isolation from ES cells see 5.1.9.

5.1.2.3 Gelatinizing dishes

Dishes for ES cell culture were gelatinized using 0.2% gelatin in DPBS. For 6cm dishes 2ml and for 10cm dishes 4ml were used and the dishes were incubated for at least 20min at 37°C. Prior to seeding of cells, the gelatin solution was sucked off leaving a gelatin coating on the plate.

5.1.2.4 Feeder depletion using sedimentation/selective adhesion

Cells were transferred to a non-gelatinized dish and incubated for 20min at 37°C. During that time, feeder cells preferentially attach to the dish and ES cells remain floating free. By gently tilting the plate and transferring the medium to a new dish, most of the feeders were removed since they stay attached.

5.1.2.5 Back freezing cells

ES cells and MEFs were back frozen the same way. In general, one dense 10cm dish was split into three vials. Cells were harvested and the pellet was resuspended in 500µl MEF, ES cell or 2i medium per future tube. 500µl of freezing medium (80% FBS, 20% DMSO) was mixed with 500µl cell suspension in cryotubes (Nunc) and immediately transferred to a precooled Mr. Frosty Freezing Container (Thermo Scientific) and put at -80°C for about 24h. For longtime storage, the tubes were transferred to a liquid nitrogen tank.

5.1.2.6 Counting cells

All cells were counted using CASY Model TT (Innovatis) with recommended settings according to the cell types.

5.1.3 Native Chromatin Immunoprecipitation (nChIP)

The protocol for native ChIP originally published by Umlauf et al. (2004) was adapted by Regha et al. (2007) aiming to minimize unspecific binding by adding a pre-clearing step, pre-treating the protein A sepharose beads and improving the washing steps (Umlauf *et al.* 2004; Regha *et al.* 2007). For the present work the protocol was further adapted for use in embryonic tissue.

5.1.3.1 Sample preparation

Embryonic tissue and cells were used as ChIP starting material. Cells frozen as a pellet in a 50ml Falcon and stored at -80°C were resuspended in 50ml cold DPBS by rotating for 5min at 4°C followed by centrifugation (272xg, 6min, 4°C). The DPBS was removed and the remaining pellet was resuspended in 50ml cold DPBS (washing step). A total of 3 DPBS washing steps were performed. Embryonic liver was grinded up under liquid nitrogen in a precooled mortar using a pestle, scraped into a 50ml Falcon tube and resuspended in 50ml cold DPBS. VE frozen as a drop was directly put into 50ml cold DPBS and resuspended by rotating for 5min at 4°C. A total of three DPBS washing steps were performed as described above.

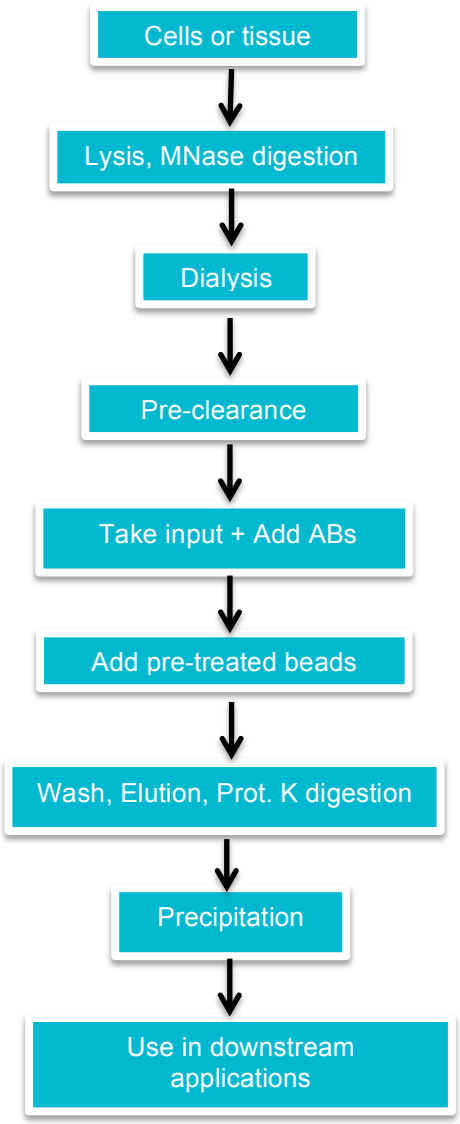

Fig. 29 Main steps of the nChIP protocol

5.1.3.2 Cell lysis, MNase digestion and dialysis

The cell pellet was resuspended in 2ml lysis buffer A and then mixed with 2ml lysis buffer B by pipetting up and down, and incubated on ice for 10min. 8ml MNase buffer A was then added and the mixture was centrifuged (3023xg, 10min, 4°C). The supernatant was removed and 12ml MNase buffer A was added and the pellet resuspended by pipetting followed by centrifugation (3023xg, 10min, 4°C). Again the supernatant was removed and the pellet resuspended in 12ml MNase buffer B (without EDTA) followed by centrifugation (3023xg, 10min, 4°C). All supernatant was then removed and the pellet resuspended in 2ml MNase buffer B. MNase was diluted to 20U/µl in MNase buffer B and 4µl was added to the sample in 2ml MNase buffer B. The sample was aliquoted 4 x 0.5ml into 1.5ml Eppendorf tubes and MNase digested for 7min at 37°C in a water bath. The digestion was stopped by adding 20µl 0.5M EDTA and quickly mixing by flicking the tubes. The sample was then centrifuged (16100xg, 15min, 4°C) and the supernatant (S1 fraction) was transferred to a new 1.5ml Eppendorf tubes. The remaining pellet was resuspended in dialysis buffer (S2 fraction), the four tubes for each fraction were then pooled and filled into dialysis bags. Both fractions were then dialyzed in dialysis buffer with gently stirring o/n at 4°C.

5.1.3.3 Testing cutting efficiency, pre-clearing and adding antibodies

The samples were removed from dialysis bags, centrifuged (16100xg, 15min, 4°C) and the supernatants from each fraction were pooled. The fractions were quantified using Nanodrop (peqlab) and the total amount of chromatin was calculated. In order to check for MNase cutting efficiency, 1µg of each fraction in 40 µl (or a maximum of 40µl) mixed with 4µl 20% SDS and 4µl loading dye were loaded on a 2% agarose gel. The S1 and S2 fractions were pooled, mixed with pre-clearing mix and 400µl protein A sepharose beads in a 15ml Falcon tube and incubated rotating for at least 4h at 4°C. The pre-cleared chromatin was then centrifuged (272xg, 10min, 4°C) and the supernatant was transferred to a new 15ml Falcon tube. The chromatin was divided evenly into new 15ml Falcons for ChIP, with 1/10th of the chromatin used per ChIP set

aside as input and stored at 4°C (the volume of the ChIP and input differed depending on the number of ChIPs conducted per sample). Antibodies were then added to each individual ChIP reaction and incubated rotating o/n. In parallel, 100µl of protein A sepharose beads (1:1 beads: water) were pre-treated by removing the water, adding 500µl of the pre-treatment mix and incubated rotating o/n at 4°C.

5.1.3.4 Adding of beads, washing, elution, proteinase K digestion

Pre-treated beads were centrifuged (200xg, 5min, 4°C) and the pre-treatment solution was removed. The beads were then resuspended in 500µl antibody/chromatin mix and added back to the antibody/chromatin mix followed by incubation rotating for 1-2h at 4°C. The antibody/chromatin/beads mix was centrifuged (188xg, 5min) and the supernatant was removed. Three different wash buffers (1, 2 and 3) were used with increasing NaCl concentration (125mM, 250mM and 500mM). Washing was done by adding 1ml wash buffer, rotating for 5min and centrifuging (188xg, 5min), performing a total of 3 washes with each wash buffer. After the last wash step the beads were resuspended in wash buffer and transferred to 1.5ml Eppendorf tubes and centrifuged (200xg, 5min). The wash buffer was then removed and 200µl pre-heated elution buffer was added and the beads incubated for 15min at 65°C in a water bath and rotating for 5min at RT. The sample was centrifuged (200xg, 5min) and the supernatant transferred to a new 1.5ml Eppendorf tube. Elution was then repeated, and the 200µl supernatant transferred to the same tube to make 400µl eluted sample. The input samples were then brought to a total volume of 400µl by adding elution buffer, and then 20µl proteinase K [10mg/ml] was added to both the ChIP sample and the input followed by incubation for at least 2h or o/n at 55°C. Finally the samples were phenol-chlorophorm extracted (see 5.1.4), precipitated (see 5.1.5) and the pellet was resuspended in 100-200µl TE buffer depending on the amount of chromatin.

5.1.3.5 Preparation of protein A sepharose beads

Protein A sepharose beads (GE Healthcare) were supplied as powder and stabilizing agents that needed to be washed away prior to using them in ChIP. Therefore, the required amount was weighted out in a 15ml Falcon tube, H_2O was added to 15ml and the tube was rotated for 5min at 4°C. The beads were then put for 1h at 4°C to allow them to settle and the water was removed. A total of 3 washing rounds were performed and finally the beads were aliquoted to 100 and 400µl (1:1 beads: water) in Eppendorf tubes.

5.1.4 Phenol-Chlorophorm-Extraction

An equal amount as the sample of phenol:chloroform:isoamyl Alcohol was added, vortexed and centrifuged (16100xg, 5min). The upper, aqueous phase was transferred to a new tube and an equal amount of phenol was added. After vortexing, the mixture was centrifuged (16100xg, 5min) and the upper, aqueous phase was transferred to a new tube.

5.1.5 Precipitation of nucleic acids

Nucleic acids were precipitated using 0.1 vol. 3M NaAc, 2.5 vol. ethanol and 2µl of glycogen. Precipitation took place o/n or longer at -20 °C. The sample was centrifuged (16100xg, 15min, 4°C), the arising pellet was washed with an appropriate amount of 70% ethanol followed by centrifugation (16100xg, 5min). All ethanol was removed, the pellet air-dried for about 3min at RT and resuspended in a volume of TE buffer depending on pellet size and end use of the DNA.

5.1.6 Quantitative Polymerase Chain Reaction (qPCR)

SYBRGreen and TaqMan assays were performed on an ABI PRISM 7000 Sequence Detection System (Applied Biosystems) in optical 96-well plates, 25µl reaction volumes each (5µl template + 20µl mastermix). The results were calculated using the ABI PRISM 7000 Sequence Detection System software,

version 2.3 (Applied Biosystems) and Microsoft Excel 2011. Standard curves were calculated using duplicates of input samples diluted in 4 steps, 1:4 each. Samples were used as triplicates and the mean and standard deviation was calculated. ChIP samples were normalized to the input samples and cut samples in feeder depletion assay were normalized relative to the uncut samples.

5.1.6.1 SYBRGreen

Mastermix SYBRGreen	final conc.	
2x Mesa Green	1x	12.5µl
fwd. and rev. primer (0.5µM)	0.05µM	2.5µl
embryo water	-	5µl
DNA template	-	5µl

Cycling conditions

50°C 2min
95°C 10min
95°C 15sec
60°C 1min } 40x
95°C 15sec
60°C 15sec

5.1.6.2 TaqMan

Mastermix TaqMan	final conc.	
2x universal Mastermix	1x	12.5µl
forward primer [100µM]	900nM	0.23µl
reverse primer [100µM]	900nM	0.23µl
probe	200nM	0.05µl
embryo water	-	7µl
DNA template	-	5µl

Cycling conditions

50°C 2min
95°C 10min
95°C 15sec
60°C 1min } 40x
60°C 1min

5.1.7 ChIP sample library preparation

Libraries for sequencing were prepared using TruSeq ChIP Sample Prep Kit (Illumina). If possible 5-10ng (but as little as 1.3ng worked fine) of input DNA was used. All steps were carried out according to the TruSeq ChIP Sample Preparation Guide (Illumina) with minor changes indicated in the following. At this point I would like to acknowledge Philipp Guenzl for showing me how to prepare the libraries and for patiently answering all my questions about the protocol. Incubation steps were done at RT, unless otherwise stated. For some library preparations, only half reaction volume was used. 1.5 and 2.0ml

DNA LoBind Tubes (Eppendorf) were used instead of the 96-well PCR plate suggested in the protocol. During the PCR steps, samples were transferred to standard 0.2ml PCR tubes (SARSTEDT).
Ethanol washes and resuspension steps: with the tube on the magnetic stand (MagnaRack, Invitrogen), 200µl of freshly prepared 80% ethanol was added, incubated for 5sec and removed. Another 200µl of 80% ethanol were added, incubated for 30sec and removed. The samples were air-dried for 15min, an indicated amount of resuspension buffer was added (see below) and the sample was incubated for 2min. The tubes were transferred to the magnetic stand, incubated for 5min and an indicated amount was transferred to a new tube (see below).

5.1.7.1 Quantification of the ChIP result, end repair

All quantifications were done using Qubit Fluorometer, dsDNA HS Assay Kit (Life Technologies). 50µl of ChIP DNA was mixed with 10µl resuspension buffer and 40µl of end repair mix and incubated for 30min at 30°C. 160µl of AMPure XP Beads (Beckman Coulter) were added and incubated for 15min. The tube was put on the magnetic stand for 15min and all supernatant was removed, followed by an ethanol wash and resuspension using 16µl resuspension buffer. Finally, 15µl were transferred to a new tube.

5.1.7.2 Adenylation of 3' ends and ligation of adapters

2.5µl of resuspension buffer and 12.5µl of A-tailing mix were added and the tube was incubated for 30min at 37°C and for 5min at 70°C. 2.5µl of resuspension buffer and 2.5µl of ligation mix were added. 2.5µl of the desired RNA adapter index was added and the tube was incubated for 10min at 30°C. 5µl of stop ligation buffer and 42.5µl of AMPure XP Beads were added and incubated for 15min. The tube was transferred to the magnetic stand, incubated for 5min and the supernatant was removed followed by an ethanol wash and resuspension using 51µl of resuspension buffer. 50µl were transferred to a new tube and another round of cleaning using 50µl AMPure

XP Beads was done, followed by an ethanol wash and resuspension using 21µl of resuspension buffer. 20µl were transferred to a new tube.

5.1.7.3 Purification of ligation products

150ml of a 2% agarose gel in TAE containing 15µl SYBRGold was prepared. The samples were mixed with 6x loading dye and loaded on the gel, leaving a gap between the ladder and samples and also between the samples if possible. DirectLoad Wide Range DNA Marker was used and the gel was run for 2h at 92V. The gel was photographed (UVsolo TS, Biometra) and bands between 250-300bp were cut using a scalpel. To control the cutting, another photo was taken. Gel extraction was done following the protocol of the QIAquick gel extraction kit (Qiagen) (using a microcentrifuge, all centrifugation steps at 16100xg, 1min), by dissolving the gel slice in 300µl buffer QG and incubation for 20min at 37°C. 10µl of 3M NaAc were added if the color of the mixture was not yellow. 83µl of isopropanol were added and the entire volume was applied to a MinElute Spin Columns 50 (QIAGEN) followed by centrifugation. The flow-through was discarded, 0.5ml Buffer QG were added and centrifuged. Washing was done using 0.75ml buffer PE and centrifugation. The flow-through was discarded and the column was centrifuged again. Elution was done by adding 21µl of buffer EB and centrifugation. Finally, 20µl eluted DNA was transferred to a new tube.

5.1.7.4 Enrichment of DNA fragments via PCR

5µl of PCR primer cocktail and 25µl of PCR master mix were added to the sample and the following PCR program was used:

98°C 5min
60°C 30sec ┐
72°C 30sec ├ 16-17x
72°C 5min ┘
4°C hold

50µl of AMPure XP Beads were added, incubated for 15min and put for 5min on the magnetic stand. The supernatant was removed, an ethanol wash and

resuspension with 17.5µl resuspension buffer was performed. 15µl were transferred to a new tube.

5.1.7.5 Validation, normalization and pooling of the libraries

This step was done by the Biomedical Sequencing Facility (BSF) at CeMM using an Experion Automated Electrophoresis System (Experion DNA 1K Analysis Kit) (Bio-Rad). The validated libraries were diluted to a concentration of 2nM using 5µl of library mixed with Qiagen EB buffer with 0.1% Tween 20 and the samples were pooled using 15µl according to the lanes they should be loaded on. Such multiplexed samples were sent for 50bp SE sequencing on a Illumina HiSeq 2000 machine, performed by the BSF at CeMM.

5.1.8 Methylation assay of *Igf2r* and *Airn* via Southern blot

Thanks to Justyna Konecka for introducing me to the Southern blot technique.

5.1.8.1 Isolation of genomic DNA

A cell pellet in a 1.5ml Eppendorf tube was resuspended in 700µl DNA lysis solution and incubated o/n at 55°C. 300µl of saturated NaCl solution was added and the sample was mixed for 20sec. The tube was centrifuged (16100xg, 10min) and the supernatant transferred to a new tube containing 600µl of isopropanol, and by inverting the tube several times, the DNA was precipitated. The tube was centrifuged (16100xg 7min), the supernatant removed and the tube centrifuged again (16100xg, 1min). The rest of the supernatant was discarded and 800µl of 70% ethanol were added to wash the pellet. The tube was incubated at RT for 15min, shaking it every 5min and then centrifuged (16100xg, 7min). The supernatant was removed, the tube was centrifuged again (16100xg, 1min) and the rest of the supernatant was removed. The pellet was resuspended in 150µl-200µl TE buffer and incubated o/n at 55°C to dissolve the genomic DNA.

5.1.8.2 Digestion of DNA

Both, single enzyme and double enzyme digests were done using EcoRI in all digests (as a control) and for double digests additionally a methylation sensitive restriction enzyme cutting at the promoter region of *Igf2r* and *Airn* was added (MluI and NotI, respectively). Digestions were done o/n at 37°C.

Single enzyme digest	
genomic DNA	30µl
EcoRI	2µl
buffer EcoRI	4µl
water	4µl

Double enzyme digest	
genomic DNA	30µl
EcoRI	2µl
methylation sensitive enzyme	2µl
buffer for double digest	4µl
water	2µl

5.1.8.3 Blotting

Samples mixed with 4µl loading buffer were loaded on a 0.8% agarose gel in 1xTBE and the gel was run for 5.5h at 80V. The gel was stained in EtBr for 1h and photographed. To denature the DNA in order to make it single stranded so it could bind to the bind to the membrane, the gel was incubated shaking 2x for 30min in denaturing solution, exchanging the solution after 30min. To prepare the blotting-sandwich a plastic tray filled with the denaturing solution used in the denaturing steps covered with a glass plate was used. Two sheets of 3MM Whatman paper cut according to the width of the gel and long enough that they reach to the denaturing buffer were put on the glass plate and the gel was put upside down on top of it. The Hybond XL blotting membrane cut to size of the sample area was activated by incubating it for 1min in H_20 followed by incubation in denaturing buffer for 5min. The membrane was put on top of the gel and the area around the gel was covered with plastic strips. Next, two more sheets of 3MM Whatman paper cut to the size of the gel were placed and 10-15cm of paper towels and a glass plate was put on top. While building the blot, a glass pipette was used to roll over the surface in order to avoid air bubbles and finally to press together. The entire blotting sandwich, a

blotting weight of about 200ml water was put on top and blotting was done over the weekend.

5.1.8.4 Probe labeling and probe cleaning

20ng of probe (Msi or En-4kb) were mixed with H_2O to a final volume of 14µl and the cold mix was prepared. In the hot lab, the probe was denatured by putting the tubes at 100°C for 5min followed by incubation on ice for 3min. To the cold mix, 2-4µl of α-^{32}P-dATP (amount depending on the age of the radioactivity) were added and this was mixed with the probe and incubated o/n at RT. A sephadex column was prepared by plugging a 1ml syringe with aquarium wool and filling it with sephadex equilibrated in a TE buffer. The syringe was put into a 15ml Falcon tube and 60µl of TE buffer were added to the labeled probe. The mixture was loaded onto the sephadex column and centrifuged (1450xg, 3min) to remove salt and single nucleotides not incorporated during labeling. 1µl of the cleaned probe was diluted in 99µl TE buffer and counts were read on a scintillation counter (Tri-Carb 2810 TR, PerkinElmer).

5.1.8.5 Hybridization

The blotting sandwich was disassembled, the membrane was incubated for 2min in Na_2HPO_4 and the membrane was pre-hybridized rotating in a small hybridization tube filled with 30ml Church buffer for 30min at 65°C. The cleaned probe was put for 5min at 100°C and for 2min on ice. Church buffer used for pre-hybridization was removed and 20ml fresh Church buffer together with the probe were added. Hybridization was done for about 18h. The next day, the Church buffer was removed, 20ml Church wash buffer pre-warmed to 65°C was added and incubated rotating for 30min in the hybridization oven. This washing step was repeated.

5.1.8.6 Exposure and scanning

The membrane was put into a plastic foil and sealed. It was exposed to an imaging plate (Fujifilm) o/n and finally scanned on a Typhoon 8600 Variable Mode Imager (GE Healthcare). The pictures were linearly adjusted using Photoshop CS5.

5.1.9 RNA extraction of ES cells/tissues

The cell pellet of a dense 10cm dish was trypsinized, centrifuged and resuspended in 2ml TriReagent (Sigma), split to 1ml each and incubated for 10min at RT. Tissues (VE and VYS) were transferred from -80°C to a RNAse free tube (Ambion) containing 1ml TriReagent and also incubated for 10min at RT. 100µl of BCP were added, incubated for 10min at RT and centrifuged (12000xg, 15min, 4°C). The upper phase was transferred to a new tube containing 500µl isopropanol and incubated for 5min at RT. The mixture was centrifuged (12000xg, 8min), the supernatant trashed and the pellet washed with 75% ethanol followed by centrifugation (7500xg, 5min). All ethanol was removed, the pellet air-dried for 5min at RT and resuspended in 100µl RNA storage solution. The yield was measured using Nanodrop.

5.1.10 cDNA preparation

For cDNA preparation, the DNAfree (Invitrogen) and the RevertAid (Thermo Scientific) kits were used. 6µg of RNA were dissolved in 28.7µl water, 3.3µl of DNAfree 10x buffer and 1µl of DNase was added. This was incubated for 30min at 37°C and 5µl of inactivation solution was added followed by incubation for 2min at RT. The sample was centrifuged (16100xg, 1.5min, 4°C) and the supernatant was transferred to a new tube. 3µl of random hexamer primers was added and incubated for 5min at 70°C. The sample was spun down and put on ice. 21µl of pre-mix was added and the sample was divided into 1x38µl (2x +RT) and 1x19µl (1x -RT). Both samples were incubated for 5min at 25°C and 2µl of reverse transcriptase was added to the

+RT sample only. Finally, both samples were incubated using the following program.
25°C 10min
42°C 60min
70°C 10min
4°C hold

5.1.11 Bioinformatics analysis

Bioinformatics analysis was done with the help of Christoph Dotter and Tomasz Kulinski.

5.1.11.1 Screenshots genome browser

All genome browser screenshots were made using the UCSC Genome Bioinformatics Site (http://genome.ucsc.edu).

5.1.11.2 Conversion of .bam files to .fastq files

All sequencing results for VE obtained from the BSF as .bam files needed to be converted to .fastq files in order to be aligned. Therefore, Picard command-line tools (http://picard.sourceforge.net) with standard settings were used.

5.1.11.3 Analysis of sequencing quality

For quick analysis of the sequencing quality, the program FastQC High Throughput Sequence QC Report version 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used.

5.1.11.4 Alignment of sequencing results

Alignment of ChIP-seq and RNA-seq data was done with GSNAP (Wu & Nacu 2010).

5.1.11.5 Imprinting pipeline to detect allele-specific chromatin state

To analyze ChIP-seq and RNA-seq data, an imprinting pipeline set up by Daniel Andergassen and refined by Christoph Dotter (former and present students in the Barlow lab, respectively) was used (Andergassen et al., in preparation). This pipeline is modified from a published pipeline (DeVeale *et al.* 2012). The pipeline uses a list of about 20 million high quality SNPs between FVB/N and Cast/EiJ recently published by the Sanger Institute (Keane *et al.* 2011). Taking advantage of these SNPs, the reads can be assigned according to their parental origin and for each SNP a SNP-score is calculated. In the RNA-seq total RNA was used, which also allows using intronic SNPs, resulting in a more robust analysis as SNP coverage is increased within a gene. Using biological replicates of the reciprocal crosses, a four way comparison is possible, resulting in a robust imprinting score and by comparing the two biological replicates (mock comparison) the false discovery rate can be set empirically. The imprinting score is calculated based on coverage in the sequencing and deviation from the expected ratio of alleles (50:50 paternal: maternal) minus standard deviation resulting from the four comparisons. Only genes that show a robust imprinting score in all four comparisons give a score over the FDR and are called.

5.1.12 Venn diagrams

Venn diagrams were made in R, Version 2.15.1 using the Venn diagram package.

5.2 Materials

5.2.1 Cell lines

MEFF and MEFB1 (Regha *et al.* 2007)
primary MEFs of E12.5 of reciprocal crosses of FVB/NJ and CAST/EiJ (derived and grown by Daniel Andergassen)

ES cells of reciprocal F1 crosses of FVB/NJ and CAST/EiJ made by Daniel Wenzel (Josef Penninger laboratory, IMBA) and Christian Theussl (Transgenic Service IMBA/IMP)

5.2.2 Mouse strains

E12.5 of reciprocal F1 crosses of FVB/NJ and CAST/EiJ
E12.5 FVB +/+

5.2.3 ChIP-antibodies

Modification	Cat. Nr.	Product name	Company	Lot. Nr.
H3K4me3	07-473	Anti-trimethyl-Histone H3 (Lys4)	Millipore	2019729
H3K9me2	ab1220	Anti-Histone H3 (di methyl K9) [mAbcam 1220] antibody	Abcam	GR124129-2
H3K9me2	CST4658	Di-Methyl-Histone H3 (Lys9) (D85B4) #4658	CST	1
H3K9me3	4861	non commercial	Jenuwein	6th bleed
H3K9me3	MAb-153-050	H3K9me3 monoclonal antibody	diagenode	001
H3K9me3	SN-146-100	Monoclonal Rabbit anti-Histone H3	diagenode	001
H3K27ac	ab4729	Anti-Histone H3 (acetyl K27) antibody (ab4729)	Abcam	GR104852-2
H3K27me3	ab6002	Anti-Histone H3 (tri methyl K27) [mAbcam 6002] antibody	Abcam	GR123526-1
H3K27me3	6523	non commercial	Jenuwein	5th bleed
Anti-chicken IgG	C2288	Anti-Chicken IgY (IgG) (whole molecule) antibody	Sigma	084K4752

5.2.4 Primer list

5.2.4.1 TaqMan

Name	Sequence	Probe
Igf2r-97_F	CACTTGCAACACTAAACATCAACCT	ACTCCATCTCGGCCACCACCGTACTGGTC
Igf2r-97_R	CGCTTCCTAACTCTCTCTTCTTCAA	
Airn-125_F	CTGAGCTTTCCCTTCCCTTTC	ACCGCAACTCAGCACAACCAAGGATC
Airn-125_R	AGCAATTCCGGTTGT	
AirnEndFwd	GGACTGGCTCAGGCAAGCT	CCTGCTCGAGTTGCCATTCCCAGA
AirnEndRev	TTCAGTCAAAAATCCAAAACATGTG	

5.2.4.2 SYBRGreen

Name	forward	reverse
IAP (LTR)	CTCCATGTGCTCTGCCTTCC	CCCCGTCCCTTTTTTAGGAGA
MusD (LTR)	CCCTTCCTTCATAACTGGTGTCGCA	TAGCATCTCTCTGCCATTCTTCAGG
Mage-a2	TTGGTGGACAGGGAAGCTAGGGGA	CGCTCCAGAACAAAATGGCGCAGA

Afp enhancer1	CGCACAGCTTCTCATTCACTCCCA	GACAGGCATTGGCTTATCGCTTCC
Atoh1 enhancer B	GCGGGCAAAGACAGGTGGT	CAGTTAGGAGCCAGAAGCAAAGGG
MyoD1 enhancer	GGCTCCAGAATGGGCTCGGTT	CTGCTAATCTGCCACTCTTTGGGC
BV1	GGGTTCCTTTCAGTTTTCCATC	AACGGCAGAAGTGACCACAAG
BV5	TGCGCCTCCCCTCTTTATTT	TTCAGCAAGTGAACAACAGACATC
BV14	GCAGCATCTTGGAAAGGAACA	GACAGAGACTGGCTTCCAGTTGA
BV21	AAGGCTTTCCCAGGATTTGC	CTTAGAGACCCCACCACCTTTG
BV25	ACCACATCATGTAAACGGTGTTTC	CCCCCTACTGCTGGTATTTGAA
BV32	GACATGACCTCCAGCTCCAAA	TGGGCATCACATGAGCTAACTAG
Gapdh	GAGTAGCTGGGCCTCTCTCA	TGCGTGCACATTTCAAAAA
Flk1	GGGATGGTCCTTGCATCAGAA	ACTGGTAGCCACTGGTCTGGTTG
Afp	GCATGCTGCAAAGCTGACAA	AATGGTTGTTGCCTGGAGGTT
Actin prom	CCATGGTGTCCGTTCTGAGT	CTTCGCTCTCTCGTGGCTAG

5.2.5 Southern blot probes

Msi (chr17:12741514-12742531)

En-4kb (chr17:12,769,447-12,773,408)

5.2.6 Solutions

Lysis buffer A	**amount per sample**
Incomplete lysis buffer (50mM Tris HCl pH 8.0, 100mM KCl, 5mM $MgCl_2$, 50% glycerol)	4ml
2-mercaptoethanol	62.5µl
cOmplete protease inhibitor (1 tablet in 1ml H_20)	160µl

Lysis buffer B	**amount per sample**
Lysis buffer A	2ml
5% saponin	80µl

MNase buffer A	**amount per sample**
Incomplete MNase buffer (50mM Tris HCl pH 7.5, 4mM $MgCl_2$, 1mM $CaCl_2$ 0.32 sucrose)	20ml
cOmplete protease inhibitor (1 tablet in 1ml H_20)	800µl

MNase buffer B	amount per sample
Incomplete MNase buffer (50mM Tris HCl pH 7.5, 4mM $MgCl_2$, 1mM $CaCl_2$ 0.32 sucrose)	15ml
EDTA-free protease inhibitor (1 tablet in 1ml H_20)	1 tablet

Dialysis buffer	
Tris HCl pH 8.0	10mM
EDTA	10mM
PMSF in ethanol	0.01M

Pre-clearing mix	amount per sample
S1 and S2 fraction plus dialysis buffer	5.25ml
10% Triton-X	525µl
20% SDS	3µl
5M NaCl	60µl
complete protease inhibitor (1 tablet in 1ml H_20)	225µl

Pre-treatment solution	
Tris HCl pH 8.0	10mM
EDTA pH 8.0	10mM
Triton-X	1%
SDS	0.01%
NaCl	50mM
BSA	0.5mg/ml
salmon sperm DNA	0.25mg/ml
cOmplete protease inhibitor (1 tablet in 1ml H_20)	50µl in 4ml

Wash buffer 1	
Tris HCl pH 8.0	50mM
EDTA pH 8.0	10mM
NaCl	125mM

Wash buffer 2	
Tris HCl pH 8.0	50mM
EDTA pH 8.0	10mM
NaCl	250mM

Wash buffer 3	
Tris HCl pH 8.0	50mM
EDTA pH 8.0	10mM
NaCl	500mM

Elution buffer	
Tris HCl pH 8.0	50mM
EDTA pH 8.0	10mM
SDS	1% SDS

DNA lysis solution	
TEN pH 9 (250mM Tris pH 9.0, 100mM EDTA pH 8.0, 200mM NaCl)	1x
SDS	1%
Proteinase K	0.5mg/ml

Denaturing solution	
NaOh	0.5M
NaCl	1.5M

Church buffer	
Na_2HPO_4	250mM
SDS	7%
EDTA	1mM

Church wash buffer	
Na_2HPO_4	20mM
SDS	1%

LS mix	
1M Hepes pH 6.6	25ml
OL (50U hexamer primer dissolved in 1.6ml TE pH 8.0	1ml
TM (250mM Tris ph8 8.0, 25mM $MgCl_2$, 50mM 2-mercaptoethanol	25ml

CTG mix	
dCTP	**100µM**
dGTP	**100µM**
dTTP	**100µM**
BSA	**2mg**

Cold mix	
LS mix	20µl
CTG mix	6µl
Klenow fragment	1µl

Freezing medium	
FBS	80%
DMSO	20%

ES cell medium	
FCS (sterile filtered)	30ml
MEM non	2ml
L- Glutamine	2ml
Gentamicin	1ml
2-mercaptoethanol	0.4ml
Na-Pyruvate	2ml
LIF	depends on batch
D-MEM+ Hepes	to 200ml

MEF medium	
FCS	20ml
L-glutamine	2ml
Gentamicin	1ml
D-MEM	to 200ml

TE buffer	
Tris HCl pH 8.0	10mM
EDTA	1mM

TAE buffer	
Tris	40mM
Acetic acid	20mM
EDTA	1mM

TBE buffer	
Tris	90mM
Boric acid	90mM
EDTA	2mM

5.2.7 Chemicals and consumables

.5% Trypsin-EDTA	Gibco
0.25% Trypsin-EDTA	Gibco
2-mercaptoethanol	Sigma
3MM Whatman paper	Whatmann
6X DNA Loading Dye	Thermo Scientific
Accutase	Millipore
Acetic acid	Applichem
Agarose LE	Biozym
AMPure XP Beads	Beckman Coulter
Boric acid	Applichem
Bovine serum albumin (BSA)	Fermentas
Bromo-chloropropane (BCP)	MRC
BSA	Ambion
Calcium chloride (CaCl)	Sigma
Chlorophorm	Sigma
cOmplete Protease Inhibitor Cocktail Tablet, EDTA-free	Roche
cOmplete Protease Inhibitor Cocktail Tablets	Roche
dCTP	Bioron
dGTP	Bioron
Dialysis tubing	Thermo Scientific
Dimethyl sulfoxide (DMSO)	Sigma
DirectLoad Wide Range DNA Marker	Sigma
Disodium hydrogen phosphate (Na2HPO4)	Merck
DMEM/F12, with HEPES, L-Glutamine	Millipore
dTTP	Bioron
Dulbecco's Modified Eagle Medium high glucose, HEPES (DMEM)	Gibco
Dulbecco's Modified Eagle Medium, high glucose, pyruvate (DMEM)	Gibco
Dulbecco's Phosphate-Buffered Saline (DPBS)	Gibco
ESGRO-2i medium	Millipore
Ethanol 96%	Merck
Ethidium bromide	Applichem
Ethylenediaminetetraacetic acid (EDTA)	Merck
Fetal bovine serum (FBS)	PAA
Gelatin	Sigma
GeneRuler 1 kb DNA Ladder	Thermo Scientific
GeneRuler 100 bp Plus DNA Ladder	Thermo Scientific
Gentamycin	Invitrogen
Glycerol	Applichem
Glycogen	Ambion
Hybond XL blotting membrane	GE Healthcare
Isopropanol	Merck
L-Glutamine	Gibco
Loading dye	Fermentas
MEM Non essential amino acids	Gibco

MESA GREEN qPCR Plus	Eurogentec
Magnesium chloride ($MgCl_2$)	Sigma
Phenol:Chloroform:Isoamyl Alcohol	Applichem
Phenylmethylsulfonyl fluoride in ethanol (PMSF)	Sigma
Potassium chloride (KCl)	Sigma
qPCR mastermix plus	Eurogentec
Random hexamer primer	GE Healthcare
RNA storage solution	Ambion
Salmon Sperm DNA	Invitrogen
Saponin	Sigma
SephadexTm G-50	Amersham
Sepharose protein A beads	GE Healthcare
Sieve 3:1 Agarose	Biozym
Sodium acetate (NaAc)	Ambion
Sodium chloride (NaCl)	Applichem
Sodium dodecyl sulfate (SDS)	Applichem
Sodium hydroxide (NaOH)	Applichem
Sodium pyruvate	Sigma
Sucrose powder	EMPROVE bio
SYBRGold Nucleic Acid Gel Stain	Invitrogen
TriReagent	Sigma
Tris	Applichem
Triton-X	Sigma
Tween 20	Sigma
Water for embryo transfer	Sigma
α-^{32}P-dATP	Hartmann Analytic

5.2.8 Buffer and enzymes

10x Buffer O	Thermo Scientific
10x Buffer R	Thermo Scientific
10x Buffer EcoRI	Thermo Scientific
EcoRI	Thermo Scientific
MluI	Thermo Scientific
NotI	Thermo Scientific
MNase	Thermo Scientific
Proteinase K	Applichem
Dispase II	Roche
Klenow fragment	Fermentas
DNase I	Ambion
Reverse transcriptase	Fermentas

5.2.9 Kits

QIAquick Gel extraction kit	Qiagen
TruSeq ChIP Sample Preparation	Invitrogen
Qubit dsDNA HS Assay Kit	Invitrogen
DNA-free Kit	Invitrogen
RevertAid First Strand cDNA Synthesis Kit	Thermo Scientific

6. REFERENCES

Barlow D.P. (1995) Gametic imprinting in mammals. *Science* **270**, 1610-3.

Barlow D.P. (2011) Genomic imprinting: a Mammalian epigenetic discovery model. *Annu Rev Genet* **45**, 379-403.

Barlow D.P., Stoger R., Herrmann B.G., Saito K. & Schweifer N. (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* **349**, 84-7.

Bartolomei M.S. & Ferguson-Smith A.C. (2011) Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol* **3**.

Bartolomei M.S., Zemel S. & Tilghman S.M. (1991) Parental imprinting of the mouse H19 gene. *Nature* **351**, 153-5.

Bell A.C. & Felsenfeld G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-5.

Ben-Arie N., McCall A.E., Berkman S., Eichele G., Bellen H.J. & Zoghbi H.Y. (1996) Evolutionary conservation of sequence and expression of the bHLH protein Atonal suggests a conserved role in neurogenesis. *Human molecular genetics* **5**, 1207-16.

Berger S.L., Kouzarides T., Shiekhattar R. & Shilatifard A. (2009) An operational definition of epigenetics. *Genes & development* **23**, 781-3.

Bryja V., Bonilla S. & Arenas E. (2006) Derivation of mouse embryonic stem cells. *Nat Protoc* **1**, 2082-7.

Cai X. & Cullen B.R. (2007) The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA* **13**, 313-6.

Caspary T., Cleary M.A., Baker C.C., Guan X.J. & Tilghman S.M. (1998) Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Molecular and cellular biology* **18**, 3466-74.

Cattanach B.M. (1986) Parental origin effects in mice. *Journal of embryology and experimental morphology* **97 Suppl**, 137-50.

Chapman V., Forrester L., Sanford J., Hastie N. & Rossant J. (1984) Cell lineage-specific undermethylation of mouse repetitive DNA. *Nature* **307**, 284-6.

Creyghton M.P., Cheng A.W., Welstead G.G., Kooistra T., Carey B.W., Steine E.J., Hanna J., Lodato M.A., Frampton G.M., Sharp P.A., Boyer L.A., Young R.A. & Jaenisch R. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-6.

Curley J.P., Barton S., Surani A. & Keverne E.B. (2004) Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. *Proc Biol Sci* **271**, 1303-9.

da Rocha S.T., Charalambous M., Lin S.P., Gutteridge I., Ito Y., Gray D., Dean W. & Ferguson-Smith A.C. (2009) Gene dosage effects of the imprinted delta-like homologue 1 (dlk1/pref1) in development: implications for the evolution of imprinting. *PLoS genetics* **5**, e1000392.

Dao D., Frank D., Qian N., O'Keefe D., Vosatka R.J., Walsh C.P. & Tycko B. (1998) IMPT1, an imprinted gene similar to polyspecific transporter and multi-drug resistance genes. *Human molecular genetics* **7**, 597-608.

DeChiara T.M., Robertson E.J. & Efstratiadis A. (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849-59.

Deltour L., Montagutelli X., Guenet J.L., Jami J. & Paldi A. (1995) Tissue- and developmental stage-specific imprinting of the mouse proinsulin gene, Ins2. *Developmental biology* **168**, 686-8.

Derrien T., Guigo R. & Johnson R. (2011) The Long Non-Coding RNAs: A New (P)layer in the "Dark Matter". *Front Genet* **2**, 107.

DeVeale B., van der Kooy D. & Babak T. (2012) Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS genetics* **8**, e1002600.

Dong K.B., Maksakova I.A., Mohn F., Leung D., Appanah R., Lee S., Yang H.W., Lam L.L., Mager D.L., Schubeler D., Tachibana M., Shinkai Y. & Lorincz M.C. (2008) DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *The EMBO journal* **27**, 2691-701.

Downs K.M. & Davies T. (1993) Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255-66.

Dunwoodie S.L. & Beddington R.S. (2002) The expression of the imprinted gene Ipl is restricted to extra-embryonic tissues and embryonic lateral mesoderm during early mouse development. *The International journal of developmental biology* **46**, 459-66.

Egelhofer T.A., Minoda A., Klugman S., Lee K., Kolasinska-Zwierz P., Alekseyenko A.A., Cheung M.S., Day D.S., Gadel S., Gorchakov A.A., Gu T., Kharchenko P.V., Kuan S., Latorre I., Linder-Basso D., Luu Y., Ngo Q., Perry M., Rechtsteiner A., Riddle N.C., Schwartz Y.B., Shanower G.A., Vielle A., Ahringer J., Elgin S.C., Kuroda M.I., Pirrotta V., Ren B., Strome S., Park P.J., Karpen G.H., Hawkins R.D. & Lieb J.D. (2011) An assessment of histone-modification antibody quality. *Nat Struct Mol Biol* **18**, 91-3.

Ferguson-Smith A.C. (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nature reviews. Genetics* **12**, 565-75.

Fleming T.P., Warren P.D., Chisholm J.C. & Johnson M.H. (1984) Trophectodermal processes regulate the expression of totipotency within the inner cell mass of the mouse expanding blastocyst. *Journal of embryology and experimental morphology* **84**, 63-90.

Gabory A., Ripoche M.A., Le Digarcher A., Watrin F., Ziyyat A., Forne T., Jammes H., Ainscough J.F., Surani M.A., Journot L. & Dandolo L. (2009) H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* **136**, 3413-21.

Giddings S.J., King C.D., Harman K.W., Flood J.F. & Carnaghi L.R. (1994) Allele specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting. *Nature genetics* **6**, 310-3.

Green K., Lewis A., Dawson C., Dean W., Reinhart B., Chaillet J.R. & Reik W. (2007) A developmental window of opportunity for imprinted gene silencing mediated by DNA methylation and the Kcnq1ot1 noncoding RNA. *Mammalian genome : official journal of the International Mammalian Genome Society* **18**, 32-42.

Gregg C., Zhang J., Weissbourd B., Luo S., Schroth G.P., Haig D. & Dulac C. (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329**, 643-8.

Guttman M. & Rinn J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339-46.

Hagan J.P., O'Neill B.L., Stewart C.L., Kozlov S.V. & Croce C.M. (2009) At least ten genes define the imprinted Dlk1-Dio3 cluster on mouse chromosome 12qF1. *PloS one* **4**, e4352.

Hagiwara Y., Hirai M., Nishiyama K., Kanazawa I., Ueda T., Sakaki Y. & Ito T. (1997) Screening for imprinted genes by allelic message display: identification of a paternally expressed gene impact on mouse chromosome 18. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9249-54.

Hnisz D., Abraham B.J., Lee T.I., Lau A., Saint-Andre V., Sigova A.A., Hoke H.A. & Young R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47.

Hudson Q.J., Kulinski T.M., Huetter S.P. & Barlow D.P. (2010) Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. *Heredity* **105**, 45-56.

Hudson Q.J., Seidl C.I., Kulinski T.M., Huang R., Warczok K.E., Bittner R., Bartolomei M.S. & Barlow D.P. (2011) Extra-embryonic-specific imprinted expression is restricted to defined lineages in the post-implantation embryo. *Developmental biology* **353**, 420-31.

Johnson D.R. (1974) Hairpin-tail: a case of post-reductional gene action in the mouse egg. *Genetics* **76**, 795-805.

Johnson M.H. (2009) From mouse egg to mouse embryo: polarities, axes, and tissues. *Annu Rev Cell Dev Biol* **25**, 483-512.

Keane T.M., Goodstadt L., Danecek P., White M.A., Wong K., Yalcin B., Heger A., Agam A., Slater G., Goodson M., Furlotte N.A., Eskin E., Nellaker C., Whitley H., Cleak J., Janowitz D., Hernandez-Pliego P., Edwards A., Belgard T.G., Oliver P.L., McIntyre R.E., Bhomra A., Nicod J., Gan X., Yuan W., van der Weyden L., Steward C.A., Bala S., Stalker J., Mott R., Durbin R., Jackson I.J., Czechanski A., Guerra-Assuncao J.A., Donahue L.R., Reinholdt L.G., Payseur B.A., Ponting C.P., Birney E., Flint J. & Adams D.J. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-94.

Kobayashi S., Wagatsuma H., Ono R., Ichikawa H., Yamazaki M., Tashiro H., Aisaka K., Miyoshi N., Kohda T., Ogura A., Ohki M., Kaneko-Ishino T. & Ishino F. (2000) Mouse Peg9/Dlk1 and human PEG9/DLK1 are paternally expressed imprinted genes closely located to the maternally expressed imprinted genes: mouse Meg3/Gtl2 and human MEG3. *Genes Cells* **5**, 1029-37.

Kulinski T.M., Barlow D.P. & Hudson Q.J. (2013) Imprinted silencing is extended over broad chromosomal domains in mouse extra-embryonic lineages. *Current opinion in cell biology* **25**, 297-304.

Kuzmin A., Han Z., Golding M.C., Mann M.R., Latham K.E. & Varmuza S. (2008) The PcG gene Sfmbt2 is paternally expressed in extraembryonic tissues. *Gene expression patterns : GEP* **8**, 107-16.

Kwon G.S., Viotti M. & Hadjantonakis A.K. (2008) The endoderm of the mouse embryo arises by dynamic widespread intercalation of

embryonic and extraembryonic lineages. *Developmental cell* **15**, 509-20.

Landgraf P., Rusu M., Sheridan R., Sewer A., Iovino N., Aravin A., Pfeffer S., Rice A., Kamphorst A.O., Landthaler M., Lin C., Socci N.D., Hermida L., Fulci V., Chiaretti S., Foa R., Schliwka J., Fuchs U., Novosel A., Muller R.U., Schermer B., Bissels U., Inman J., Phan Q., Chien M., Weir D.B., Choksi R., De Vita G., Frezzetti D., Trompeter H.I., Hornung V., Teng G., Hartmann G., Palkovits M., Di Lauro R., Wernet P., Macino G., Rogler C.E., Nagle J.W., Ju J., Papavasiliou F.N., Benzing T., Lichter P., Tam W., Brownstein M.J., Bosio A., Borkhardt A., Russo J.J., Sander C., Zavolan M. & Tuschl T. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-14.

Latos P.A., Pauler F.M., Koerner M.V., Şenergin H.B., Hudson Q.J., Stocsits R.R., Allhoff W., Stricker S.H., Klement R.M., Warczok K.E., Aumayr K., Pasierbek P. & Barlow D.P. (2012) Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science* **338**, 1469-72.

Latos P.A., Stricker S.H., Steenpass L., Pauler F.M., Huang R., Senergin B.H., Regha K., Koerner M.V., Warczok K.E., Unger C. & Barlow D.P. (2009) An in vitro ES cell imprinting model shows that imprinted expression of the Igf2r gene arises from an allele-specific expression bias. *Development* **136**, 437-48.

Lerchner W. & Barlow D.P. (1997) Paternal repression of the imprinted mouse Igf2r locus occurs during implantation and is stable in all tissues of the post-implantation mouse embryo. *Mech Dev* **61**, 141-9.

Lewis A., Mitsuya K., Umlauf D., Smith P., Dean W., Walter J., Higgins M., Feil R. & Reik W. (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nature genetics* **36**, 1291-5.

Luger K., Mader A.W., Richmond R.K., Sargent D.F. & Richmond T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60.

Lyle R., Watanabe D., te Vruchte D., Lerchner W., Smrzka O.W., Wutz A., Schageman J., Hahner L., Davies C. & Barlow D.P. (2000) The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nature genetics* **25**, 19-21.

Mager D.L. & Freeman J.D. (2000) Novel mouse type D endogenous proviruses and ETn elements share long terminal repeat and internal sequences. *Journal of virology* **74**, 7221-9.

Mancini-DiNardo D., Steele S.J., Ingram R.S. & Tilghman S.M. (2003) A differentially methylated region within the gene Kcnq1 functions as an imprinted promoter and silencer. *Human molecular genetics* **12**, 283-94.

McGrath J. & Solter D. (1983) Nuclear transplantation in mouse embryos. *J Exp Zool* **228**, 355-62.

McGrath J. & Solter D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**, 179-83.

Menheniott T.R., Woodfine K., Schulz R., Wood A.J., Monk D., Giraud A.S., Baldwin H.S., Moore G.E. & Oakey R.J. (2008) Genomic imprinting of

Dopa decarboxylase in heart and reciprocal allelic expression with neighboring Grb10. *Molecular and cellular biology* **28**, 386-96.
Mikkelsen T.S., Ku M., Jaffe D.B., Issac B., Lieberman E., Giannoukos G., Alvarez P., Brockman W., Kim T.K., Koche R.P., Lee W., Mendenhall E., O'Donovan A., Presser A., Russ C., Xie X., Meissner A., Wernig M., Jaenisch R., Nusbaum C., Lander E.S. & Bernstein B.E. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60.
Miri K. & Varmuza S. (2009) Imprinting and extraembryonic tissues-mom takes control. *International review of cell and molecular biology* **276**, 215-62.
Mitiku N. & Baker J.C. (2007) Genomic analysis of gastrulation and organogenesis in the mouse. *Developmental cell* **13**, 897-907.
Mohamed O.A., Clarke H.J. & Dufort D. (2004) Beta-catenin signaling marks the prospective site of primitive streak formation in the mouse embryo. *Developmental dynamics : an official publication of the American Association of Anatomists* **231**, 416-24.
Moore T. & Haig D. (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in genetics : TIG* **7**, 45-9.
Morison I.M., Paton C.J. & Cleverley S.D. (2001) The imprinted gene and parent-of-origin effect database. *Nucleic acids research* **29**, 275-6.
Morison I.M., Ramsay J.P. & Spencer H.G. (2005) A census of mammalian imprinting. *Trends in genetics : TIG* **21**, 457-65.
Nagano T., Mitchell J.A., Sanz L.A., Pauler F.M., Ferguson-Smith A.C., Feil R. & Fraser P. (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717-20.
Nagy A., Gertsenstein M., Vintersten K. & Behringer R. (2003) *Manipulating the Mouse Embryo: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Okae H., Hiura H., Nishida Y., Funayama R., Tanaka S., Chiba H., Yaegashi N., Nakayama K., Sasaki H. & Arima T. (2012) Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Human molecular genetics* **21**, 548-58.
Pauler F.M., Barlow D.P. & Hudson Q.J. (2012) Mechanisms of long range silencing by imprinted macro non-coding RNAs. *Current opinion in genetics & development* **22**, 283-9.
Pauler F.M., Sloane M.A., Huang R., Regha K., Koerner M.V., Tamir I., Sommer A., Aszodi A., Jenuwein T. & Barlow D.P. (2009) H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome research* **19**, 221-33.
Pauler F.M., Stricker S.H., Warczok K.E. & Barlow D.P. (2005) Long-range DNase I hypersensitivity mapping reveals the imprinted Igf2r and Air promoters share cis-regulatory elements. *Genome research* **15**, 1379-87.
Peach S.E., Rudomin E.L., Udeshi N.D., Carr S.A. & Jaffe J.D. (2012) Quantitative assessment of chromatin immunoprecipitation grade antibodies directed against histone modifications reveals patterns of co-occurring marks on histone protein molecules. *Mol Cell Proteomics* **11**, 128-37.

Peters J. & Williamson C.M. (2007) Control of imprinting at the Gnas cluster. *Epigenetics* **2**, 207-13.
Popp C., Dean W., Feng S., Cokus S.J., Andrews S., Pellegrini M., Jacobsen S.E. & Reik W. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101-5.
Prickett A.R. & Oakey R.J. (2012) A survey of tissue-specific genomic imprinting in mammals. *Molecular genetics and genomics : MGG* **287**, 621-30.
Proudhon C. & Bourc'his D. (2010) Identification and resolution of artifacts in the interpretation of imprinted gene expression. *Briefings in functional genomics* **9**, 374-84.
Regha K., Sloane M.A., Huang R., Pauler F.M., Warczok K.E., Melikant B., Radolf M., Martens J.H., Schotta G., Jenuwein T. & Barlow D.P. (2007) Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. *Molecular cell* **27**, 353-66.
Reik W. & Walter J. (2001) Genomic imprinting: parental influence on the genome. *Nature reviews. Genetics* **2**, 21-32.
Ribet D., Harper F., Dewannieux M., Pierron G. & Heidmann T. (2007) Murine MusD retrotransposon: structure and molecular evolution of an "intracellularized" retrovirus. *Journal of virology* **81**, 1888-98.
Richmond T.J. & Davey C.A. (2003) The structure of DNA in the nucleosome core. *Nature* **423**, 145-50.
Rossant J. & Cross J.C. (2001) Placental development: lessons from mouse mutants. *Nature reviews. Genetics* **2**, 538-48.
Rossant J., Sanford J.P., Chapman V.M. & Andrews G.K. (1986) Undermethylation of structural gene sequences in extraembryonic lineages of the mouse. *Developmental biology* **117**, 567-73.
Rossant J. & Tam P.P.L. (2002) *Mouse development : patterning, morphogenesis, and organogenesis*. Harcourt Academic, San Diego, Calif. ; London.
Royo H., Bortolin M.L., Seitz H. & Cavaille J. (2006) Small non-coding RNAs and genomic imprinting. *Cytogenetic and genome research* **113**, 99-108.
Royo H. & Cavaille J. (2008) Non-coding RNAs in imprinted gene clusters. *Biol Cell* **100**, 149-66.
Santoro F. & Barlow D.P. (2011) Developmental control of imprinted expression by macro non-coding RNAs. *Semin Cell Dev Biol*.
Seidl C.I., Stricker S.H. & Barlow D.P. (2006) The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *The EMBO journal* **25**, 3565-75.
Sleutels F., Zwart R. & Barlow D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810-3.
Smallwood S.A. & Kelsey G. (2012) De novo DNA methylation: a germ cell perspective. *Trends in genetics : TIG* **28**, 33-42.
Surani M.A. & Barton S.C. (1983) Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. *Science* **222**, 1034-6.

Surani M.A., Barton S.C. & Norris M.L. (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548-50.
Szabo P.E. & Mann J.R. (1995) Allele-specific expression and total expression levels of imprinted genes during early mouse development: implications for imprinting mechanisms. *Genes & development* **9**, 3097-108.
Takagi N. & Sasaki M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* **256**, 640-2.
Tanaka S., Kunath T., Hadjantonakis A.K., Nagy A. & Rossant J. (1998) Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072-5.
Tarkowski A.K., Witkowska A. & Opas J. (1977) Development of cytochalasin in B-induced tetraploid and diploid/tetraploid mosaic mouse embryos. *Journal of embryology and experimental morphology* **41**, 47-64.
Terranova R., Yokobayashi S., Stadler M.B., Otte A.P., van Lohuizen M., Orkin S.H. & Peters A.H. (2008) Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Developmental cell* **15**, 668-79.
Tilghman S.M. (1985) The structure and regulation of the alpha-fetoprotein and albumin genes. *Oxf Surv Eukaryot Genes* **2**, 160-206.
Tran D.A., Bai A.Y., Singh P., Wu X. & Szabo P.E. (2014) Characterization of the imprinting signature of mouse embryo fibroblasts by RNA deep sequencing. *Nucleic acids research* **42**, 1772-83.
Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J. & Pachter L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**, 511-5.
Umlauf D., Goto Y., Cao R., Cerqueira F., Wagschal A., Zhang Y. & Feil R. (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nature genetics* **36**, 1296-300.
Varmuza S. & Mann M. (1994) Genomic imprinting--defusing the ovarian time bomb. *Trends in genetics : TIG* **10**, 118-23.
Wagschal A., Sutherland H.G., Woodfine K., Henckel A., Chebli K., Schulz R., Oakey R.J., Bickmore W.A. & Feil R. (2008) G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Molecular and cellular biology* **28**, 1104-13.
Wang H. & Dey S.K. (2006) Roadmap to embryo implantation: clues from mouse models. *Nature reviews. Genetics* **7**, 185-99.
Wang X., Soloway P.D. & Clark A.G. (2011) A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics* **189**, 109-22.
Wang X., Sun Q., McGrath S.D., Mardis E.R., Soloway P.D. & Clark A.G. (2008) Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PloS one* **3**, e3839.
Weaver J.R. & Bartolomei M.S. (2013) Chromatin regulators of genomic imprinting. *Biochim Biophys Acta*.

Weisstein A.E., Feldman M.W. & Spencer H.G. (2002) Evolutionary genetic models of the ovarian time bomb hypothesis for the evolution of genomic imprinting. *Genetics* **162**, 425-39.

Whyte W.A., Orlando D.A., Hnisz D., Abraham B.J., Lin C.Y., Kagey M.H., Rahl P.B., Lee T.I. & Young R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19.

Williamson C.M., Blake A., Thomas S., Beechey C.V., Hancock J., Cattanach B.M. & Peters J. (2014) MRC Harwell, Oxfordshire. World Wide Web Site - Mouse Imprinting Data and References - http://www.har.mrc.ac.uk/research/genomic_imprinting/.

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

Wu T.D. & Nacu S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873-81.

Wutz A. (2011) Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nature reviews. Genetics* **12**, 542-53.

Wutz A., Smrzka O.W., Schweifer N., Schellander K., Wagner E.F. & Barlow D.P. (1997) Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745-9.

Xie W., Barr C.L., Kim A., Yue F., Lee A.Y., Eubanks J., Dempster E.L. & Ren B. (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816-31.

Ying Q.L., Wray J., Nichols J., Batlle-Morera L., Doble B., Woodgett J., Cohen P. & Smith A. (2008) The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-23.

Zentner G.E. & Henikoff S. (2013) Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20**, 259-66.

Zhang Y., Liu T., Meyer C.A., Eeckhoute J., Johnson D.S., Bernstein B.E., Nusbaum C., Myers R.M., Brown M., Li W. & Liu X.S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137.

Zwart R., Sleutels F., Wutz A., Schinkel A.H. & Barlow D.P. (2001) Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. *Genes & development* **15**, 2361-6.

7. CURRICULUM VITAE

Philipp Bammer

EDUCATION

2010-2014	Master's program 'Genetics and Developmental Biology', University of Vienna
08/2012- 02/2013	Erasmus exchange program, Lund University (Sweden)
2007-2010	Bachelor's program 'Biology', University of Vienna
1999-2007	High school Kirchdorf an der Krems

RESEARCH EXPERIENCE

02/2013-02/2014	Master's project on 'Allele-specific chromatin state in embryonic and extra-embryonic mouse tissues', group of Denise Barlow, CeMM Research Center for Molecular Medicine, Vienna
08/2012- 02/2013	Research project on 'Identification and functional analysis of the *EST1* gene in *Saccharomyces castellii*', group of Marita Cohn, Lund University
03- 05/2012	Research project on 'The role of *Candida albicans* Hat1 and Asf1 in histone processing', group of Karl Kuchler, Medical University of Vienna
11/2010- 01/2011	Research project on 'Intraspecific variations in the *Rps5* gene in myriapods', group of Günther Pass, Department of Evolutionary Biology, University of Vienna
11/2009	Bachelor's project on 'DNA Barcoding in papilionoidean caterpillars', Department of Population Ecology, University of Vienna

SCHOLARSHIPS

Performance Fellowship of the University of Vienna 2012

Erasmus Scholarship 2012

8. ACKNOWLEDGEMENTS

First of all I would like to express my gratitude to Denise Barlow for giving me the unique opportunity to work in such a great and welcoming scientific surrounding. Thanks for sharing your enthusiasm for science and for your helpful support during writing this thesis and when I was looking for a PhD position. It was really a pleasure working in such a well-organized and supporting environment. This holds true for entire Ce-M-M-, which I now consider a perfect place to work.

Great thanks to my supervisor Quanah Hudson for his guidance and support, but also for letting me work independently and appreciating my own ideas. He did an amazing job while checking this thesis and patiently answered all my questions. Thank you for reading my thesis even on the weekend and for being there when the deadline was looming.

A huge thank you to Christoph Dotter for making me familiar with bioinformatics tools and helping with problems and also to Tomasz Kulinski for showing some bioinformatics basics. Christoph was there whenever a problem with analyzing the data came up and despite he was very busy with his own work, he always found time to help me. He became a friend and his support was not only appreciated in the lab but also when belying me during our climbing sessions. Thank you also for your great taste in music, a raclette set is waiting for you!

A big thanks is dedicated to Sabine and Sarah, without whom the project would not have been possible. They really worked hard to breed the mice, which indeed was not easy all the time. At this point I would also like to thank Rita for helping with dissections and for good talks while doing so.

I would also like to express my appreciation to Florian Pauler, for his helpful and relaxed attitude, which makes it really easy to talk to him. Thanks for kindly and patiently answering all my questions and for mixing scientific facts with a bit of humor, making them way easier to understand.

Thanks to Christian Seiser for being my official supervisor and to Gerda Egger for her support during the time I was looking for a master's position.

Of course the entire project would not have been as much fun without the company of Aleksandra, Daniel, Daniela, Justyna, Markus and Philipp. Thanks a lot for good talks during coffee breaks and for cheering one up after an experiment failed. Especially Alexandra was very understanding and with her unique attitude she always managed somehow to put a smile on my face.

Last but certainly not least thanks to my best friend Julia for sharing experiences and enthusiasm for science and my family and all my friends for giving support whenever needed. Thanks also to everyone not specifically mentioned in this text, but supporting me, there are many people who helped and encouraged me throughout the project, which I am very grateful for.