PhD degree in Molecular Medicine (curriculum in Molecular Oncology and Human Genetics)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: Bio/18

Neuronal Transdifferentiation: Uncovering The Role of MLL1 and MLL2 During Lineage Conversion

Giulia Barbagiovanni

European Institute of Oncology, Milan

Matricola n. R09851

Supervisor: Prof. Giuseppe Testa

European Institute of Oncology, Milan

University of Milan, Milan

Anno accademico 2015-2016

1. TABLE OF CONTENTS

2. FIGURES AND TABLES INDEX

3. LIST OF ABBREVIATIONS

4-OHT 4-hydroxytamoxifene

5-azaC 5-azacytidine

Achetylcolinesterase Ache

Adcy1 Type-1 Adenylyl cyclase

Agap2 ArfGAP with GTPase domain, ankyrin repeat and PH domain 2

ALS Amyotrophic Lateral Sclerosis

ApoE Apolipoptrotein E

Arnt2 Aryl hydrocarbon receptor nuclear translocator 2

Bad Bcl-2 associated agonist of cell death

Bax Bcl-2 associated X protein

Bcl-2 B-cell lymphoma 2

bp base pairs

BP Basal Progenitor

Cadm3 Cell adhesion molecule 3

Cbx7 Chromobox 7

CDKI Cyclin Dependent Kinase Inhibitors

C/EBPα CCAAT/enhancer binding protein α

CFP1 CXXC finger protein 1

ChIP-seq Chromatin Immunoprecipitation coupled with sequencing

ChIP-qPCR Chromatin Immunoprecipitation coupled to detection by quantitative real-time PCR

Chst10 Carbohydrate sulfotransferase 10

Clstn3 Calcyntenin 3

DMEM Dulbecco's Modified Eagle Medium

Dock4 Dedicator of cytokinesis 4

Dox doxycycline

DTT Dithiothreitol

EBs Embryoid Bodies

Efnb2 Ephrinb2

EPSC excitatory post-synaptic currents

ESCs Embryonic Stem Cells

FACS Fluorescence-Activated Cell Sorting

FAD Familial Alzheimer's Disease

FBS Fetal Bovine Serum

FC Fold Change

FDR False Discovery Rate

Foxl2 Forkhead box L2

FPKM Fragments Per Kilobase of exon per Million fragments mapped

FYRC phenylalanine and tyrosine-rich C-terminal

FYRN phenylalanine and tyrosine-rich N-terminal

GluA1 Glutamate A1

Gnao1 Guanine nucleotide binding protein, alpha O

GO Gene Ontology

H3K4me3 trimethylated histone 3 lysine 4

H3K27me3 trimethylated histone 3 lysine 27

hCG human Chorionic Gonadotropin

HEK293T Human Embryonic Kidney 293T

HOX genes Homeotic genes

IMDM Iscove's Modified Dulbecco's Medium

iN induced Neuronal cell

iNPC induced Neural Precursor Cell

IPSC Inhibitory Post-Synaptic Current

iPSCs induced Pluripotent Stem Cells

JNK c-Jun N-terminal kinase

Jph3 Junctophilin 3

Jph4 Junctophilin 4

Kif5a Kinesin family member 5A

KMT2 histone-lysine N-methyltransferase 2

KO Knock-out

Lrfn1 Leucine-rich repeat and fibronectin type III domain-containing protein 1

MAP2b Microtubule associated protein 2b

MEF Mouse Embryonic Fibroblast

Morc2a MORC family CW-type finger protein 2A

mRNA messenger RNA

Myo16 Myosin 16

MYOD1 Myogenic Differentiation 1

Neurl1a Neuralized homolog 1 a

NSCs Neural Stem Cells

NYAP Neuronal tYrosine-phosphorylated Adaptor for the PI 3-kinase

PCA Principal Component Analysis

PcG Polycomb group proteins

Pdcd5 Programmed cell death 5

PDZ PSD-95/Discs large/ZO-1

Phactr2 phosphatase and actin regulator 2

PI Propidium Iodide

PI3 phosphoinositide 3

PI3 kinase Phosphatidylinositol 3 kinase

PIC Protease Inhibitor Cocktail

PMSG Pregnant Mare Serum Gonadotropin

Ppp1r9a Protein phosphatase 1 regulatory subunit 9A

PRC1 Polycomb Repressive Complex 1

PRC2 Polycomb Repressive Complex 2

PRE Polycomb Responsive Element

Prka1b type I-beta regulatory subunit of protein kinase A

PSA-NCAM Polysialylated-neural cell adhesion molecule

Ptch1 patched homolog 1

RG Radial Glia

RNA-seq RNA-sequencing

RPKM Reads per Kilobase per Million mapped reads

rRNA ribosomal RNA

Rtn1 Reticulon 1

RT-qPCR real-time quantitative PCR

Runx1t1 Runt-related transcription factor 1 translocated to 1

SAM S-adenosyl-methionine

SAM Sterile Alpha Motif

SCNT Somatic Cell Nuclear Transfer

SD Standard Deviation

SEM Standard Error of the Mean

Set1 SET domain-containing 1

Sh Sonic hedgehog

Sh3pxd2a SH3 and PX domain-containing protein 2A

Shank3 SH3/ankyrin domain gene 3

Slc2a3 solute carrier family 2 member 3

SPF Specific Pathogen Free

Srsf12 Serine/Arginine-Rich Splicing Factor 12

Sult4a1 Sulfontransferase 4A member 1

Suv39h1 Suppressor of variegation 3-9 homolog 1

SVZ Subventrincular Zone

Syngr Synaptogyrin

SYP Synaptophysin

TF Transcription Factor

TrxG Trithorax group proteins

TSS Transcription Start Site

Tug1 Taurine upregulated 1

WAVE WASP-family verprolin homologous protein

wt wild type

XAF1 X-linked IAP-associated factor-1

XIAP X-linked inhibitor of apoptosis

Zdbf2 Zinc finger, DBF-type containing 2

4. ABSTRACT

Transdifferentiation entails the direct conversion between terminally differentiated cells, without passing through the pluripotent state. Strikingly, through the exogenous expression of few transcription factors (TFs), the starting cell epigenome is restructured, allowing the generation of a cell that can be embryologically distant. This is the case of BAM (*i.e.*, Brn2, Ascl1, Myt1l) factors, whose transduction in mouse embryonic fibroblasts (MEFs), deriving from the mesoderm, allows their transdifferentiation into induced neuronal cells (iNs), which derive instead from the ectoderm.

The gene networks that characterize each MEFs-to-iNs transdifferentiation phase have been recently described, but to date it is still unclear how BAM factors guide such a conspicuous epigenetic remodelling.

MLL1 and MLL2 are two H3K4 trimethylases, belonging to the Trithorax protein family, discovered for their role in the regulation of *Hox* genes, which are fundamental for cell identity specification. Although during neuronal development and differentiation the function of the Trithorax axis has not been clearly identified, some hints suggest the involvement of MLL1 and MLL2 in both *in vivo* and *in vitro* neuronal differentiation.

Therefore, I studied the role of MLL1 and MLL2 during MEF-to-iNs transdifferentiation.

Whether transdifferentiation could envisage the same temporal epigenetic changes of physiological differentiation is still an open debate. What is evident is that in these two types of lineage specification (normal differentiation and transdifferentiation) starting cells have very different epigenomes, so a complete overlap of the epigenetic remodelling steps seems implausible. However, transdifferentiation remains a good tool to gain more insights both on neuronal differentiation and on the driving factors of transdifferentiation itself, which are often “terminal selector genes” or TFs fundamental during cell specification in development.

First, I extensively characterized the impact of the deletion of either *Mll1*, *Mll2* or both, measuring transdifferentiation efficiency and iNs maturation. In particular, I showed that the absence of MLL1 does not affect neither transdifferentiation efficiency nor iNs neuronal morphology. *Mll1* deletion is only detrimental for the survival rate of transdifferentiating MEFs. On the contrary, transdifferentiation efficiency is compromised in the absence of *Mll2,* and *Mll2*$^{-/-}$ iNs show an impairment in neurite elongation. The co-deletion of *Mll1* and *Mll2* impinges on cell viability as the knock-out of *Mll1* and it further exacerbates the *Mll2*$^{-/-}$ transdifferentiation defect. Moreover the few *Mll1*$^{-/-}$*Mll2*$^{-/-}$ iNs generated have very short neurites. These results suggest a role for MLL2-mediated H3K4 methylation in the control of transdifferentiation. Therefore, I defined the direct and indirect MLL2 targets through the integrative analysis of: i) the RNA-seq on iNs, ii) the ChIP-seq for MENIN, the specific common subunit of MLL1 and MLL2, for the delineation of the direct MLL2 targets which were not compensated by MLL1 and iii) the H3K4me3 ChIP-seq, to identify which are the genes that need to be trimethylated to allow transdifferentiation. I showed that in the absence of *Mll2* a conspicuous fraction of the transcriptome is down-regulated and/or loses the H3K4me3 mark. Therefore, MLL1, despite being the MLL2 homolog, is not capable of compensating for the absence of MLL2 during transdifferentiation. Moreover, many deregulated genes (either differentially expressed or differentially marked by H3K4me3) are linked to neuronal differentiation and maturation, as expected by the phenotype analysis. Finally, among the others, also the Polycomb axis is dysregulated, suggesting its possible involvement in the *Mll2*$^{-/-}$ defective transdifferentiation.

5. INTRODUCTION

5.1 Cell fate plasticity

5.1.1 Cell fate plasticity: a historical perspective

From an evolutionary perspective we could expect that the genome size increases along with the organism complexity. Paradoxically, higher order organisms do not necessarily have both a higher DNA content and a greater number of genes. This is the so-called C-value enigma or, indeed, paradox (C.A. Thomas Jr, 1971). However, the layers of epigenetic control of gene expression and gene accessibility co-evolved with genome size[1-3]. This, on one side, could partially reconcile the C-value enigma. Moreover, the epigenetic configuration, since the cells of an organism have the same DNA content, defines the identity/the lineage commitment of the considered cell.

Conrad Hal Waddington was the father of the term epigenetics, defined as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being"[4]. He also stated, regarding "the "*whole complex of developmental processes*" that connects genotype and phenotype that "*It is convenient to have a name for this complex; 'epigenotype' seems suitable*"[5,6].

With the growth of the knowledge in this topic, epigenetics, nowadays, comprehends all the inheritable modifications at the DNA level, but not to its sequence, and at the histone proteins associated to DNA.

Conrad Hal Waddington also fathered the seminal metaphor of the epigenetic landscape. He depicted the pluripotent cell as a marble rolling down the epigenetic landscape (Figure 1). Every valley and hill that the marble encounters represents a different lineage choice, respectively more or less energetically favourable. Progressively the marble becomes more and more committed until it reaches its final valley: the terminally differentiated state. We could envisage the cell/marble modifying its epigenome along the descent and therefore

we could consider that each step of differentiation is defined by a different epigenome configuration (Figure 1).

Due to the unfavourable energetic condition, the marble was considered unable to get back on top of the epigenetic landscape (*i.e.*, back to pluripotency) and to cross the epigenetic hills that separate the diverse lineage fates (*i.e.*, change commitment). Although the metaphor of differentiation remains valid, the notion of irreversibility of the fate choice has by now been surpassed.

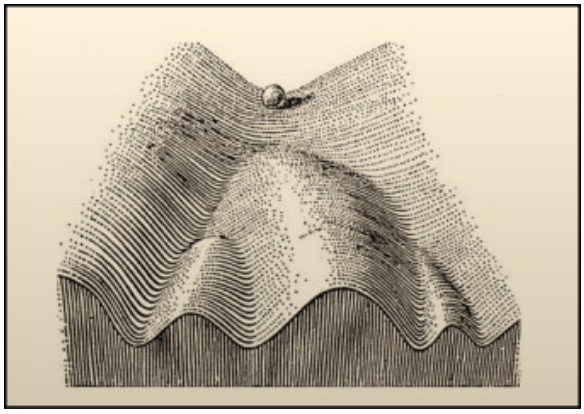


Figure 1: The Waddington landscape model

In the Waddington model the pluripotent cell is depicted as a marble rolling down the epigenetic landscape until its final differentiated state is reached (From Waddington, 1942).

5.1.2 Epigenetics: crossing the barriers

In 1950s, Briggs, King and Gurdon, taking advantage of the Somatic Cell Nuclear Transfer (SCNT) technique, demonstrated that some components of the cytoplasm of an enucleated frog oocyte, at that time unidentified, were able to "reprogram" the transferred nucleus of somatic, terminally differentiated *Xenopus laevis* cells, back to pluripotency, allowing the formation of new tadpoles[7,8]. SCNT was demonstrated effective also in different organisms, such as the *Mus musculus* (Figure 2d). This discovery set the foundations for the seminal study of Yamanaka and Takahashi of reprogramming of differentiated cells

into induced Pluripotent Stem Cells (iPSCs) in 2006 (Figure 2b)[9]. Yamanaka and Takahashi showed that through the ectopic expression of Oct3/4, Sox2, Klf4 and c-Myc, four transcription factors (TFs) important for the maintenance of pluripotency and for the proliferation of Embryonic Stem Cells (ESCs), they were able to reprogram mouse fibroblasts, both embryonic and adult, into pluripotent cells. iPSCs expressed pluripotency markers, formed teratomas if transplanted in nude mice and contributed to mouse development when they were introduced into the blastocyst, three fundamental features of ESCs and hallmarks of pluripotency[9]. The reprogramming protocol has been adapted also to human cells[10,11] and to cells of different lineages[12-15], revolutionizing the stem cell field. The discoveries of Gurdon and Yamanaka showed that lineage-specific TFs are able to reprogram and reset the epigenome of a differentiated cell[16]: the epigenome, therefore is not fixed in a static and incontrovertible configuration (Figure 2b).

Another proof that lineage specification of a cell is not irreversible came from cell fusion studies. Heterokaryons are non-dividing multinucleated cells that originate by fusion of cells with different identities. In the heterokaryons formed by mouse myotubes and differentiated human cells, human muscle cell markers are up-regulated in the latter[17] (Figure 2e). This evidence shows that the identity of differentiated cells, and therefore their epigenetic configuration are actively maintained by instructive factors, able to orient the fusion partner to their lineage. These TFs are called also “terminal selectors genes”, which, by controlling a plethora of effector genes, directly specify the cell identity during development and preserve it in mature cells.

In 1979 Taylor and Jones demonstrated that treating mouse fibroblasts with the demethylating agent 5-azacytidine (5-azaC) they formed myotubes[18]. In the following years the helix-loop-helix TF Myogenic Differentiation 1 (MYOD1) was identified as the main driver of myogenic conversion[19]. The direct passage, without passing through the pluripotent state, between differentiated cells of diverse lineages is called transdifferentiation or cell conversion (Figure 2c and 2f). Nowadays there are many other

examples of experimental transdifferentiation, also between cells that derive, unlike fibroblasts and muscle cells, from different germ layers and that therefore do no share the majority of the epigenetic marks.

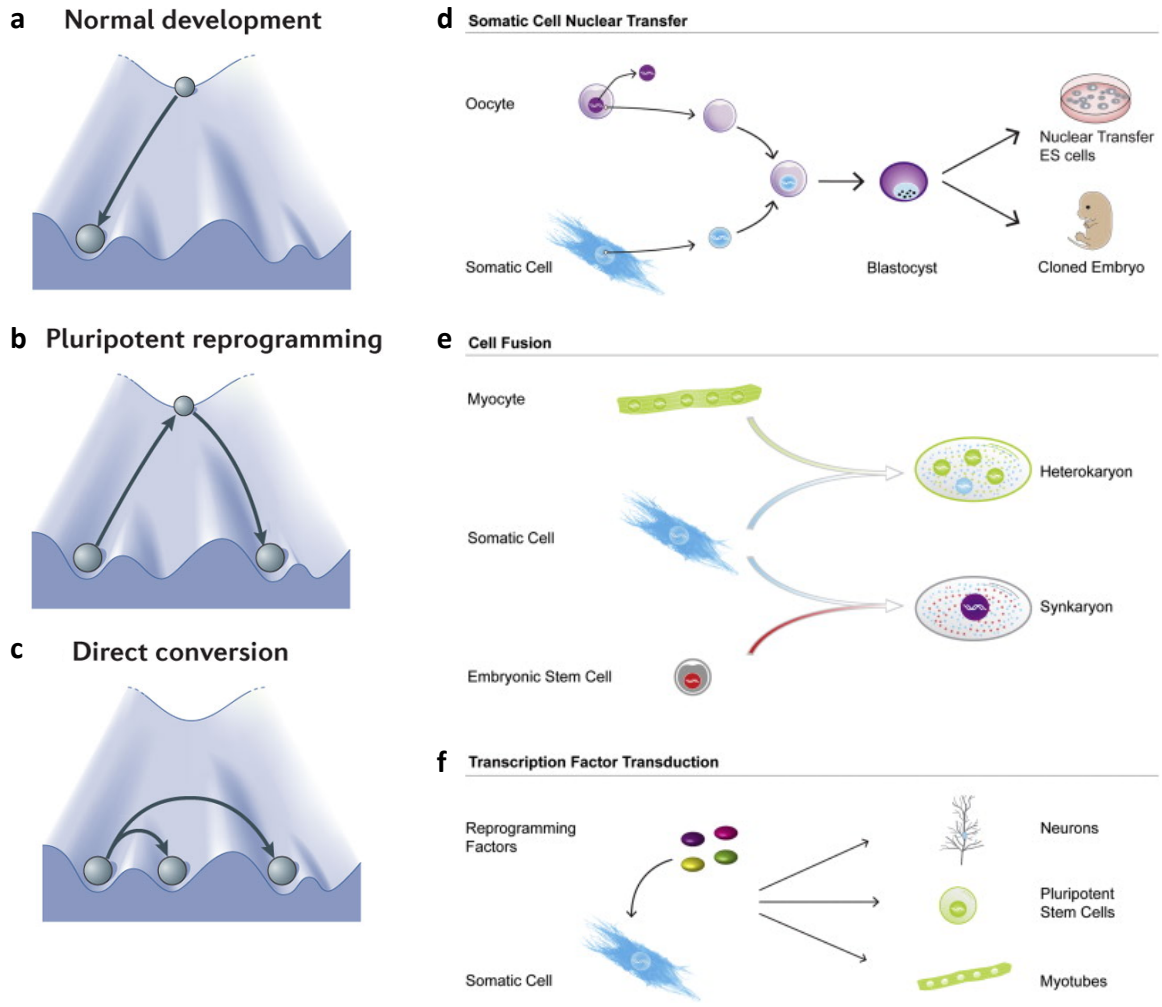


Figure 2: The exceptions to the Waddington's model

a. Cell development path according to Waddington's model. **b.** Reprogramming to pluripotency. Unexpectedly the marble, from the differentiated state, can also roll back to the top (the pluripotent state). **c.** Transdifferentiation**.** Cells can convert into other differentiated cells, both deriving from the same progenitor (short arrow) and with different embryological origins (long arrow), without passing through the pluripotent state. **d.** Schematic of somatic cell nuclear transfer. Through the transfer of a somatic nucleus into an enucleated oocyte, the nucleus-receiving cell is reprogrammed to pluripotency, from which also an entire organism can be derived. **e.** Schematic of Cell Fusion

between different cell types, which leads to the formation of heterokaryons. **f.** Schematic of transcription factors-mediated reprogramming towards different cell fates.
(Adapted from Ladewig et al., Nature Reviews Molecular Cell Biology 2013 and Vierbuchen and Wernig, Molecular Cell 2012)

These breakthrough studies amplified the vision of Waddington. Cell identity is dynamically preserved by lineage instructor factors that, on one side, keep lineage specific cascades active (*i.e.*, SCNT and heterokaryons studies). On the other side they epigenetically maintain non-lineage specific signalling switched off (*i.e.*, 5-azaC experiments). This equilibrium is, however, perturbable, precisely thanks to the ectopic expression of specific TFs for different lineages. Cells are therefore more plastic than previously thought. Although the hierarchical epigenetic landscape is still perfect to depict cell differentiation, it does not apply to the description of reprogramming and transdifferentiation. One alternative representation is the "epigenetic disc" proposed by Brüstle and colleagues (Figure 3)[20]. Since it is now almost equally simple to reprogram cells to pluripotency and to differentiate or to transdifferentiate them, the epigenetic disc is a model where all the cell identities are holes (the unstable pluripotent state is an open hutch) lining the border of a disc. In this representation, a slight tilt of the disc (the expression of specific TFs) is sufficient to make the marble (*i.e.*, the cell) roll towards the desired cell lineage, abolishing the concept of hierarchies.

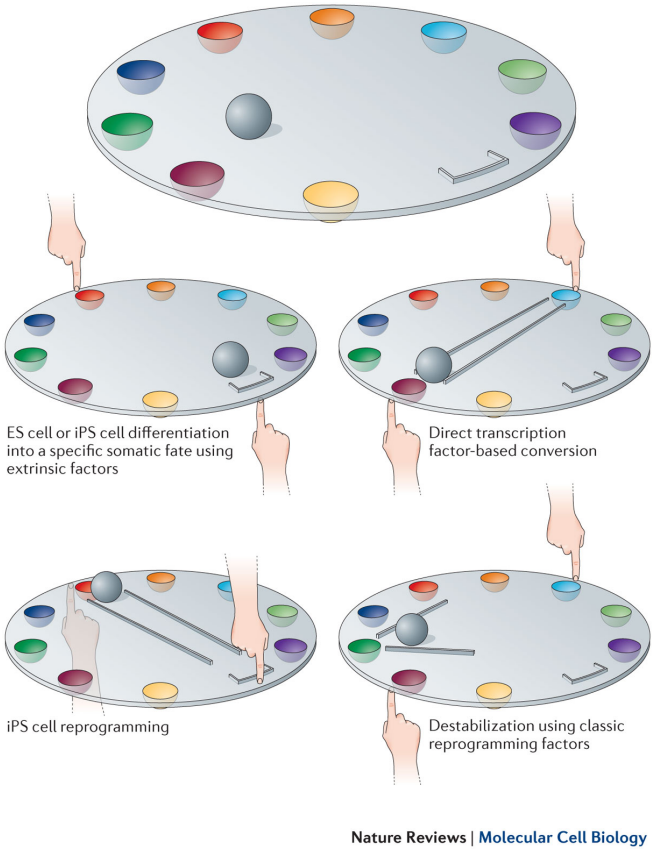


Figure 3: The epigenetic disc model

The cell is depicted as a marble, as in Waddington's model, but all the cell identities are represented as holes along the margins of a disc, thus eliminating the hierarchy embedded in Waddington's model. By tilting the disc (*i.e.,* the transduction of specific TFs), the marble can reach other differentiated states including the pluripotent one. Since pluripotency is a metastable condition, it is represented as an open hutch rather than a hole.

(From Ladewig et al., Nature Reviews Molecular Cell Biology 2013)

5.1.3 Stability of a differentiated cell

The identity of a cell is differently maintained depending on the cell type, through networks that are more or less stable.

In cells that are highly plastic (*e.g.,* B cells) this network could be as simple as a cross-inhibitory regulation between two instructor TFs: the active one directs and maintains the current lineage, inhibiting the other that would be responsible for the alternative

specification (Figure 4a). In this case the ablation of the expressed TF would lead to the activation of the alternative pathway, and consequently to lineage switching. One example of this univocal type of cross-regulation is the misexpression of the *CCAAT/enhancer binding protein* α (*C/EBP*α) and *C/EBP*β in B cells. These two TFs inhibit *Pax5,* the instructive factor for B cell commitment, leading to macrophages transdifferentiation[21]. Another case is the ablation of *forkhead box L2* (*Foxl2*) gene, a TF fundamental for female gonad development, in granulosa and theca cells, which induces their transdifferentiation into Sertoli and Leydig cells[22]. Thus on one side, in the hierarchical tree of differentiation, the choice between two closely related cell lineages branches is exerted by the mutually exclusive expression of driver TFs. On the other side, since these cell types are developmentally proximal and in most of the cases deriving by the same progenitor, their epigenomes present only minor differences. Therefore the resetting of the lineage-specific epigenetic marks is not massive as in the case of transdifferentiation between cells deriving by different germ layers.

In the case of stable/less plastic cells, instead (*e.g.,* neurons) the sole ablation of the lineage specific TF or misexpression of the alternative TF are not sufficient to drive transdifferentiation in the mutually exclusive cell-fate, because of a higher internal regulation of the differentiation/maintenance pathway and a stronger blockade of the alternative pathway by epigenetic modifications (Figure 4b). One example is *Nurr1*, the fundamental TF for dopaminergic differentiation, which leads only to the down-regulation of some dopaminergic markers upon ablation in adult mice. On the contrary if it is embryonically deleted, this results in the totally absence of dopaminergic neurons[23]. The tighter regulation in the adult implicates a fine preservation of the integrity of the tissue.

A third case comprises stable cells in which the two driver TFs activate pathways that have both a high internal regulation and that do not intersect (Figure 4c). This is also the case of cells deriving from different germ layers.

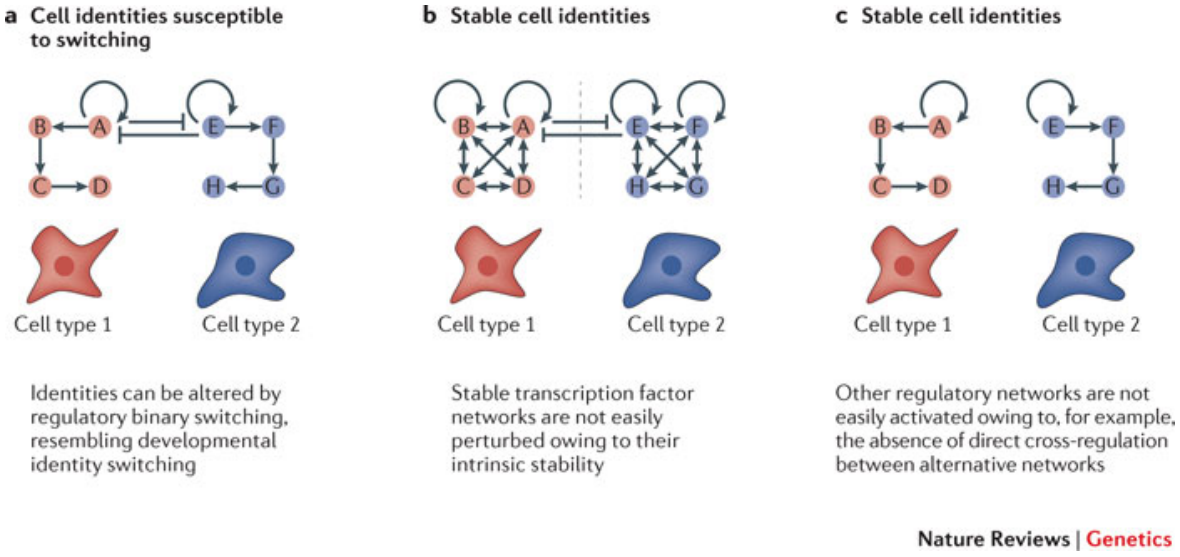


Figure 4: The Networks preserving cell identity

a. Cell identity is controlled only by one TF (A), which inhibits the mutually exclusive activity of the TF (B) governing the alternative cell fate acquisition. **b.** The deletion of the TF (A or B), that maintains the cell identity, is not sufficient to induce the alternative cell fate, because of internal cross regulatory mechanisms. **c.** The deletion of the TF (A or B), that maintains the cell identity, does not lead to the acquisition of the alternative cell fate, because the networks depending from A and B do not intersect.

(From Holmberg and Perlmann, Nature Reviews Genetics 2012)

5.2 Epigenetics and cell fate acquisition

5.2.1 Basic notions of epigenetics

Eukaryotic DNA is wrapped around nucleosomes, multiprotein complexes composed by eight histone subunits: two histones 2A (H2A), two H2B, two H3 and two H4. The complexes of DNA with nucleosomes and other DNA-bound proteins constitute chromatin. When chromatin is in a closed conformation and is not accessible for instance to TFs binding, it is called heterochromatin, whereas in an open conformation is called

euchromatin. As written above the term epigenetics comprises all the inheritable modifications to DNA, but not in its sequence, and to the nucleosomes.

Up to date several post translational modifications of histone tails have been identified, including: phosphorylation, acetylation, ubiquitination, and sumoylation. These "histone marks" constitute a chromatin signature drawn and recognized by specific epigenetic remodellers and adaptors whose binding influences the chromatin structure itself, gene transcription and DNA replication. Among the major epigenetic modifiers, Polycomb group (PcG) and Trithorax group (TrxG) protein complexes have been characterized.

These two complexes were discovered as the causative agents, when mutated, of the transformation of one *Drosophila melanogaster* (*D. melanogaster*) body segment into another (*i.e*., homeotic transformation)[24-26]. During development the identity of each *D. melanogaster* body segment is specified by the spatial and temporal expression pattern of *Homeobox* genes (*HOX* genes), determined by maternal and zygotic TFs that, however, are present only during the early stages. At later time points only PcG and TrxG preserve the memory of *HOX* genes transcriptional states, by depositing specific histone marks. Hence their perturbation leads to homeotic transformations.

PcG proteins are mainly transcriptional repressors responsible for the trimethylation of the lysine 27 of histone H3 (H3K27me3), in the case of the Polycomb repressive complex 2 (PRC2) and for the ubiquitination of the lysine 119 of histone H2A (H2AK119ub,) in the case of Polycomb repressive complex 1 (PRC1). On the other side, TrxG proteins deposit an activating mark, methylating the lysine 4 of the histone H3 (H3K4me1, H3K4me2, H3K4me3).

The epigenome and, consequently, the transcriptome of a cell define its identity. Indeed the epigenetic configurations at enhancers (*i.e.,* genomic sequences, proximal or distal with respect to the transcription start site (TSS), recognized by specific TFs, which positively or negatively modulate gene transcription) and promoters, highly influenced by TrxG and PcG proteins, regulate gene expression[27,28]. In particular, H3K4me3 has been found to be

associated with active gene promoters, while H3K27me3 with repressed ones. H3K4me3 and H3K27me3 are co-present in promoters of developmental regulator genes, allowing low mRNA transcription. This specific epigenetic configuration, called bivalent chromatin domain[29], has been discovered in ESCs and confers plasticity to cells: during lineage specification, if the trimethylation of H3K27 is removed, the gene is activated. On the contrary, if H3K4me3 is erased the gene is permanently silenced.

Also enhancers can be classified on the basis of the different TrxG and PcG epigenetic marks distribution. In particular the presence of only H3K4me1 characterizes the enhancers primed for activation and the co-presence of H3K4me1 and H3K27me3 defines the poised enhancers. In the latter configuration if the H3K27me3 mark is replaced by H3K27 acetylation the enhancer is considered active, whereas if only the H3K27me3 is present the enhancer is silenced. TrxG and PcG are therefore antagonistic epigenetic remodellers, finely tuned during development and differentiation.

5.2.2 Polycomb group Proteins

The catalytic components of the PRC1 are RING1A and RING1B, the E3 ubiquitin ligases responsible for the deposition of the H2AK119ub, an epigenetic mark linked to chromatin compaction and transcriptional repression. RING1A and RING1B were found to be associated with different components, forming the canonical and noncanonical PRC1 complexes. In the canonical complex they associate with CBX, which allows the recruitment of PRC1 to specific targets marked by H3K27me3, making the chromatin binding PRC2-dependent[30] and with BMI1, which is fundamental for the complex formation and activity[31]. On the contrary noncanonical PRC1 complexes contain RYBP and bind to chromatin in a PRC2-independent manner[32].

The catalytic subunits of PRC2 are instead EZH2 or its homolog EZH1 that through their SET domain deposit the H3K27me3 mark. EED and SUZ12, two components of PRC2 complexes, support the catalytic activity of EZH2/1[33]. The binding of PRC2 complexes is modulated by RBAP46/48, responsible of nucleosome recognition[34] and JARID2, that is also both a binding partner and a substrate for PRC2 itself[35,36]. In *D. melanogaster* specific genomic sequences responsible for PRC2 binding have been identified and named Polycomb responsive elements (PRE). To date an equivalent of PRE has not been identified in mammals. However, recently, unmethylated CpG islands associated with silenced enhancers have been suggested as PRC2 recruiting sites[37].

5.2.3 Trithorax group Proteins

The first H3K4 mono-, di- and trimethylation enzyme was reported in *Saccaromices cerevisiae (S. cerevisiae)*: the SET domain-containing 1 (Set1)[38]. To be functional Set1 has to be part of the COMplex of Proteins ASSociated with Set1 (COMPASS) and it is SET the domain that possesses the catalytic methylating activity. TrxG proteins, as mentioned before, were discovered in *D. melanogaster* thanks to homeotic transformation observed when they were mutated. In the fly, differently from *S. cerevisiae*, three H3K4 methylases are present: the homolog of Set1-COMPASS, the trithorax COMPASS-like and the trithorax-related COMPASS-like complex.

Similarly to PcG, TrxG complexes in *D. melanogaster* recognize a specific sequence in the genome: the TrxG responsive elements (TRE). As for PRE, up to date this type of sequence has not been found in mammals.

Because of the gene duplication that occurred during evolution, each complex acquired two paralogs in mammals[39-43]. The Set1-COMPASS can contain either SET1A (also known as KMT2 histone-lysine N-methyltransferase 2F (KMT2F)) or SET1B (also known

as KMT2G). The catalytic protein present in the COMPASS-like complex can be MLL1 (also called KMT2A) or MLL2 (also known as KMT2B, MLL4, TRX2 and WBP7), while in the case of trithorax-related complex the methylating protein can be MLL3 (also called KMT2C) or MLL4 (also known as KMT2D, MLL2 and ALR) (Figure 6). To avoid misunderstanding, from now on KMT2B will be referred as MLL2 and KMT2D as MLL4. The three different COMPASS sub-classifications have some shared and some specific components (Figure 6). WDR5, ASH2, RBBP5 and DPY30 (WARD) are the four proteins falling into the commonly found components of all KMT2 complexes that function as complex stabilizers and enhancers of KMT2 activity[44,45]. Set1-COMPASS complexes associate specifically with WDR82 and CFP1 (CXXC finger protein 1), the trithorax COMPASS-like complexes with MENIN and the trithorax-related COMPASS-like with UTX, PTIP, PA1 and NCOA6.

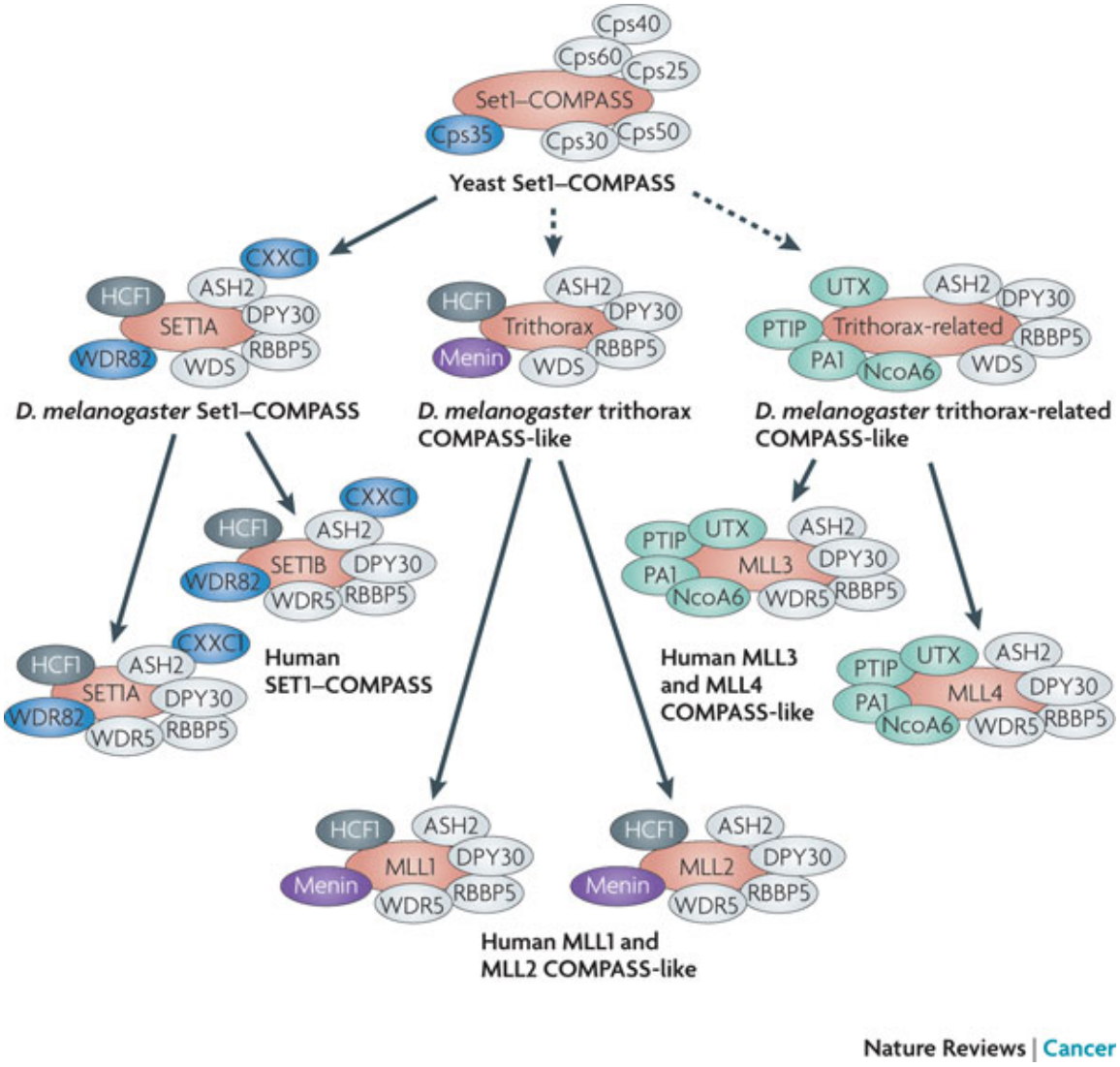


Figure 5: The COMPASS complexes in yeast, *D. melanogaster* and Human

In yeast only one COMPASS complex is present, while *D. melanogaster* has three COMPASS complexes. The components with the methylase activity in *Drosophila* are: SET1A, Trithorax and Trithorax-related. During the evolution each of them acquired two homologs. Grey subunits are shared among all complexes. In green the specific subunits for the trithorax-related COMPASS-like complexes, in purple the specific subunit of trithorax COMPASS-like complexes, while in blue the subunits of SET1-COMPASS complexes. (From Mohan et al., Nature Reviews Cancer 2010)

5.2.4 Functions and localization of Trithorax group Proteins

SET1A and SET1B are the two trithorax members responsible for the majority of H3K4 methylation[46]. In particular, as their ancestor Set1 in *Caenorhabditis elegans,* SET1A and SET1B can catalyse mono-, di-, and trimethylation (H3K4me1, H3K4me2 and H3K4me3)[47]. As mentioned before their specific components are WDR82 and CFP1. WDR82 is fundamental for proper H3K4 methylation upon H2B monoubiquitination[46] and it associates with the phosphorylated (on serine-5 of the C-terminal domain) Polymerase II allowing the co-transcriptionally deposition of H3K4me3[48]. On the contrary, CFP1 guides SET1A and SET1B binding to unmethylated CpG islands. Hence these two methylases preferentially associate with promoters and TSS (Figure 7). In the WDR82-dependent binding they are recruited after transcription has initiated[49] as their counterpart Set1 in *C. elegans*[48], whose deletion, indeed, leads to modest variation in the yeast transcriptome[50]. In the alternative case, their transcription-independent recruitment could most likely influence transcription.

MLL1 and MLL2 are able to mono-, di-, and trimethylate (even if at a lower rate) H3K4 as SET1A and SET1B[47], but they have a restricted set of targets (*i.e.*, MLL1 is responsible for the trimethylation of less than 5% H3K4 trimethylated promoter genes[51]). Both MLL1 and MLL2 can bind to promoters-associated unmethylated CpG islands through the CXXC domain, a type of zinc finger, whereas they can also be recruited to enhancers through the interaction with specific TFs (Figure 7)[52,53,54]. Moreover MENIN, their exclusive component, is the subunit that is mainly responsible for the interaction between TFs and trithorax COMPASS-like complexes. In particular MENIN is a tumor suppressor gene, which favours MLL1 and MLL2 localization at their most known targets, such as *HOX* genes[43,55-57]. Indeed mice knock-out for the *Mll1* SET domain are viable, but show an impairment in *Hox* expression[58]. Hence, differently from SET1A and SET1B, MLL1 and

MLL2 are recruited exclusively before transcription has initiated and they can actively influence gene expression.

MLL1 and MLL2 are cleaved by taspase 1, which creates an N-terminal fragment containing the phenylalanine and tyrosine-rich region FYRN (phenylalanine and tyrosine-rich N-terminal) and a C-terminal fragment containing both FYRC (phenylalanine and tyrosine-rich C-terminal) and the highly conserved SET domain (Figure 8)[59,60]. FYRN and FYRC heterodimerize to generate a functional complex. Indeed *taspase* deficient mice present homeotic transformation[61].

MLL2 augments *in vitro* chromatin transcription, only if S-adenosyl-methionine (SAM), the co-substrate of methylation, is present, directly linking the enhancement of transcription to H3K4 methylation[62]. Moreover the MLL2-deposited H3K4me3 prevents H3K27me3 deposition[63,64] and CpG methylation, playing an anti-silencing role[65].

MLL3 and MLL4, instead, are mostly monomethylases[47] that preferentially bind to enhancers (Figure 7)[66,67]. PTIP and NCOA6, the MLL3 and MLL4 specific subunits, interact with the PAX family of TFs and nuclear receptors[42,68,69], while CFP1 (CXXC finger protein 1) recruits Set1-COMPASS complexes to unmethylated CpG dinucleotides[70]. Since UTX, one of the demethylases of H3K27me3, is a component of the trithorax-related COMPASS-like complexes[68], the monomethylation of H3K4 occurs concomitantly with the removal of H3K27 methylation[71]. Afterwards P300/CBP is recruited to acetylate H3K27. The co-occurrence of H3K4me3 and H3K27ac, as aforementioned, is a signature of an active enhancer.

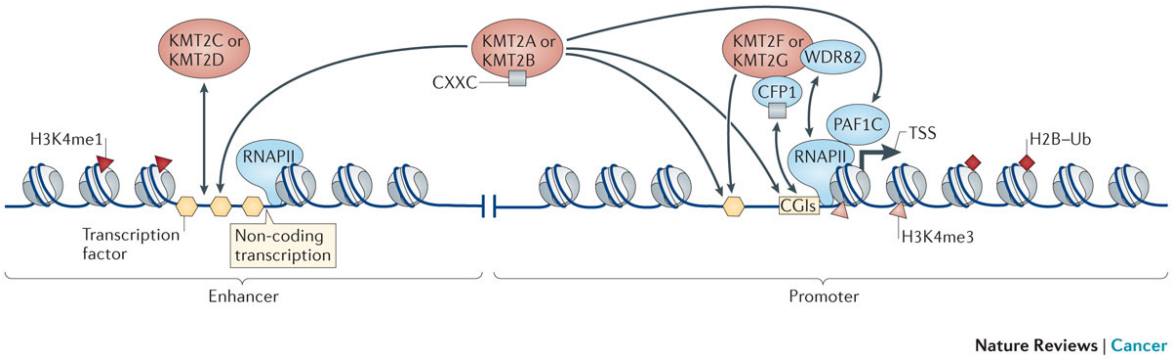


Figure 6: Functions and localization of KMT2 proteins

Schematic representation of KMT2 binding. KMT2C and KMT2D (MLL3 and MLL4), mostly bind at enhancer (through the interaction with TFs) and are responsible for H3K4me1. KMT2F and KMT2G (SET1A and SET1B) are the main responsible for H3K4 trimethylation at promoters. They can interact with TFs, with RNApolII and with CpG islands through the CFP1 subunit. KMT2A and KMT2B (MLL1 and MLL2), instead, can bind both to promoters (through their CXXC domain or through the interaction with TFs) and to enhancers (through the interaction with TFs) and deposit mostly the H3K4me3 mark.

(From Rao and Dou., Nature Reviews Cancer 2015)

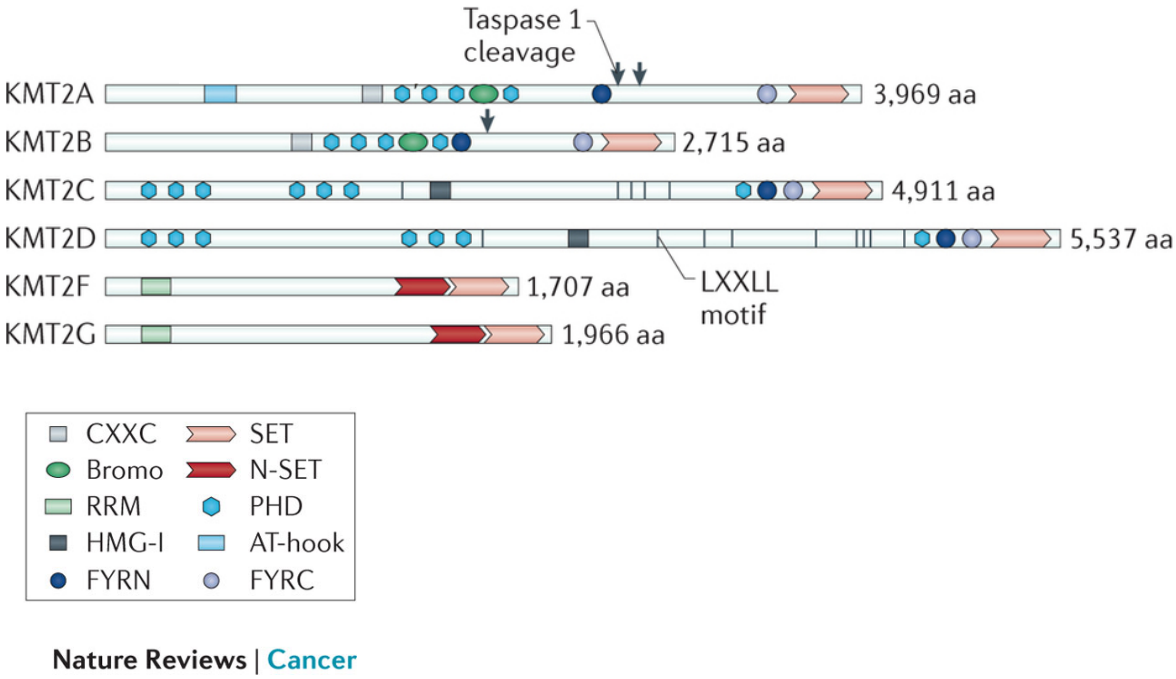


Figure 7: Protein domains of KTM2 enzymes

The specific and shared protein domains are depicted. In particular KMT2A and KMT2B (MLL1 and MLL2) possess a taspase cleavage site. Upon taspase cleavage the two protein fragments form a heterodimer through FYRN and FYRC, in order to be functional. In pink the catalytic SET domain.

(Adapted from Rao and Dou., Nature Reviews Cancer 2015)

5.2.5 Trithorax group Proteins and redundancy

In the case of TrxG genes, evolution corresponds to gene duplication. This on one side is synonymous with redundancy, but on the other side it increases the level of complexity and specification of H3K4 methylation deposition by each COMPASS member. Despite gene duplication, some specific and non-redundant functions still pertain to each trithorax member. Indeed the single knock-out mice for *Mll1* and *Mll2* are embryonic lethal, pointing to the absence of a complete compensation among the two paralogs[55,56]. Moreover, although the components composition and the protein domains are the same in each of the three sub-classifications (Figure 8), they display different subnuclear localizations[72,73]. In the current view each TrxG protein has a specific role, which is temporally and spatially modulated. *Mll2* represents a clear example, since it is fundamental during gastrulation until day E.8.5/9.5 of gestation[64]. Only the deletion in this time frame results in embryonic lethality. Then, MLL2 is again essential during oogenesis[74] and spermatogenesis[64].

Mll1, instead, is required later during development, especially for the hematopoietic lineages acquisition[75].

Moreover during *in vitro* retinoic acid (RA)-ESC differentiation, MLL1 and MLL2 regulate different *Hox* clusters[76].

Recently, MLL2 has been found to be required for H3K4me3 deposition on bivalent promoters in ESCs while MLL1 is redundant[76,77].

Hence, these two paralogs present non-redundant activities, regulated spatially (*i.e.*, they are the main H3K4 methylases of specific cell types and they have specific targets) and temporally (*i.e.*, they are fundamental in specific developmental time frames).

5.2.6 Trithorax group Proteins, cell cycle and cell mortality

TrxG proteins regulate and are regulated by the cell cycle. MLL1 presents a biphasic expression: it peaks during G1-S and G2-M transitions and it is degraded during late S and M phases by respectively the E3 ubiquitin ligase SCF bound to S-phase kinase-associated protein 2 (SCF^{SKP2}) and by the anaphase-promoting complex bound to its coactivator CDC20 (APC/C^{CDC20}) (Figure 9)[78].

MLL1, however, remains bound to DNA during replication and mitosis, such as its fly homolog Trx, preserving the transcription of cell-cycle genes and maintaining histone marks in the newly assembled nucleosomes[79,80] (Figure 9). Since an epigenetic modification is inheritable by definition, this observation makes the H3K4me3 a *bona fide* epigenetic mark. The turnover of other MLL proteins has not been investigated yet.

In the case of DNA damage, the phosphorylation on serine 516 of MLL1 reduces its degradation by SCF^{SKP2} [81]. Consequently H3K4me3 deposition at late replication origins increases, impeding the binding of CDC45, a component of the pre-replication complex, and therefore the replication of the damaged DNA (Figure 9).

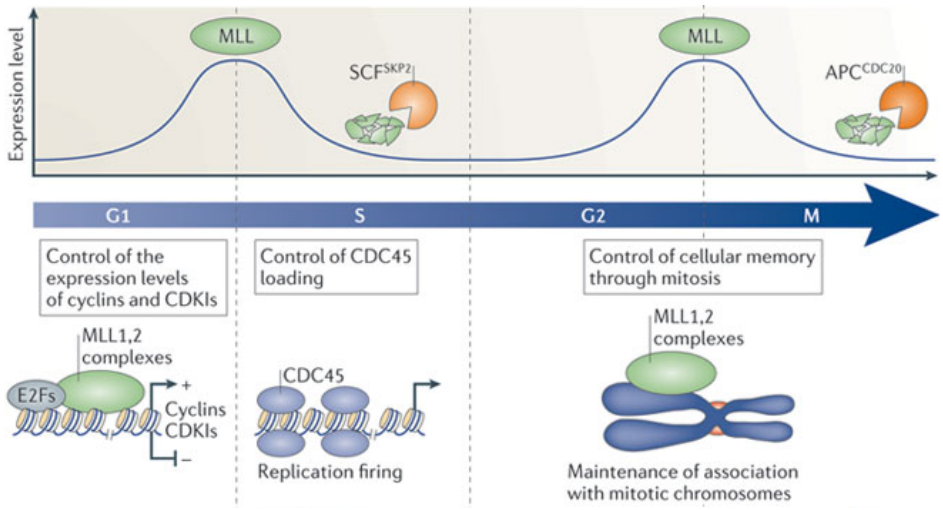


Figure 8: Cell cycle dependent degradation of MLL1

MLL1 is degraded during late S and M phases, while it peaks at the G1/S and G2/M transitions. MLL1 remains associated to chromosomes during mitosis and replication. In case of DNA damage, MLL1 phosphorylation of serine 516 reduces its degradation, increasing the H3K4me3 deposition at late origins and inhibiting the binding of CDC45. This impedes the replication of damaged DNA. Moreover both MLL1 and MLL2 interact with E2F proteins regulating the cell cycle.
(Adapted from Schuettengruber et al., Nature Reviews Molecular Cell Biology 2011)

On the other side, in the presence of DNA damage, one of the targets of MENIN-containing complexes is *p21*, the primary downstream effector of *p53,* and *p16*[82].

In the absence of damage, cell cycle regulator genes, such as *p27* and *p18* are transcriptionally switched on by MLL1 and MLL2 COMPASS-like complexes in fibroblasts[83]. Furthermore these complexes modulate also the expression of cyclin dependent kinase inhibitors (CDKI)[82] and of cyclins[61].

Moreover multiple cell types *Mll2*$^{-/-}$ present a higher cell death rate than the controls. *Mll2*$^{-/-}$ embryos die before E.10.5 due to a growth retardation that culminates in a high cell death[56]. *Mll2*$^{-/-}$ ESCs show a higher apoptotic rate than wild-type (wt) controls, likely due to *Bcl2* down-regulation, while cell cycle length and cell cycle phases distribution are not affected[84]. If *Mll2* is deleted in 2 months-old male mice, this results in infertility associated with an increase apoptosis of spermatogonia[64]. Finally if *Mll2* is conditionally knocked-out in oocytes, pro-apoptotic genes, such as *Bax*, *Casp6* and *Setd7*, which stabilizes p53, are up-regulated[74].

MLL1 and MLL2 associate also with cell-cycle regulatory TFs as E2F to regulate G1-phase cell cycle genes[85]. Moreover the two proteins modulate different E2F genes suggesting again their non-redundancy.

Cellular senescence is associated with the activation of p16, whose H3K4 trimethylation increases during cellular ageing, counteracting the role of PcG proteins,[82] fundamental in young cells[55,86].

5.3 Transdifferentiation

5.3.1 Models of TF mediated cell reprogramming and conversion

TFs that drive cell fate change frequently are terminal selector genes in the resultant cell.

Four models have been proposed to describe how a TF can start the process of lineage conversion[87]. In the first case (Figure 5a), despite the fact that the promoters of the genes important for transdifferentiation or reprogramming are in an epigenetically closed configuration, their enhancers are accessible to the TF. The TF binding can induce chromatin remodelling at the promoter, and subsequently the activation of transdifferentiation/reprogramming effector genes. The majority of polycomb targets, the genes trimethylated at H3K27, have exactly this configuration[88].

Pioneer factors, instead, are able to bind also to closed chromatin, displacing the nucleosomes and recruiting other chromatin modifiers and TFs thus initiating the process of cell conversion (Figure 5b).

In other cases the TF can access its binding sites stochastically, for example during their temporary unwinding from the nucleosome or during nucleosome turnover (Figure 5c). During cell division, instead, there could be a time window in which the repressive marks are still not re-deposited and therefore the TF could bind (Figure 5d).

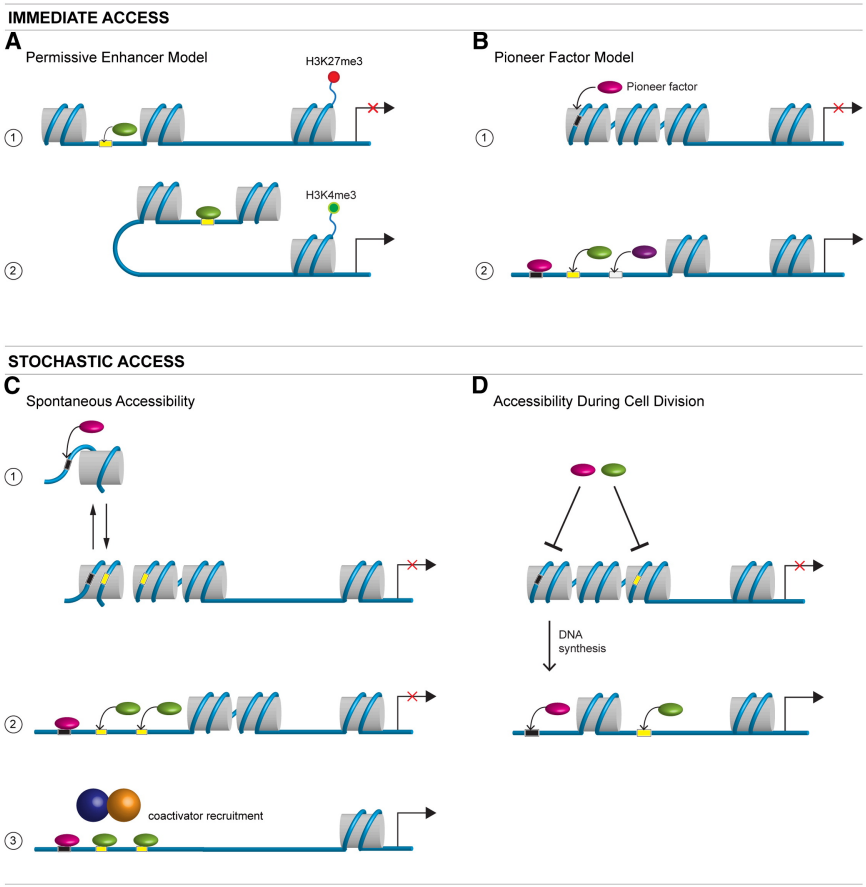


Figure 9: Models of reprogramming/transdifferentiation TF-binding

a. If the enhancer is in a permissive configuration the TF can bind, triggering chromatin remodelling at the promoter. **b.** Pioneer factors are able to bind to chromatin even if it is in a close conformation, leading to its subsequent remodelling. **c.** In the stochastic model TFs can bind to chromatin for example when, by chance, a specific sequence becomes accessible upon temporary nucleosome shifting. **d.** During cell division nucleosome are displaced and the new ones are initially not modified, giving to the TF a time window opportunity for binding.

(From Vierbuchen and Wernig, Molecular Cell 2012)

5.3.2 Neuronal transdifferentiation

In 2002 the group of Götz reported one of the first cases, in literature, of neuronal transdifferentiation[89]. They showed that the forced expression of *Pax6* in astrocytes, one of

the main drivers of neurogenesis *in vivo*, orients them towards neuronal fate. Subsequently, other works demonstrated that astrocytes can be converted into neurons also through the expression of different genes, such as *Ngn2*[90]. In these protocols, however, both the original cell and the transdifferentiated resultant derive from ectoderm.

In 2010 Wernig and colleagues screened 19 genes for their ability to induce cell conversion of mouse embryonic fibroblasts (MEFs) into neurons, hence cells deriving from two completely different germ layers[91]. Criteria for including genes were their neuronal expression and their role in neural development and in the reprogramming to pluripotency.

They showed that the combination of three TFs, *Brn2*, *Myt1l* and *Ascl1* (BAM pool), was able to transdifferentiate MEFs (ca. 20% efficiency) and mouse tail tip fibroblasts (ca. 6% efficiency) into functional neurons (*i.e.*, able to fire action potential and to form synapses) in 14 days. They named the transdifferentiated cells induces Neuronal cells (iNs) to distinguish them from induced neurons (*i.e.*, brain-derived cells). Also the BAZ pool, in which *Mytl1* was replaced by *Zic1,* was efficient in MEFs transdifferentiation, but BAZ-iNs presented a less complex morphology.

When they functionally and morphologically analysed BAM-iNs, they observed that the majority of them were excitatory, expressing *vGLUT1* and *Tbr1*, markers of cortical excitatory neurons and that, if co-cultured with astrocytes, iNs released excitatory post-synaptic currents (EPSC). Only a small fraction of iNs, instead, expressed *GAD*, a GABAergic marker and no inhibitory post-synaptic currents (IPSC) were recorded.

Hence, Wernig and colleagues demonstrated that cells deriving from the mesoderm can transdifferentiate into cell deriving from the ectoderm, implying that the BAM pool can drive a conspicuous and efficient epigenetic remodelling.

Just a year later, the same group demonstrated that also hepatocytes, deriving from endoderm, upon BAM pool transduction, transdifferentiate into functional iNs with an efficiency similar to that of tail tip fibroblasts[92].

Once the new cell identity is well-established, a fully transdifferentiated cell should become independent from the driver instructor factor(s). BAM genes were administered through lentiviral vectors, whose expression was induced and sustained by doxycycline (dox) administered in the medium throughout the protocol. They indeed showed that 5 days of dox administration was enough to generate both hepatocytes- and MEF-iNs, even if at a low percentage[92]. Removing dox 11 days after transduction did not change transdifferentiation efficiency at 22 days. This is likely due to the up-regulation of the endogenous levels of *Ascl1*, *Brn2* and *Myt1l* genes. At day 11-12 *Ascl1* reaches a plateau, while the levels of *Brn2* and *Myt1l* increase continuously until day 24.

5.3.3 Human induced neuronal cells

When Wernig and colleagues used the BAM pool to transdifferentiate human fetal fibroblasts into neurons, they observed β-III tubulin-positive ($Tuj1^{+}$) cells after 7-10 days, but these human iNs appeared functionally immature[93]. Adding *NEUROD1* to the pool, they finally succeeded in the generation of firing action potential iNs, that, like mouse iNs, were mostly excitatory. However the efficiency of conversion remained low (ca. 2-4%, 10 time less than MEFs) and transdifferentiation lasted around 35 days. Similar results were obtained using the five factors effective in mouse (5F pool: *Brn2*, *Olig2*, *Ascl1*, *Myt1l, Zic1*)[91,94].

Alternatively miR-9/9* and miR-124 alone, specifically expressed in postmitotic neurons, were used as factors to transdifferentiate human fibroblasts, but the resulting iNs were functionally immature[95]. If *NEUROD2*, *ASCL1* and *MYT1L* were added to the miRNA, they acted synergistically: transdifferentiation efficiency increased and iNs were functional. Moreover also the combination of *miR-124*, *BRN2* and *MYT1L* led to neuronal transdifferentiation, even if iNs generated presented less competent synapses[96].

In all the aforementioned reports, however, iNs were composed of a mixed population.

5.3.4 Disease modelling using iNs

In the case of 5F pool-iNs, the starting population consisted of fibroblasts derived from control and patient with Familial Alzheimer's Disease (FAD), carrying a *PSEN* mutation. FAD-iNs presented a disease-specific phenotype, showing amyloid precursor protein (APP) puncta in endosomes, rescued by the wt expression of PSEN1[94]. This was the first demonstration that, when affected cell types, such as neurons, are not easily accessible, transdifferentiation can be exploited with the purpose of disease modelling.

Another proof of principle that transdifferentiation can be exploitable to model disease came from Südhof's laboratory. They showed that the transdifferentiation of MEFs with the autism-related neuroligin-3 mutation is not affected, but that mutated iNs resemble *in vivo* affected hippocampal neurons. Indeed, mutated iNs present a reduction in the surface level of AMPA receptors with the consequent decrease of miniature EPSC both in amplitude and in frequency. This suggests that, even though the process of transdifferentiation does not follow the canonical physiological differentiation path, the resulting neurons are comparable to the primary ones[97].

To model disease and to study development, many successful attempts have been made to transdifferentiate fibroblasts into specific neuronal types involved in pathologies, such as dopaminergic neurons for Parkinson's disease. Pfisterer and colleagues demonstrated that the addition of *LMX1a* and *FOXA2* to BAM factors was effective for the transdifferentiation of human fetal fibroblasts into dopaminergic iNs (purity: 25% of all iNs generated)[98]. A higher purity was reached using the combination of *Ascl1*, *Nurr1* and *Lmx1a*, effective both for mouse and human fibroblasts (efficiency 18%; purity 85% and 60% of iNs were dopaminergic, in mouse and human respectively)[99]. However the transcriptome of dopaminergic iNs was different from *bona fide* dopaminergic neurons.

Alternatively *Ascl1*, *Pitx3*, *Lmx1a*, *Nurr1*, *Foxa2*, *En1*, *SHH* and *FGF8* were used to generate iNs, which in this case were also able to compensate for the absence of dopamine *in vivo*[100]. Despite the usage of eight TFs and their effectiveness when transplanted, the iNs gene expression profile was still different from the one of dopaminergic neurons.

Another fundamental type of neurons is constituted by motor neurons. Through the combination of BAM factors with *Lhx3*, *Hb9*, *Isl1* and *Ngn2* (and *NEUROD1* for human fibroblasts) mouse and human fibroblasts were transdifferentiated into motor neuron iNs[101]. iNs expressing the mutated gene SOD1, that causes amyotrophic lateral sclerosis (ALS), phenocopied the disease-survival defects of affected motor neurons.

These are only some examples of transdifferentiation towards specific neuronal types. A perfect combination of factors for the generation of iNs that fully recapitulate (by their function and their expression profiles) *in vivo* neurons has not yet been discovered. These studies can anyway give more insights on neuronal development and on the role of the TFs used for transdifferentiation. iNs, however, remain a promising, but still perfectible, tool to study disease-affected neurons. When the right "factors cocktail" will be found, hopefully in the near future, iNs could be exploited in regenerative medicine. With this purpose non-integrative transdifferentiation protocols, such as transfection of the factors with poly(amidoamine)[102] and adenoviral gene delivery[103], have been established. Since iNs are not expandable and the efficiency of transdifferentiation of human fibroblasts is low, protocols have been adapted to overcome these limitations. With the usage of small molecules inhibiting glycogen synthase kinase-3β and SMAD signaling, MEF-to-iNs conversion, driven by ASCL1 and NGN2, increases up to 200%[104]. Alternatively, transdifferentiation towards expandable induced neural precursor cells (iNPCs), prior to iNs, was achieved to overcome limitation on abundance[105].

5.3.5 How the BAM pool works

BAM factors act synergistically for the generation of iNs, inducing a conspicuous transcriptome remodelling: 2522 genes are differentially expressed between MEFs and iNs[106]. The resetting of the transcriptome should reflect a major restructuring of the epigenome; however, to date, no data are present on how the epigenetic landscape changes in iNs and on how BAM factors guide this remodelling.

If the MEFs-to-iNs transdifferentiation system were to be classified in the aforementioned four TFs-mediated reprogramming systems, it would fall in the "immediate access" category. ASCL1 is in fact a pioneer factor that can immediately bind chromatin leading, in a 5F pool transduced MEFs, to the first Tuj1^{+} (Neuronal- specific Class III β-tubulin) cells appearance already 3 days after infection[91]. In particular ASCL1 functions as a transcriptional activator, recognising a newly discovered epigenetic mark: the trivalent chromatin state, characterized by the co-presence of the H3K4me1, the H3 lysine 9 trimethylation (H3K9me3) and the H3K27 acetylation (H3K27ac)[106]. Accordingly, it has been demonstrated that cell types with an enrichment of the trivalent chromatin state have a higher transdifferentiation efficiency. Moreover cell proliferation is dispensable for the occurrence of direct cell conversion. One day after transgene induction only a minority of cells still proliferates[91] and blocking DNA replication does not affect the appearance of iNs[107]. Single cell RNA-sequencing (RNA-seq) of MEFs *Ascl1*-transduced, showed that upon transgene induction the switched-off or down-regulated genes are the ones linked to mitosis, confirming that cell division is not required for transdifferentiation[108]. In parallel ASCL1 up-regulates genes associated to cytoskeletal reorganization, synaptic activity and neurite extension, confirming its essential role throughout transdifferentiation[108].

Zfp238, a transcriptional repressor whose expression increases upon BAM pool dox-induction, was discovered as one of the main targets of ASCL1[106]. When *Zfp238* is expressed in combination with *Myt1l* and other neuronal TFs up-regulated during

transdifferentiation (*i.e.*, *Rfx1*, *Lmo2* and *Tcfl5*) it induces the formation of neuronal-like structures in MEFs, without the need of the expression of ASCL1. Hence *Zfp238* is one of the key effectors of ASCL1.

Interestingly, a portion of MEFs infected with *Ascl1* alone does not become iNs and activates the myocyte transcriptional program, while concomitantly down-regulating *Ascl1*[108]. This transcriptional deviation is however blocked in the presence of BRN2 and MYTL1L. Moreover, iNs at day 22, transduced only with *Ascl1,* fail to mature and mostly acquire the myogenic expression profile. Hence BRN2 and MYT1L present a major involvement in neuronal maturation in later stages of transdifferentiation[106,108]. It is ASCL1 that recruits BRN2 to some of its targets, otherwise not accessible, confirming its role as pioneer factor[106].

The definition of transdifferentiation excludes the passage through the pluripotent state. In the case of neuronal transdifferentiation this has been so far ruled out because of the rapidity of appearance of $Tuj1^+$ cells and of the absence of cells positive for pluripotency markers[91]. Recently it has been shown that during the first phases of MEFs conversion, their transcriptome clusters with the one of NPCs[108]. Then, as soon as iNs maturation progresses, their expression signature detaches from the one of NPCs and gets closer to the one of iNs[108]. However the two main NPC markers, *Sox2* and *Pax6*, are never induced during the process, implying that MEFs pass through an intermediate stage that only resembles NPCs.

All transduced MEFs are equally competent for the activation of the transdifferentiation program: the transcriptional analysis of starting MEFs does not show the presence of subgroups that could be less or more prone to cell conversion. Infected MEFs homogenously induce *Ascl1*-targets and silence MEF-genes[108]. Hence the 20% efficiency should be ascribed to the aforementioned silencing of *Ascl1* and activation of myogenic pathway, but the reason why this occurs only in some cells remains unknown.

5.3.6 PcG, TrxG and transdifferentiation

In certain species transdifferentiation occurs physiologically during both development and/or regeneration. For instance, the lens of the newt regenerates through the transdifferentiation of pigmented epithelial cells. As stated earlier, this entails a fine-tuned balance between plasticity and cell identity stability. To allow the cell to change fate, the activation of target cell genes and the repression of starting cell ones are essential. Hence the orchestrated activity of TrxG and PcG proteins is likely to be one of the main drivers.

In *D. melanogaster* each imaginal disc, a fly larval developmental structure, gives origin to a specific segment with a cuticular structure. The pre-determined imaginal disc clusters of cells can transdifferentiate during regeneration, upon the ectopic expression of morphogens and in the case of deregulation of PcG genes. Hence PcG proteins have a role in the suppression of the alternative cell fate acquisition, as demonstrated for the homeotic transformations. Fly cells, during regeneration, down-regulate PcG through the c-Jun N-terminal kinase (-) pathway and PcG mutant flies show a higher rate of transdifferentiation[109]. If transdifferentiation is artificially induced, members of the PcG and the TrxG families change their expression, and when some of them are mutated, transdifferentiation frequency is increased[110]. Hence, in *D. melanogaster*, the two families of histone methylases are the guardians of the cell fate identity that is not however, irreversibly crystallized.

In *C. elengans,* during development, a specific hindgut cell (*i.e.,* the Y cell) physiologically transdifferentiates to form a motor neuron. Mutations of the demethylase of H3K27, Jmjd3-3.1, and of the ortholog of Set1-COMPASS complex, Set-2, lead to transdifferentiation defects[111]. Moreover *C. elegans* germ cells knock-out for PRC2 show a higher rate of TF-mediated transdifferentiation toward both neurons and muscle cells[112]. Hence, also in *C. elegans*, the two antagonistic families play a major role during transdifferentiation and the preservation of cell identity.

Furthermore some of the TFs that drive, in mammals, *in vitro* transdifferentiation (*e.g., MyoD1, Cebpa, Pdx1,* etcetera*)* are repressed by H3K27me3 in the starting cell while they are active in the resultant cell, pointing to the role of PcG proteins in the silencing of terminal selector genes of other cell identities[113].

5.3.7 Histone methylation and corticogenesis

In the currently accepted model of corticogenesis, epigenetics plays a major role[114]. Radial glia (RG) cells, specific neural stem cells that reside in the ventricular zone, either directly or indirectly (through the generation of basal progenitors (BPs)) give rise to cortical neurons and only later to glia. The six cortical layers are generated in a temporally programmed fashion as well. It has been demonstrated both *in vitro* and *in vivo* that at each time point neural progenitors are "committed" to generate defined neurons in a largely cell-autonomous manner[115-117]. The sequential modification of the epigenetic profile of the RG cells could be responsible for the acquisition of their different differentiation potential. One of the most attractive hypotheses is the presence of a repressor, which is down-regulated during RG cell cycles, leading to the derepression of targets relevant for the differentiation into that time-specific cell type. *Ezh2* shows indeed this expression pattern: it is down-regulated both along the 11 cell cycles of RG cells and during neurons generation[118]. However, the ablation of this enzyme through different methods *in vivo* led to interestingly opposite results[118,119], further underscoring the importance of precise timing in the execution of its molecular function. Moreover the highly dynamic H3K27 methylation mark during neuronal differentiation *in vitro*[120] and the aforementioned theory of progressive Polycomb target derepression in the RG suggest a role for one of the main H3K27 demethylating enzymes, JMJD3. Our group has already demonstrated that this demethylase is fundamental for neural lineage commitment[121] and is

also currently unravelling its role during corticogenesis *in vivo* with the generation of constitutive and conditional knock-out mice[122].

Conversely, the role of H3K4me3 is less well understood, but it has been still associated with neurogenesis. It has been recently shown in *C. elegans* that the Set1 complex participates to the process of dedifferentiation of the hindgut cells that will subsequently redifferentiate in motor neurons[111].

MLL1 and MLL2 have been shown to be essential in neural commitment. Indeed the differentiation of *Mll1*$^{-/-}$ Subventrincular Zone Neural stem cells (SVZ NSCs) into neurons is impaired, while their proliferation rate is not altered[123]. This deficit is not attributable to a reduced trimethylation at MLL1 targets because the level of H3K4me3 is not altered. This implies a possible compensation operated by the homolog MLL2. The authors described the absence of the erasure of H3K27me3 at important neuronal targets as the cause of the defect in neuronal differentiation. Furthermore it has been recently shown that MLL1 is essential for neural development in *Danio rerio*[124]. In particular the absence of *Mll1* in zebrafish leads to a reduction in the proliferation of neural progenitor cells and to an anticipated neuronal differentiation. On the other side *Mll2*$^{-/-}$ ESCs have a severe delay in ectodermal *in vitro* differentiation[84], but they are able to differentiate and they are positive for mature neuronal markers upon retinoic acid treatment. However, the expression of maturation genes in these cells is highly affected. It has been shown recently that in ESCs MLL2 is responsible of H3K4 trimethylation at bivalent promoters[67,76]. During neuronal differentiation this defect is only partially overcome, also through the compensation of MLL1, but *Mll2*$^{-/-}$ ESCs can anyway respond to retinoic acid treatment[67,76]. Moreover the *Mll2* deletion in the excitatory neurons of the forebrain impairs both short and long term memory[125]. 161 genes have been found to be differentially expressed in *Mll2*$^{-/-}$ dorsal dentate gyrus with respect to controls by microarray analysis. The majority of them are down-regulated and linked to memory formation and neuronal plasticity. The levels of H3K4me3 and H3K4me2 it these genes

are reduced, as assessed through Chromatin Immunoprecipitation coupled to detection by quantitative real-time PCR (ChIP-qPCR), further defining them as MLL2 targets and excluding the compensation by other TrxG proteins. However, surprisingly, *Mll2*$^{-/-}$ brains and neurons do not present an altered morphology.

Hence, despite the fact that *Mll1* and *Mll2* arise from gene duplication[56,126,127] and share the classical Trithorax core components, plus specific components such as MENIN and HCF1O, they have precise and non-replaceable roles during neuronal differentiation.

5.4 Aim of the thesis

The aim of this study is the characterization of the role of MLL1 and MLL2 during BAM factors-driven MEFs-to-iNs transdifferentiation. This protocol offers three main advantages: i) it fosters the transition between cell lineages that derive from different germ layers, therefore requiring the transition across a "high epigenetic hill"; ii) it allows the complete erasure of the MEF epigenetic signature and the establishment of the neuronal one; iii) it is a remarkably efficient and fast process. Hence this system is appropriately suited for investigating how a massive TF-mediated epigenome resetting occurs at the molecular level. Moreover BAM factors are fundamental during physiologic neuronal differentiation, but it is unclear how they coordinate the massive epigenetic remodelling which occurs during direct cell conversion and what are the epigenetic remodellers with which they cooperate to achieve the complete epigenome restructuring. On this perspective we decided to unravel whether and how MLL1 and MLL2 are relevant during transdifferentiation, since they actively participate to both *in vivo* and *in vitro* neuronal differentiation. This will allow first to dissect a possible involvement of these two methylases during transdifferentiation process *per se*. Furthermore it will give more insights on their role during neuronal specification.

6. MATERIALS AND METHODS

6.1 Mouse Models

- Conditional knock-out for *Mll*: mouse models used in this study were conditional knock-out for *Mll1*[128], for *Mll2*[56] or for both. In these mice the exon 2 of *Mll1* or/and *Mll2* gene was flanked by two *LoxP* sites and the mice also harboured the *ROSA*26-*CreERT2* allele to enable tamoxifen-induced Cre-mediated recombination. They were crossed with mice presenting the *YFP* in the *ROSA*26 locus, downstream to a stop cassette between *LoxP* sites. Hence *Cre* was maintained in heterozygosis, reducing its toxicity. Control MEFs were derived from $Mll^{+/+}YFP^{+}Cre^{+}$ mice and in the case of *Mll1* also from $Mll1^{fl/+}$ $YFP^{+}Cre^{+}$ ones.

- NOD SCID IL2R$\gamma^{-/-}$ were bought from Charles River Laboratories and bred in house.

All mice were housed and bred in a specific pathogen free (SPF) animal house.

Mice were genotyped extracting DNA from tail biopsies. In particular tails, in 400µl of Lysis Buffer (100mM Tris-HCl pH 8.5, 5mM EDTA, 0.2% SDS, 200mM NaCl, 100 µg Proteinase K/ml), were digested at 55°C in agitation. Afterwards DNA was precipitated with isopropanol and washed in 70% ethanol. After ethanol evaporation, DNA was resuspended in ddH_2O. PCR were run on TAE 1% Agarose gel (2% in the case of *Mll2*).

Primers used are listed in Table 1.

	Forward	Reverse
YFP	TGCAGTGCTTCGCCCGCTACC	CCGTCGCCGATGGGGGTGTTC
Cre	GCCTGCATTACCGGTCGATGCAACGA	GTGGCAGATGGCGCGGCAACACCATT
Mll1	GAGGTAAGGAGAGTTTTTGCT	GGTAACACCTTAAAATGCCCT
Mll2	CGGAGGAAGAGAGCAGTGACG	GGACAGGAGTCACATCTGCTAGG
EGFP	GACAACCACTACCTGAGCAC	CAGGACCATGTGATCGCG

Table 1: Sequences of primers used for genotyping

6.2 Derivation of transdifferentiation starting cells

6.2.1 MEFs derivation

Embryos were harvested at E.13.5 of development. After removal of both the head and the liver, cells were enzymatically and mechanically dissociated. In particular, the embryo body was finely cut with scissor and afterwards 1ml of trypsin was added. Cells were incubated at 37° C, 3% O_2, 5% CO_2 for 30 minutes pipetting every 10 minutes. Dissociated MEFs were seeded in MEFs medium and cultured at 37 °C, 3% O_2 and 5% CO_2.

MEFs medium:

Dulbecco's Modified Eagle Medium without L-Glutamine (DMEM) (Lonza, catalog number BE12-614F)
10% Fetal Bovine Serum (FBS) (HyClone, catalog number SH30088.03)
1% Penicillin/Streptomycin (Life Technologies, catalog number 15140-122)
1% L-Glutamine 200mM (Life Technologies, catalog number 25030-024)
1% Sodium pyruvate (Lonza, catalog number BE13-115E)
1% Non-essential aminoacids (Lonza, catalog number BE13-114E)
0.2% β-mercaptoethanol 50mM (Life Technologies, catalog number 31350-010)

6.2.2 MEFs 4-OHT treatment

To allow exon 2 deletion MEFs, at passage 2 maximum, were treated for 5 days with 4-Hydroxytamoxifen (4-OHT), added to the MEFs medium in a final concentration of 1000ng/ml, and then left either 2 or 7 days in medium not supplemented with 4-OHT. MEFs *Mll*$^{+/+}$ 4-OHT-treated were used as control for MEFs *Mll*$^{flox/flox}$ 4-OHT treated, to rule out Cre and 4-OHT toxicity in all the experiments. However, for *Mll1* ScanR experiments, MEFs *Mll1*$^{flox/+}$ 4-OHT treated MEFs were used as control.

6.2.3 Assessment of exon2 deletion of both Mll1 and Mll2

To determine the entity of *Mll* exon 2 deletion, TaqMan assays were performed collecting the MEFs the day were plated for transdifferentiation. In particular, TaqMan was executed on 7900HT Fast Real-Time PCR system (Applied Biosystems), analyzing each sample in triplicate and using *Tert* as housekeeping. The Primers and reporters sequences are listed in Table 2.

Assay ID	Forward Primer Seq.	Reverse Primer Seq.	Reporter 1 Sequence
MLL2EX2	GGTCCCCTAAATCAGGAGTTTCAG	GACCGAAGCGCAGAGC	AAGATGTGGCCCCCAGTTC
MLL1EX2	GCAGTTCTTAGGTTTTGGCTCAGA	GCTTTATTGGCCATACCTGAAGGA	CTTCGCACTCTGACTTCTTCA

Table 2: Sequences of primers and probes used for Taqman assays

DNA was extracted through DNeasy Blood and Tissue kit according to manufacturer's instruction (Qiagen, catalog number 69504). Briefly, cells were lysed in the appropriate buffer (with proteinase K added) and loaded onto the DNeasy Mini spin column. Through centrifugation only DNA was bound to the column. After the washing of the remaining contaminants, DNA was eluted in AE buffer.

6.2.4 Western blot for MLL2

6.2.4.1 Protein extraction

Pellets, of at least 8 x 10^6 cells, were collected in the day MEFs were plated for transdifferentiation. After washing each 15-cm dish 2 times with cold PBS, 5 ml of cold PBS were added and MEFs were scraped and collected in a tube maintained on ice. Cell suspensions were centrifuged at 4 °C, 1100 rpm for 5 minutes. The supernatants were removed and the pellets were immediately put in dry ice and stored in –80 °C until proteins were extracted.

When proteins were extracted, pellets were thawed on ice. 4 volumes of RIPA buffer were added and tubes were put at 4 °C on a rotating wheel for 30 minutes. Extracts were centrifuged for 30 minutes at 13000 rpm at 4 °C and supernatants were transferred in new tubes and store at -80 °C.

RIPA buffer composition:

10 mM Tris-HCl pH 8
1% Triton X-100
0.1% SDS
0.1% Sodium Deoxycholate
140 mM NaCl
1 mM EDTA

Before usage the protease inhibitor cocktail (PIC) (Sigma Aldrich, catalog number P8340) and Phenylmethanesulfonyl fluoride (PMSF) (0.5 mM) (Sigma Aldrich, catalog number P7626) were added.

Proteins were quantified at the spectrophotometer (λ 595nm), through the Bradford protein assay as followed:

200µl of Bio-rad protein Assay Dye reagent Concentrate
800µl of ddH_2O
1µl of protein extract
using BSA (NEB) do derive a standard curve.

6.2.4.2 Western blot

Western blot was performed loading at least 70 µg of protein extract, diluted in RIPA buffer plus PIC and PMSF. The NuPAGE LDS Sample Buffer (4x) supplemented with Dithiothreitol (DTT) (50 mM) was used (ThermoFisher Scientific, catalog number NP0007). After denaturation at 95 °C for 3 minutes, proteins were loaded on NuPAGE Novex 3-8% Tris-Acetate Proteins Gels (ThermoFisher Scientific, catalog number

EA0378BOX). The Novex Sharp Pre-stained Protein Standard (ThermoFisher Scientific, catalog number LC5800) was used as marker. Run was executed in the NuPAGE Tris-Acetate SDS Running buffer (20X) (ThermoFisher Scientific, catalog number LA0041) at 4°C, initially at 80 V and, when the bands started to separate, at 100 V.

Wet transfer was performed at 4 °C, 30V with the Immobilon-P, 0.45 μm, PVDF membrane (Merck Millipore, catalog number IPVH00010) activated with methanol.

Transfer buffer was prepared as followed (700ml):

200 ml methanol

100 ml Towbin Buffer (TB) 10X (0.25 M Tris base, 1.9 M glycine)

dH_2O to 700 ml

The quality of transfer was checked through Ponceau staining (VWR) and after the washing with TBS-T, the membrane was incubated with the primary antibodies overnight at 4 °C. The step of blocking was performed only for the housekeeping (1h in 5% (w/v) skimmed milk powder in TBS-T at room temperature).

TBS-T composition (500ml):

50 ml TBS 10X (final concentration 25 mM Tris, 150 mM NaCl, 2 mM KCl)

500 μl Tween-20 (final concentration 0.1%)

dH_2O to 500 ml

The rabbit anti-MLL2 antibody was provided by Stewart Lab.[56]. The antibody was raised against amino acids 864-980. The 2 expected bands were of 284 (the full length) and 225 (the taspase product) kDa. The anti-MLL2 antibody was diluted 1:1000 in TBS-T 3% BSA.

The mouse anti-VINCULIN antibody (Sigma Aldrich, catalog number V9131) was diluted 1:400 in TBS-T with milk (expected band 116 kDa).

After 3 washes in TBS-T, the membrane was incubated with the secondary antibodies, diluted 1:10000 in 5% milk, 1h in agitation at room temperature.

Afterwards the membrane was washed 3 times with TBS-T and then the ECL Prime Western Blotting detection reagent was used (Sigma-Aldrich, catalog number GERPN2236).

Images were acquire at ChemiDoc with the Image Lab Software of Bio-Rad.

6.3 Transdifferentiation protocol

6.3.1 Vector production

BAM vectors were generated through calcium phosphate transfection of human embryonic kidney 293T (HEK293T) cells and ultracentrifugation. The following transfer plasmids, deposited by the Wernig Lab in Addgene, were used:

Tet-O-FUW-Brn2;
Tet-O-FUW-Ascl1;
Tet-O-FUW-Myt1l;
UbC-rtTA

BAM vectors were produced using a third generation system (envelope plasmid: pMD2-VSV-G; packaging plasmids: pMDLg/pRRE and pRSV-REV).

All plasmids were extracted through the Nucleobond Xtra Maxi kit (Macherey-Nagel, catalog number 700414-10) according to manufacturer's instructions.

$9x10^6$ of HEK293T cells (harboring the mutant gene of SV40 Large T Antigen) were plated in a 15-cm dish in Iscove's Modified Dulbecco's Medium (IMDM) (Sigma Aldrich, catalog number I3390), 10% FBS, Penicillin and Streptomycin (25U/mL each) and 1% L-Glutamine 200mM and incubated at 37°C, 21% O_2, 5%CO_2. The day after, 1h before transfection, medium was replaced with 22.5 ml of fresh IMDM. In the meanwhile the following mix was prepared:

transfer vector: 36 µg

pMD2-VSV-G: 9 µg
pMDLg/pRRE: 12,5 µg
pRSV-REV: 6,25 µg
0.1X TE/ddH_2O (2:1) to a final volume of 1125 µl.

Finally, 125 µl of 2.5 M $CaCl_2$ were added to the mix and the tube was put on a rotating wheel for at least 20 minutes.

DNA precipitate was obtained by drop wise addition, on vortex at full speed, of 1250 µl 2X HBS solution (281 mM NaCl, 100mM HEPES, 1.5 mM Na_2HPO_4 pH 7.12, 0.22 µM filtered) to the 1250 µl mix previously prepared. This preparation was immediately added to HEK293T cells supernatant, maintaining the pipette on the medium surface, and cross movements were executed.

Cells were successively incubated at 37°C, 21% O_2, 5% CO_2 for 14 hours and afterwards medium was replaced with 16 ml of fresh IMDM medium. 30 hours after medium changing, the supernatant was filtered through a 0.22 µm pore nitrocellulose filter and ultracentrifuged at 20000 rpm in SW32Ti rotor (Optima L-60 preparative Ultracentrifuge; Beckman) for 2 hours at 20 °C. Pellets were resuspended, without doing bubbles, in a volume of sterile PBS representing 1/400 of the starting medium volume. Finally vector was aliquoted and stored at -80°C. Usually, to have enough vector for a single experiment, 14 15-cm dishes per vector were transfected.

6.3.2 Transdifferentiation protocol

Transdifferentiation was performed according to the protocol published in the original paper[91].

2 or 7 days after 4-OHT treatment MEFs were seeded in MEFs medium on Matrigel Basement Membrane Growth Factor Reduced (BD Biosciences, catalog number 354230) coated dishes. Matrigel was thawed on ice and diluted 1:50 in DMEM/F-12 (1:1) (F-12

Nut Mix (1X) + GlutaMAX; Gibco, catalog number 31765-027). Polymerization was performed either at 37 °C for 30 minutes or overnight at 4 °C. In the second case, before plating the cells, dishes and/or multichambers were left at room temperature for at least 1 h.

The excess of matrigel was removed and the desired number of MEFs was seeded. The following day MEFs were transduced with the four vectors (0.5 µl of each vector for each 50000 cells).

Doxycycline (1:1000, mother stock 2 µg/µl) was administered the day after transduction and then every other day until the end of the analysis.

The 5th day after plating, MEFs medium was replaced with Neurobasal medium (plus doxycycline).

Neurobasal medium composition:

Neurobasal minus phenol red (ThermoFisher Scientific, catalog number 12348017)
1% Penicillin/Streptomycin (Life Technologies, catalog number 15140-122)
1% L-Glutamine (Life Technologies, catalog number 25030-024)
1% N-2 Supplement 100X (ThermoFisher Scientific, catalog number 17502048)
2% B-27 Supplement, serum free 50X (ThermoFisher Scientific, catalog number 17504044)

6.4 Transdifferentiation outcome evaluation

6.4.1 ScanR images acquisition

6.4.1.1 Immunofluorescence

For ScanR experiments MEFs were seeded in different concentrations in Nunc Lab-Tek permanox chamber slides (2 wells, Sigma Aldrich, catalog number C6682).

One day after doxycycline and 13 days after plating, cells were fixed with 4% paraformaldehyde, 20 minutes on ice.

Afterwards cells were washed 3 times with PBS and permeabilized with PBS, 10% FBS, 0.1% Triton X-100 30 minutes at room temperature.

Fixed cells were incubated with primary antibodies, diluted in PBS, 10% FBS, at 4°C overnight in agitation.

The day after 3 washes with PBS were executed followed by 1 h incubation with the secondary antibodies (diluted in PBS, 10% FBS) at room temperature, in agitation. DAPI was added to the secondary antibodies mix.

Subsequently slides were mounted with Vectamount AQ Mounting Medium (Vector Laboratories, catalog number H-5501) and imaged the day after.

The list of antibodies is reported in Table 3 and Table 4.

Name	Raised in	Company	Catalog number	Dilution
Doublecortin	rabbit	Cell Signaling	4604S	1:250
MAP2B	mouse	BD Biosciences	610460	1:500
NeuN	mouse	Millipore	MAB377	1:500
Oct3/4	mouse	Santacruz	sc-5279	1:450
Oct3/4	rabbit	Santacruz	sc-9081	1:200
Olig2	rabbit	Millipore	AB9610	1:500
TUJ1	mouse	Covance	MMS-435P	1:400
TUJ1	rabbit	Covance	MRB-435P	1:400
Vimentin	mouse	abcam	ab20346	1:200

Table 3: Primary antibodies used for immunofluorescence

Name	Raised in	Company	Catalog number	Dilution
Cy3 anti rabbit	goat	Jackson ImmunoResearch	115-165-144	1:400
Cy3 anti mouse	goat	Jackson ImmunoResearch	111-165-146	1:400
A647 anti rabbit	donkey	Life Technologies	A31573	1:400
A647 anti mouse	donkey	Life Technologies	A31571	1:400
DAPI 10 mg	/	SIGMA	D9542	1:5000

Table 4: Secondary antibodies and DAPI used for immunofluorescence

6.4.1.2 Images acquisition

A grid of 100 images per embryo in duplicate was acquired with the BX61 upright microscope equipped with a motorized stage from Olympus.

The software name was ScanR (Olympus) and the objective utilized was the 20x with a 0.75 Numerical Aperture.

6.4.1.3 Images analysis

One day after doxycycline the number of cells per image was manually evaluated as number of DAPI. Each concentration was analysed and, at the end, only conditions with the same cell number at this time point were compared. At 13 days after plating, cells positive for Class III β-tubulin (Tuj1) were classified as iNs and manually quantified. The efficiency of transdifferentiation was calculated as the percentage of $Tuj1^+$ cells among the number of DAPI at 3 days after plating.

To calculate neurite length an Imagej plugin called NeuriteTracer was used[129]. This plugin traces neurites in images stained with a neuronal marker, present also in dendrites and axons, such as Tuj1. Then it estimates the area covered by them subtracting from this value the area of nuclei that intersect them. Because this calculation depends on the number of neurons the plugin-retrieved value was divided for the number of $Tuj1^+$ cells.

For the experiments executed with ScanR unpaired t test was performed as statistical analysis.

6.4.2 Cytofluorimetric analysis

6.4.2.1 Samples preparation

Cells were detached with Accutase (Sigma-Aldrich, catalog number A6964) and counted with the Countess Automated Cell Counter (ThermoFisher Scientific) in duplicate, 5, 7, 9, 13 and 21 days after they were plated for transdifferentiation. After their centrifugation, in a tabletop refrigerated centrifuge at 4 °C for 10 minutes, at the minimum speed, they were resuspended in up to 10^6 cells per 100 µl of cold PBS-FACS (Fluorescence-Activated Cell Sorting) (PBS 0.5% FBS, 2nM EDTA). 10µl of Polysialic Acid Neural Cell Adhesion Molecule (PSA-NCAM) antibody (Miltenyi, catalog number 130-093-273) were added and the suspension was mixed and kept in the dark on ice for 10 minutes. Afterwards cells were washed with 1 ml of cold PBS-FACS and centrifuged in a tabletop refrigerated centrifuge at 4 °C for 10 minutes at the minimum speed. Pellet was resuspended in 300µl of cold PBS-FACS and 3 minutes before each tube was acquired, 5-10 µl of Propidium Iodide (PI) (Sigma Aldrich, catalog number P4170-1G. stock 50µg/ml in dH_20) were added and mixed. Tubes were kept in the dark on ice.

6.4.2.2 Samples acquisition and analysis

Samples were acquired at BD FacsCalibur with the BD CellQuest Pro Software and analysed with the Flowjo software.

As statistical analysis unpaired t test was executed.

6.5 High-throughput experiments

6.5.1 FACS sorting

FACS sorting was executed on 4-OHT treated MEFs (5 days) left in culture 7 days before transdifferentiation. 10^7 MEFs were plated on matrigel coated dishes both for the RNA-seq and the ChIP-seq on iNs at 13 days. Cells were sorted for PSA-NCAM positivity after a short detachment with Accutase. The protocol of staining was the same as the one used for the cytofluorimetric analysis (paragraph 6.4.2.1).

Cells were sorted at MoFlo Astrios (Beckman Coulter) with the Software Summit v6.2.

6.5.2 RNA-seq

6.5.2.1 RNA extraction

RNA was extracted immediately after sorting with RNeasy micro kit (QIAGEN 74004). This kit was chosen because it is exploitable for low cell amounts and it is optimum for the extraction of messenger RNA (mRNA) (due to the prevalent selection of RNA molecules longer than 200 nucleotides).

Briefly, cells were resuspended in a denaturing buffer containing guanidine-isothiocyanate (RLT buffer) plus β-mercaptoethanol, to inactivate the RNases, and then passed 5 times through a blunt 20-gauge needle. Afterwards ethanol was added to the solution to favour the exclusive binding of RNA to the RNeasy MinElute spin column. DNA contamination was eliminated adding the DNase I directly to the column and incubating for 15 minutes. After some washes, RNA was eluted in 14 µl of RNase free water and quantified at Nanodrop Spectrophotometer. RNA quality was evaluated at Agilent 2100 Bioanalyzer with the Agilent RNA 6000 Nano kit (catalog number 5067-1511).

6.5.2.2 Libraries preparation

Libraries were prepared following the TruSeq Stranded Total RNA Sample Preparation manufacturer's guidelines (Illumina) starting from 100 ng of RNA per sample.

Briefly, the steps that this protocol entailed were the following. First, through the usage of biotinylated oligos that recognized ribosomal RNA (rRNA) and of *ad hoc* beads (Ribo-Zero rRNA magnetic removal beads), mRNA was purified.

Afterwards mRNA was fragmented using divalent cations at 94 °C and retrotranscribed with random primers. The second cDNA strand was synthesised using the DNA polymerase I and the RNAse H. Beads (AMPure XP beads) were used to purify the ds cDNA. Overhangs resulting from fragmentation were converted into blunt ends using an End Repair Mix. Afterwards a single adenine was added to the 3' ends of cDNA and specific adapters were ligated to it, through their 3' overhang thymine. Finally a step of PCR amplification was performed, using primers specific for the adapters, to enrich only for ligated cDNA.

The quality of the libraries was tested at Agilent 2100 Bioanalyzer with the High Sensitivity DNA kit (catalog number 5067-4626).

Libraries were sequenced with the Illumina HiSeq machine at a read length of 100 bp, paired end, and a coverage of 120 millions of reads.

6.5.2.3 RNA-seq bioinformatic analysis

6.5.2.3.1 Alignment and quantification

RNA-seq reads were aligned with TopHat v2.0.10[130], first to mm10 Refseq transcriptome, and genes without a perfect alignment were realigned to genome (*i.e.*, --read-edit-dist 2). Quantification was performed on the Refseq transcriptome using Cuffquant v2.2.1, with

multi-read correction (-u)[130]. This pipeline was selected because it was one of the highest-performing quantification methods in our recent benchmark[131].

6.5.2.3.2 Differential expression analysis

For the differential expression analysis, we first excluded genes which had an average read count across samples below 50. We used edgeR v.3.12.1 (which outperformed other methods according to our benchmark) on the TMM-normalized estimated fragment counts. Specifically, edgeR fitted a generalized linear model that takes sex into account (~sex+genotype) on the normalized log-transformed count of each gene, and then tested, using a likelihood ratio test, whether the model including the genotype variable significantly improved the fit over the model not including it. We considered differentially expressed with high confidence genes with a FDR below 0.01 and an absolute log2(foldchange) greater than log2(1.5).

6.5.2.3.3 Alternative splicing/Differential exon usage

Alternative splicing is notoriously difficult to test statistically, and the most robust method for doing so is based on differential exon usage[132]. Differential exon usage is not exactly the same as alternative splicing: strictly speaking, it detects changes in the coverage of an exon that are associated with the independent variable, and that are discordant with variations in the other exons of the same gene. This means that alternative TSS or transcription end site results in differential exon usage, although they are generally not considered alternative splicing. However, all alternative splicing events result in differential exon usage.

Differential exon usage is based on the number of fragments overlapping each exon, we which obtained using featureCounts v.1.4.4[133] with the -O -f options on the TopHat

alignments. We then tested for significant differences using DEXseq v.1.16.7 as described in the manual.

6.5.2.3.4 Enrichment analyses

Gene Ontology enrichment analyses were performed using the goseq v.1.22.0 R package in order to correct for RNA-seq transcript length bias[134], using Fisher's test and excluding genes without annotation. Categories with at least 10 genes and maximum 1500 genes were considered. When the number of enriched categories was large, we focused on the most specific categories by removing any category with enriched children categories. (based on the graphs of the GO.db R package). Quilts were generated using the treemap R package.

6.5.2.5 RNA-seq validation

The RNA was retrotranscribed through the SuperScript VILO cDNA Synthesis Kit (Invitrogen, catalog number 11754-050).

Real Time Quantitative PCR (RT-qPCR) was performed on 7500 Fast Real-Time PCR system (Applied Biosystems) using Sybr green (Applied Biosystem) as detecting reagent. Each sample was analyzed in triplicate and normalized to *Gapdh*. Relative mRNA quantity was calculated by the comparative cycle threshold (Ct) method using the formula $2^{-\Delta Ct}$. cDNA was amplified (in triplicate) in a reaction volume of 20 µl containing 9µl of cDNA in ddH_2O and 11µl of 10 µM primers and Sybr green. The protocols envisaged the following steps: 20 seconds at 95°C, followed by 40 cycles of 3 seconds at 95°C, 30 seconds at 60°C and 15 seconds at 95°C. Afterwards a dissociation stage was present: 1 minute at 60°C, 15 seconds at 95°C and 15 seconds at 60°C. Each primers pair was checked for its efficiency of amplification. If y= - ax + q was the trendline equation,

obtained by amplification of quadratic increasing quantity of cDNA, the efficiency was calculated as (10^(-1/a)-1)*100. Only pairs of primers with efficiency between 95 and 110 were used for further analysis.

Primers are listed in Table 5.

	Forward	Reverse	Efficiency
Mll2	TGTTCGCATGAAAACGCCC	TGCAAGTGGCAGCAAAGGA	Andreu-Vieyra et al., Plos Biology 2010
Zfp238	TCTCCACTTTGCATCTGTCTC	TGCTGTGGTCTGGAAACTC	107
Syt1	TCCATACGTCAAAGTCTTCCTG	TGTCTTGCCACCTAATTCCG	103
Kif5a	GCTGGTTACAATGGCACAATC	TGGAGTAGATGTGGTTGAAGATG	107
ApoE	CAATTGCGAAGATGAAGGCTC	TAATCCCAGAAGCGGTTCAG	103
Ptprn	ATGACCTTACCCAGCATGTG	CCTGAGTTAGCAATTCCCCTG	106

Table 5: Primers used in qRT-PCR

6.5.3 Chromatin immunoprecipitation coupled to deep-sequencing (ChIP-seq)

6.5.3.1 ChIP-seq libraries preparation

10^7 MEFs were plated for the ChIP-seq for MENIN and H3K4me3 at 13 days, while, in the case of the ChIP-seq for H3K4me3 at 5 days, 5 millions of MEFs were seeded, in both cases in 15-cm dishes.

At the desired time point, medium was removed and cells were fixed with 15 ml of PBS 1% formaldehyde for 10 minutes. By this step proteins and DNA were cross-linked. To stop fixation 1 ml of glycine 2 M (final concentration 0.125 M) was added for 5 minutes. Plates were washed twice with PBS and then cells were scraped and collected in SDS Buffer. At this point cells could be stored at -80 °C until they were processed. Thawing occurred in water bath at room temperature. Cells were centrifuged in a tabletop centrifuge for 10 minutes at 400 g and then resuspended in 3 ml of ice-cold IP buffer.

Sonication was performed on ice through the Digital Sonifier 450 (Branson) with the 102C CONVERTER (Branson).

ChIP-seq for Menin (5 days) 3 cycles 30'' on/30'' off at 30% amplitude, length 700 bp.

ChIP-seq for H3K4me3 (5 days) 7 cycles 30'' on/30'' off at 30% amplitude, length 200 bp.

ChIP-seq for H3K4me3 (sorted cells, 13 days) 4 cycles 30'' on/30'' off at 30% amplitude, length 200 bp.

Part of the sonicated material was decrosslinked with the decrosslinking buffer for 1h at 65 °C, purified with the QiaQuick PCR Purification kit (Qiagen, catalog number 28104), according to manufacturer's guidelines, and run on a TAE 1% agarose gel to evaluate the DNA fragments length.

Sonicated chromatin was centrifuged at full speed for 30 minutes and the supernatant was transferred to a new tube. The quantity of chromatin was estimated using Bradford assay, using BSA (NEB) to derive a standard curve. 2 µl of sonicated chromatin were diluted in 800 µl of water plus 200 µl of Biorad Protein Assay and the absorbance at 595 nm was measured by a spectrophotometer.

Immunoprecipitation was performed in 1 ml of IP buffer on a rotating wheel at 4°C overnight using the antibodies and the conditions listed in the Table 6.

Name	Raised in	Company	Catalog number	Dilution	IP
Histone H3K4me3	rabbit	Active motif	39159	1 ug Ab each 100 ug Chromatin	100 ug
Menin	rabbit	Abcam	ab31902	1 ug Ab each 100 ug Chromatin	1 mg

Table 6: Antibodies used for ChIP-seq

10 µl of sonicated chromatin was kept as input control (1% input).

The following day 30 µl of protein G dynabeads (ThermoFisher Scientific, catalog number 10003D) (pre-equilibrated with IP Buffer) were added and the mix was incubated for 4 h on a rotating wheel at 4°C. Afterwards beads were washed 3 times with 1 ml of 150 mM Wash Buffer and once with 1 ml of 500 mM Wash Buffer with the use of a Dynamag magnet (ThermoFisher Scientific, catalog number 12321D).

Afterwards beads (and the 1% input) were resuspended in 120µl of Decrosslinking Buffer and put in agitation (at full speed) at 65 °C overnight.

Finally DNA was purified with the QiaQuick PCR Purification kit (Qiagen, catalog number 28104), eluted in 40µl of ddH_2O and quantified with Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, catalog number Q32851) at the Qubit 2.0, following manufacturer's instructions.

The purified DNA was given to the sequencing facility of the IFOM/IEO Campus that prepared the libraries as in paragraph 6.5.2.2. The starting amount of DNA was 5 ng for MENIN and H3K4me3. Libraries were sequenced on a HiSeq 2000 instrument (Illumina) following manufacturer's protocol. Sequencing was performed in single end, 50 bp, with a coverage of 30 millions of reads for inputs, H3K4me1 and menin ChIP-seq.

Buffers:

SDS Buffer:

100mM NaCl
50mM Tris-HCl pH 8.1
5mM EDTA pH 8
0.5% SDS

Triton Dilution Buffer:

100mM NaCl
100mM Tris-HCl pH 8.6
5mM EDTA pH 8
5% Triton X-100

IP Buffer:

2 volumes SDS Buffer: 1 volume Triton Dilution Buffer

150mM Wash Buffer:

150mM NaCl

20mM Tris-HCl pH 8

2mM EDTA pH 8

0.1% SDS

1% Triton X-100

500mM Wash Buffer:

500mM NaCl

20mM Tris-HCl pH 8

2mM EDTA pH 8

0.1% SDS

1% Triton X-100

Decrosslinking Buffer

1% SDS

0.1M $NaHCO_3$

6.5.3.2 ChIP-seq bioinformatic analysis

6.5.3.2.1 Alignment and peak calling

ChIP-seq reads were trimmed for potential adapter contamination using scythe 0.981 (min 4 nucleotides) before being aligned to the mm10 genome using bowtie 1.0[135] with -v 2 -m 1, and peaks were called using MACS 2.0.9[136] with default settings. MACS uses sliding windows to look for significant enrichment over the input, and relies on a shifting model (detecting the average fragment size from the gap between sub-peaks originating from each end of the fragments) to refine the boundaries of the modification or binding event. Although we relied mostly on quantitative analyses (below) to compare across replicates and conditions, when comparing peaks directly we considered peaks as overlapping if they shared at least one nucleotide.

6.5.3.2.2 Identification of target genes

Proximally bound genes were defined as having a peak within a -2.5kb/+1kb window around any of their RefSeq TSS. To identify the putative targets of intergenic sites bound by MENIN but in contact with a TSS only through chromatin conformation, we relied on Hi-C data from the most similar cell type available, namely neural progenitors. We downloaded already processed, statistically significant interactions from the Gene Expression Omnibus entry GSE68582, extracted a bed file containing each region interacting with a RefSeq TSS, and intersected it (using intersectBed) with our regions of interest to find distal targets.

6.5.3.2.3 Quantitative analysis of ChIP-seq data

Peak calling is very sensitive to coverage and technical variation, and quantitative analysis of read distribution yields considerably more robust findings. To find differences across conditions, we therefore worked on the distribution of reads falling within relevant genomic windows. For each protein/mark, the windows were defined by merging the enriched regions across samples (*i.e.,* with BedTools: cat *.bed | sortBed -i - | mergeBed -i -). In this way, all regions enriched in at least one sample were considered for statistical testing, without duplicate genomic regions. The read counts were then compared across conditions using edgeR v.3.12.1[137], using the total number of mapped reads as library size and the TMM method for normalization. Differentially enriched regions were identified using the classical dispersion model of edgeR (based on a negative binomial model) and the exact test.

For the MENIN ChIP-seq, given the low quality of the peak calling (very low and variable number of peaks across samples, probably owing to a low and highly variable coverage), we performed a peak-call-agnostic differential enrichment analysis using diffReps[138],

which uses sliding windows to directly identify regions of significant difference between sets of enrichment profiles.

6.5.3.2.4 Data representation (RNA-seq and ChIP-seq)

Principal component analyses were performed on the normalized, log-transformed FPKM (for RNA-seq) or read counts (for ChIP-seq). Unless specified otherwise, all heatmaps show row z-scores of log-transformed normalized counts. Heatmaps were produced using the pheatmap R package. +1 was added before log-transform to avoid errors on null values.

6.5.3.2.5 External data (RNA-seq and ChIP-seq)

For external datasets, we used the authors' original peak calls (ChIP-seq) or quantification (RNA-seq) that are available from the respective GEO entries.

6.6 Embryonic Stem cells (ESCs) MLL2-EGFP

6.6.1 ESCs culture

ESCs were cultured on 0.2% gelatin coated dishes. Medium, supplemented with G418 (200 µg/ml), was changed every day. Cells were passaged by Accutase.

As control for in vitro differentiation studies the ESCs E14TG2alpha have been used and cultured as ESCs MLL2-EGFP.

ESCs Medium composition:

DMEM without L-Glutamine (Lonza, catalog number BE12-614F)
15% FBS (Euroclone, catalog number ECS0196D)
0.2% LIF (produced by the Transgenic Unit of the IFOM-IEO campus)
1% L-Glutamine 200mM (Life Technologies, catalog number 25030-024)
1% Non-essential aminoacids (Lonza, catalog number BE13-114E)

0.2% β-mercaptoethanol 50 mM (Life Technologies, catalog number 31350-010)

plus or minus 2 inhibitors (2i):
MEKi PD-0325901 1 μM (mother stock 10mM Sigma-Aldrich, catalog number PZ0162)
GDK3i CT-99021 3 μM (mother stock 10mM Sigma-Aldrich, catalog number SML1046)

Mll2 exon 2 copy number was tested through Taqman as in the paragraph 6.2.3.

Immunofluorescence was performed as in the paragraph 6.4.1.1.

6.6.2 ESCs differentiation toward fibroblasts

6.6.2.1 In vitro differentiation

Differentiation was performed adapting the protocol of the paper of Xu and coworkers[139].

Briefly 3.7 10^6 ESCs were plated in 10cm nonadherent cell culture dishes in:

DMEM without L-Glutamine (Lonza, catalog number BE12-614F)
1 mM L-Glutamine (Life Technologies, catalog number 25030-024)
0.1 mM β- mercaptoethanol (Life Technologies, catalog number 31350-010)
10% FBS (HyClone, catalog number SH30088.03)
1% Nonessential aminoacids (Lonza, catalog number BE13-114E)

After 4 days the clear appearance of embryoid bodies (EB) was observed. EBs were replated on 0.2% (w/v) gelatin coated dishes for 9 days. The EBs outgrowth cells were dissociated in 2mg/ml collagenase type II for 30 minutes at 37°C and replated in MEFs medium (paragraph 6.2.1). Afterwards cells were passaged in trypsin.

Immunofluorescence was performed as in paragraph 6.4.1.1.

6.6.2.2 MEFs derivation from chimeras

ESCs injection into the blastocysts was performed by the Transgenic Unit of the IFOM-IEO campus.

Briefly, blastocysts were harvested 3.5 days post coitum from 3 weeks C57Bl/6 female mice treated with Pregnant Mare Serum Gonadotropin (PMSG) (5 UI/female) and human Chorionic Gonadotropin (hCG) (5 UI/female), subsequently bred with C57Bl/6 males.

Pseudo-pregnancy was induced through the breeding of C57Bl/6 females with vasectomized mice. Once plugged they were anesthetized (Avertin 1.25%, analgesic Rimadyl (5mg/kg) or Tramador (20µg/g)) and the chirurgic embryo transfer was performed (20-28 blastocysts per female).

Embryos were harvested at E.13.5 and MEFs preparation was executed as in paragraph 6.2.1.

MEFs selection was performed supplementing MEFs medium with G418 (100 µg/ml or 200 µg/ml). *Mll2* exon 2 copy number was tested through Taqman as in the paragraph 6.2.3.

6.6.2.3 Fibroblasts derivation from teratomas

10^6 ESCs per side were injected subcutaneously in the flank of NOD SCID IL2R$\gamma^{-/-}$ mice.

Teratomas were harvested and digested in collagenase I (750U/ml) for 1h. Then they were filtered through 0.1µm pore nitrocellulose filter and centrifuged at 500g for 3 minutes. Red blood cells were lysed 1 minute in ACK lysing buffer (ThermoFisher Scientific, catalog number A1049201). MEFs medium was added to stop the lysis and cells were centrifuged at 500g for 3 minutes.

Pellet was dissociated in trypsin, subsequently inactivated by the medium addition. Finally cells were centrifuged at 500g for 3 minutes and plated in MEFs medium at 37°C, 5%CO_2 3%O_2.

Parts of the teratoma were not dissociated, but washed in PBS and fixed overnight in 4% formalin solution. The following day they were processed by an automatized tissue processor and included in paraffin blocks. 4 µm thick sections were cut with a Leica microtome and stained with haematoxilin and eosin to assess the trilineage specification.

The immunohistochemistry staining was performed by the IEO hospital Molecular Pathology Unit.

Sections were washed in Bioclear (Bio Optica), hydrated through descending graded alcohol series and then wash in ddH_2O. Subsequently the antigen retrieval (in EDTA 1mM, Tween-20 0.05% pH 8) at 95°C was performed and endogenous peroxidase was quenched with 3% H_2O_2 for 5 minutes.

Sections were then incubated in blocking buffer (2% BSA in TBST) for 20 minutes and, subsequently with the rabbit anti-GFP antibody (Santa Cruz, sc-8334), diluted in blocking buffer, for 1h. After 3 washes, the secondary antibody (anti-rabbit Dako EnVision+ System-HRP Labeled Polymer) was added for 30 minutes.

Afterwards they were incubated with Dako Liquid DAB + Substrate Chromogen System, counterstained with hematoxylin (10 seconds), dehydrated through ascending graded alcohol series and mounted with Eukitt (Bio Optica). Signals were revealed using Dako EnVision+ Kit.

7. RESULTS

7.1 Establishment of the experimental system

The mouse models used in this study are conditional knock-out for *Mll1*[128], *Mll2*[56] or both, since the straight knock-out for these genes is embryonic lethal[55,56]. The strategy envisaged in *Mll1*, *Mll2* and *Mll1*/*Mll2* conditional knock-out is the same: mice harbour a tamoxifen inducible *Cre* (*CreERT2*) in the *ROSA26* locus, and the exon 2 of the gene encoding the specific methylase is flanked by *LoxP* sites (Figure 10a). Upon 4-hydroxytamoxifene (4-OHT) administration the exon 2 is deleted causing a frameshift mutation. In particular *Mll1* and *Mll2* mRNA translation is immediately blocked, due to a new stop codon being formed at the beginning of exon 3 (Figure 10b). These mouse models bear also the *YFP* gene in the *ROSA26* locus, downstream to a stop cassette flanked by *LoxP* sites. Initially we used the YFP positivity as an indirect indication of the entity of CRE recombination, but we later replaced this assay by *ad hoc* TaqMan assays.

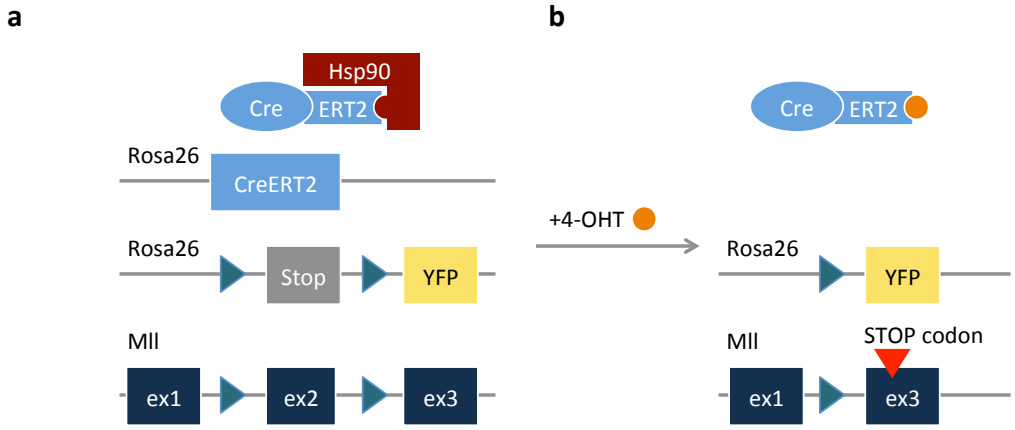


Figure 10: Schematic of the mouse models used in this study

a. Transgenic alleles of the mouse models. All the transgenic mice used in this study harbour the *CreERT2* and *YFP* genes in the *Rosa26* locus. The latter is downstream to a stop cassette between *LoxP* sites (blue triangles). Moreover the *Mll* gene, depending on the considered conditional knock-out analysed, has the exon 2 (ex2) flanked by *LoxP* sites. **b.** Deletions occur upon 4-OHT

administration. Once 4-OHT is added to the culture medium, CRE is activated leading to the deletion of both the stop cassette and the *Mll* exon 2. This causes a frameshift mutation in the *Mll* gene.

To obtain the complete deletion of exon 2, reducing to the minimum the toxicity of 4-OHT itself and of Cre activity, I conducted several experiments to define the optimal 4-OHT concentration and length of the treatment, that resulted in 5 days 1000ng/ml of 4-OHT (Figure 11a). As controls we used MEFs $Mll1^{+/+}Mll2^{+/+}CreERT2^{+}YFP^{+}$ 4-OHT-treated, to discriminate between the effects due to the absence of the specific MLL protein and the toxicity and aspecific effects of 4-OHT administration and Cre activity. Before MEFs were replated on matrigel for transduction with BAM factors they were left to recover in the medium without 4-OHT for either 2 or 7 days (Figure 11a). By leaving MEFs 7 additional days in culture before plating them, any residual MLL1 and/or MLL2 protein should be degraded and/or diluted. Hence, also the H3K4me3 mark at MLL1 and MLL2 targets should cease to be maintained, given our current knowledge of the dynamics of this mark. The degradation of MLL1 is timely regulated by the cell cycle. Indeed, MLL1 remains associated with DNA during replication, preserving the configuration of the epigenome[79,80] and it is degraded during late S and M phases[78]. It is therefore reasonable to expect that 5 days of 4-OHT treatment plus 2 days of normal medium should be already sufficient for the degradation of the residual MLL1 and for the removal of the MLL1-H3K4me3-deposited mark. The ESCs used to derive the *Mll2* conditional knock-out strain used in this study show the disappearance of MLL2 with as little as 2 days of 4-OHT treatment[64]. Hence, as for *Mll1*, 5 days of 4-OHT plus 2 days in normal medium should be sufficient to eliminate the MLL2 translated prior to the induction of Cre recombinase activity. Moreover we hypothesised that waiting an additional week before plating the MEFs for transdifferentiation would let us be more confident about MLL1 and MLL2 degradation. Even though the compensation by the homolog protein (which is, however, not present in

the double knock-out MEFs) or by other TrxG members can not be excluded by this strategy, any observed deficiency in transdifferentiation can be ascribed to a role of the specific MLL protein during cell conversion.

In all the experiments cells were collected and tested for the efficiency of exon 2 deletion on the day when they were plated for transdifferentiation. Specifically the deletion was confirmed at the genomic level, with TaqMan assays (Figure 11c), and in the case of *Mll2* also at the protein level (Figure 11d).

To understand whether the absence of the two MLLs plays a role during transdifferentiation I performed two types of experiments: immunofluorescence with images acquired at a modular microscope-based imaging platform (ScanR experiments) and cytofluorimetric analysis (FACS experiments) (Figure 11b). The aim of ScanR experiments was the semi-quantitative assessment of transdifferentiation efficiency and cell mortality, but more importantly of iNs morphology/maturation, not appreciable with other techniques. FACS kinetics was instead performed to measure the percentage of transdifferentiating MEFs and the mortality rate during the entire process.

For what concerns ScanR analysis, the experiments were carried out as follows. Different concentrations of cells were plated for experimental (flox/flox 4-OHT treated, named -/- in the rest of the thesis) and control conditions (+/+ 4-OHT treated or fl/+ 4-OHT treated, in the case of *Mll1*, named respectively +/+ and +/- in the rest of the thesis). The number of DAPI cells was counted at day 3, one day after the first doxycycline administration (*i.e.*, the plausible onset of transdifferentiation) and at the end only the conditions that at this time point were in the same number were compared to calculate transdifferentiation efficiency (Figure 11b). In fact, since any of the two epigenetic modulators is important for the regulation of cell viability/cycle[61,84,85,140] despite plating the same number of cells there could be a different number of cells in the control and the experimental condition at the moment of induction of the TFs. This could influence the final number of iNs. Since it has been demonstrated that in the day after the first doxycycline administration the majority of

the cells are postmitotic[91], it was considered as the starting point. By using this strategy any difference in transdifferentiation efficiency is imputable to an effect of MLL on the process *per se* or on the cell viability, both testable by our analysis.

In the case of FACS experiments only *Mll*$^{+/+}$ was included as control, and Polysialylated-neural cell adhesion molecule (PSA-NCAM) was used as a neuronal surrogate marker[141], whereas mortality was evaluated with Propidium Iodide (PI) staining. Moreover at the same time points of FACS analysis, cells were also counted to better assess the cell death rate.

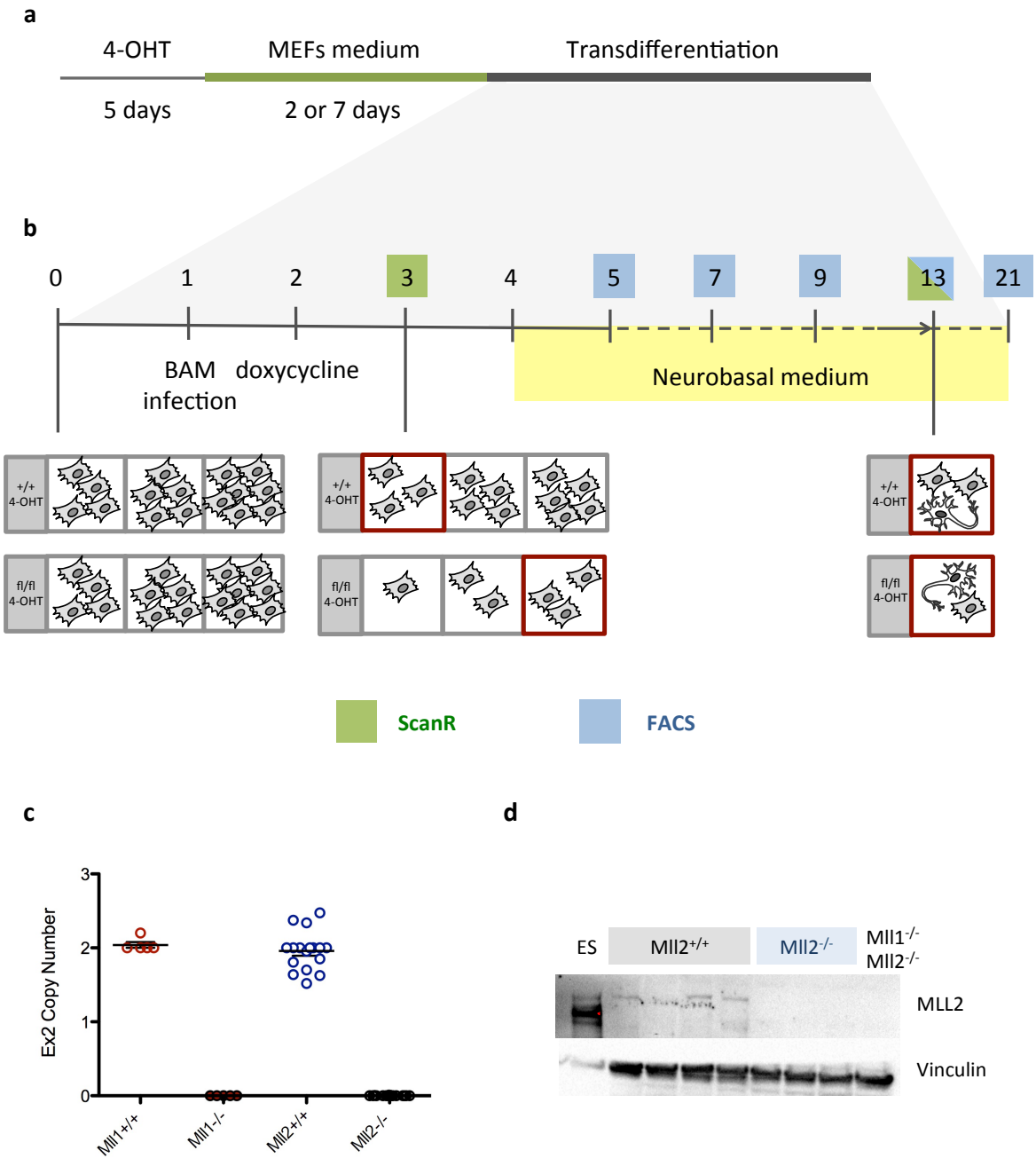


Figure 11: Experimental set-up

a. After 4-OHT treatment MEFs were plated for transduction with BAM factors either after 2 or 7 days in normal media**. b.** For ScanR experiments (time points in green boxes) experimental and control MEFs were plated in multichambers in different concentrations and imaged 3 and 13 days after plating. Only conditions that at 3 days after plating were in the same number were compared at the end (red boxes). For FACS experiments (time points in blue boxes) cells were analysed 5, 7, 9, 13 and 21 days after plating. **c.** Representative TaqMan results for exon 2 copy number of both *Mll1* and *Mll2*. Means +/- Standard Error of the Mean (SEM) are reported. **d.** Western blot for MLL2 (bands at 284 kDa and 225 kDa) with VINCULIN as housekeeping (116 kDa) in ESCs and 4 $Mll1^{+/+}Mll2^{+/+}$, 3 $Mll2^{-/-}$ and 1 $Mll1^{-/-}Mll2^{-/-}$ MEFs treated 5 days with 4-OHT and left 7 days in MEFs medium without 4-OHT.

7.2 MLL1 is dispensable during transdifferentiation

Two ScanR experiments were performed plating MEFs 2 days after the last 4--OHT administration and using $Mll1^{+/-}$ MEFs as control.

In the first experiment we observed a higher mortality of $Mll1^{-/-}$ than controls (Figures 12a and 12c on the left) and, consequently, both a lower percentage of $Tuj1^{+}$ cells at day 13 (Figure 12d) and an overall lower transdifferentiation efficiency in the experimental condition (Figure 12f on the left). On the contrary, in the second experiment, the mortality rate of $Mll1^{+/-}$ was higher (Figures 12b and 12c on the right), at 13 days there was no difference in the percentage of $Tuj1^{+}$ cells between the experimental condition and the control (Figure 12e) and the transdifferentiation efficiency was higher in $Mll1^{-/-}$ (Figure 12f on the right).

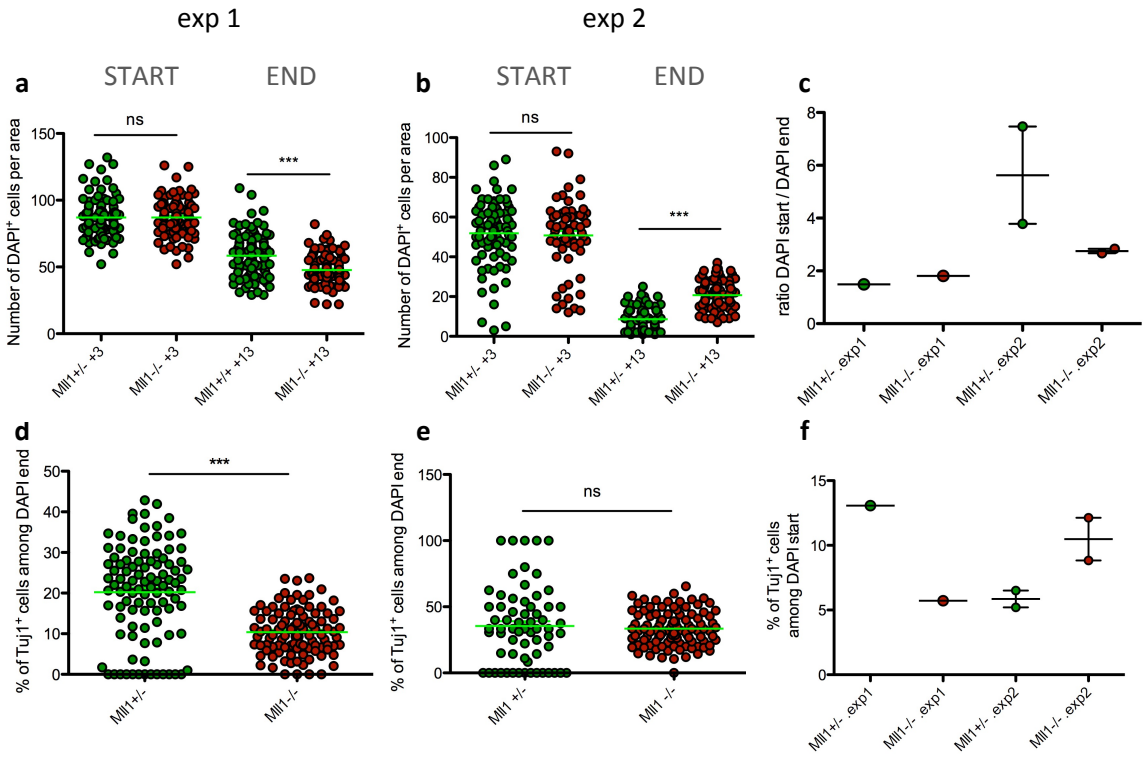


Figure 12: Transdifferentiation in the absence of MLL1: ScanR experiments

a. and **b.** Cell mortality in the absence of *Mll1* in experiments 1 and 2 respectively (exp1 and exp2). The number of DAPI in 100 images per genotype is reported. On the left the two conditions in the same number, one day after the first doxycycline administration (starting point, +3) and on the right the number of DAPI in the chosen concentration at 13 days (end of the protocol, +13). Means +/- SEM are reported. **c.** Cell mortality reported as the ratio between the sum of DAPI calculated 1 day after doxycycline administration at the chosen concentration (DAPI start) and the sum of DAPI of the corresponding chamber at 13 days (DAPI end) in the two experiments (exp1 and exp2). Means +/- SEM are reported. **d.** and **e.** Percentage of $Tuj1^+$ cells among total DAPI in 100 images acquired for the chosen concentration at 13 days, for the experiment 1 and the experiment 2 respectively (exp1 and exp2).). Means +/- SEM are reported. **f.** Efficiency of transdifferentiation in the two experiments (exp1 and exp2). The number of $Tuj1^+$ cells for the chosen concentration at 13 days was calculated and related to the number of DAPI quantified 1 day after doxycycline administration. Means +/- SEM are reported. *** $p<0.0001$; ** $p<0.001$; *$p<0.01$; ns not significant $p>0.01$.

These results suggest that the differences observed in transdifferentiation efficiency are the consequence of a difference in mortality rate among experiments. However, having only one copy of the *Mll1* gene could be sufficient for altering the cell mortality rate and/or the transdifferentiation efficiency. Afterwards, FACS experiments were performed using

Mll1$^{+/+}$ as a control and leaving MEFs in culture one additional week before plating them for BAM factors transduction. By means of this strategy we could exclude both a role of haploinsufficiency and that the phenotype observed was due to some residual MLL1 present in the knock-out MEFs at the onset of transdifferentiation. Unfortunately there are no effective commercial anti-MLL1 antibodies to test its presence.

The mortality rate was slightly higher during all analysed time points in the experimental condition with respect to the control (Figures 13a and 13b). The percentage of PSA-NCAM^{+} cells, instead, was similar in *Mll1*$^{+/+}$ and *Mll1*$^{-/-}$ throughout the experiment and as long as 21 days after plating (Figure 13c).

However, when transdifferentiation efficiency was calculated as the percentage of PSA-NCAM^{+} cells generated at 13 days with respect to the number of MEFs plated for BAM transduction it resulted lower in the *Mll1*$^{-/-}$ condition compared to the control (Figure 13d).

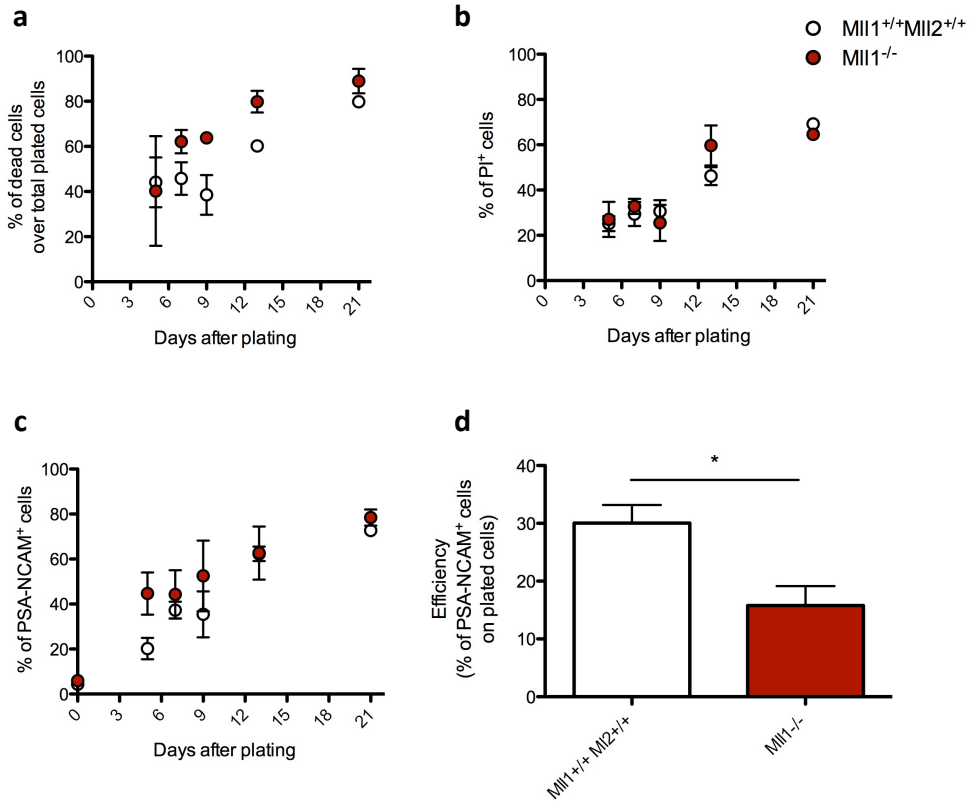


Figure 13: Transdifferentiation in the absence of MLL1: FACS experiments

a. The total number of cells was calculated 5, 7, 9, 13 and 21 days after plating with the Countess Automated Cell Counter. The percentage of dead cells was calculated over the initial number of plated MEFs. Means +/- SEM are reported. *Mll1*$^{-/-}$ n=3; *Mll1*$^{+/+}$ n=3 at all time points, but day 21: n=2. **b.** Mortality rate as percentage of PI^{+} cells 5, 7, 9, 13 and 21 days after plating, assayed with FACS analysis. Means +/- SEM are reported. *Mll1*$^{-/-}$ n=3 at all time points, but day 21 in which n=2; *Mll1*$^{+/+}$ n=3 at all time points, but day 21 in which n=1. **c.** Percentage of PSA-NCAM^{+} cells 5, 7, 9, 13 and 21 days after plating, assayed with FACS analysis. Means +/- SEM are reported. *Mll1*$^{-/-}$ n=3 at all time points, but day 21 in which n=2; *Mll1*$^{+/+}$ n=3 at all time points, but day 21 in which n=1. **d.** Efficiency of transdifferentiation calculated as the percentage of PSA-NCAM^{+} iNs cells at 13 days with respect to the number of plated cells. Means +/- SEM are reported. *Mll1*$^{-/-}$ n=3; *Mll1*$^{+/+}$ n=9. *** p<0.0001; ** p<0.001; *p<0.01; ns not significant p>0.01.

Since the percentage of PSA-NCAM^{+} cells for the two genotypes remained the same throughout the experiment, the observed lower efficiency should be attributable to the higher cell death rate in *Mll1*$^{-/-}$ condition.

Hence, also with the inclusion of the *Mll1*$^{+/+}$ cells as control, the lower transdifferentiation efficiency observed in *Mll1*$^{-/-}$ cells corresponds to a higher cell death.

Finally, in 2 out of 2 experiments, neurites of *Mll1*$^{-/-}$ iNs extended as in the *Mll1*$^{+/-}$ control (Figures 14a and 14c), further suggesting that MLL1 has no appreciable role during transdifferentiation. Although an *ad hoc* experiment would be required to prove it numerically, it is highly unlikely that this was due to haploinsufficiency because neurite length in the *Mll1*$^{+/-}$ condition was visually comparable to *Mll1*$^{+/+}$ one. Moreover both *Mll1*$^{-/-}$ and *Mll1*$^{+/-}$ iNs at 13 days were positive for the Microtubule associated protein 2b (MAP2b) (Figure 14b), a more mature neuronal marker, underlining that the absence of *Mll1* does not impair transdifferentiation.

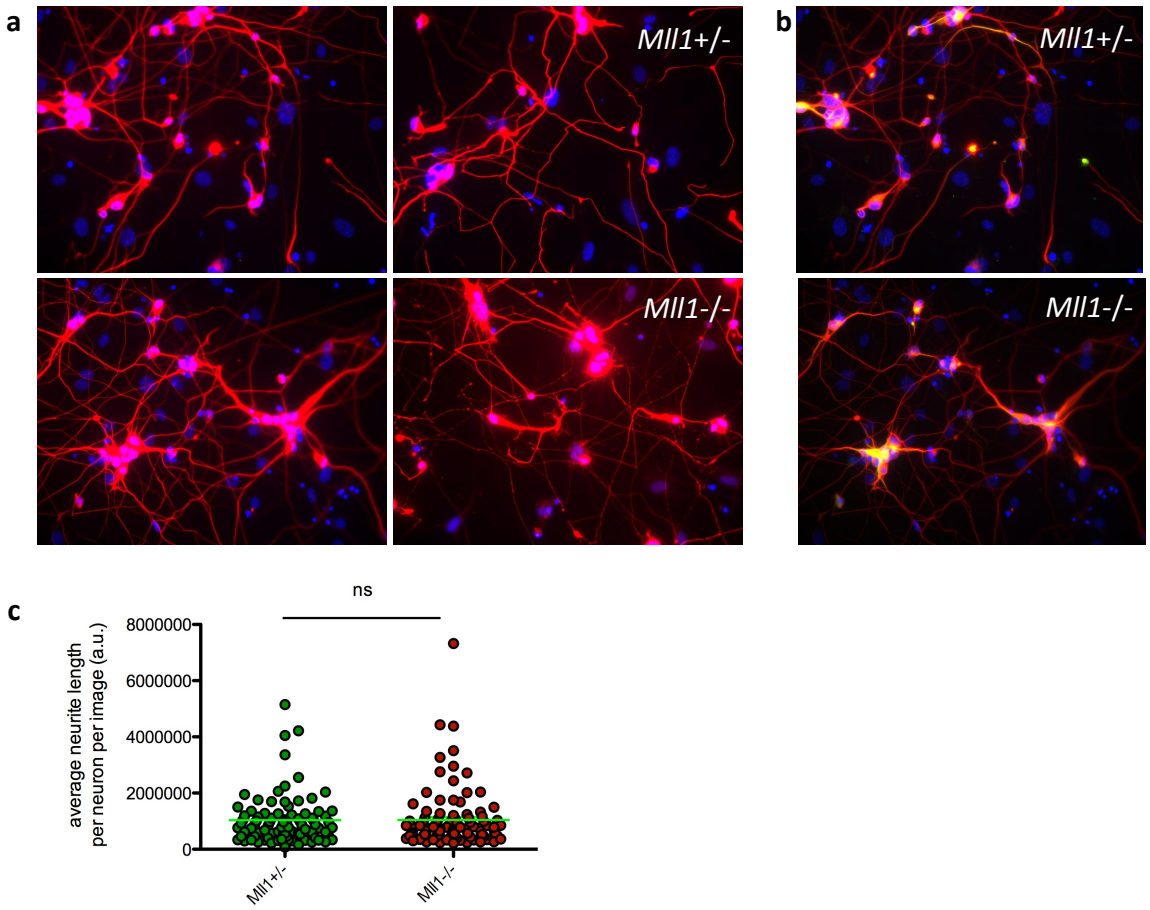


Figure 14: The absence of *Mll1* does not impair iNs morphology

a. Representative images of *Mll1*$^{+/-}$ (upper panels) and $^{-/-}$ (lower panels) iNs (DAPI is in blue and Tuj1 in red) at 13 days. **b.** Representative images of *Mll1*$^{+/-}$ (upper panels) and $^{-/-}$ (lower panels) iNs (DAPI is in blue, Tuj1 in red, MAP2b in green) at 13 days. **c.** Average neurite length per neuron calculated with Neuritetracer. Means +/- SEM are reported. ns: not significant (p>0.01).

7.3 MLL2 is necessary during transdifferentiation

The absence of MLL2 could impair the number of iNs generated because MLL2 has an active role during transdifferentiation, but also because *Mll2* deletion could lead to a higher cell mortality. However the cell viability of *Mll2*$^{-/-}$ transdifferentiating MEFs did not result affected, as demonstrated by ScanR experiments (3 out of 3 experiments) (Figures 15a and 15d on the left) and by PI staining (Figure 16d).

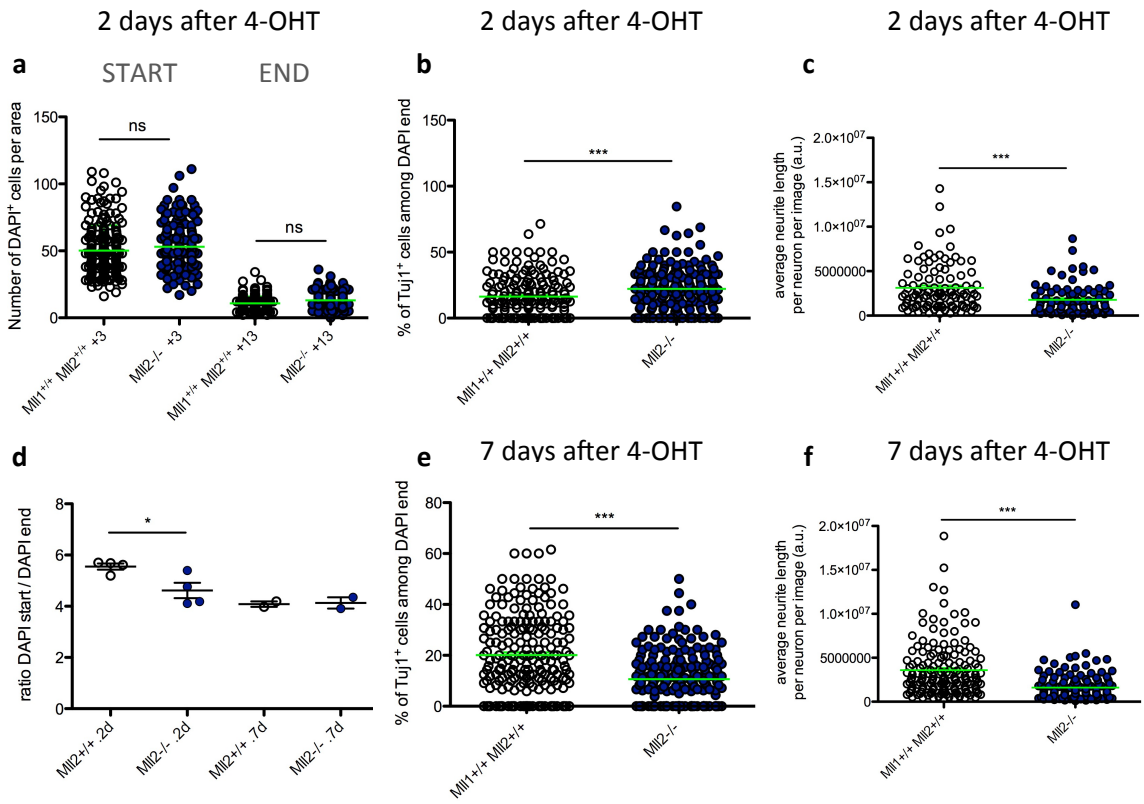


Figure 15: Transdifferentiation in the absence of MLL2: ScanR experiments

a. Cell mortality in the absence of *Mll2* in one representative experiment. The number of DAPI in 100 images per genotype is reported. Means +/- SEM are reported. On the left the two conditions in the same number one day after the first doxycycline administration (starting point, +3) and on the right the number of DAPI in the chosen concentration at 13 days (end of the protocol, +13). **b.** and **e.** Percentage of $Tuj1^{+}$ cells among total DAPI in 100 images acquired for the chosen concentration at 13 days, respectively plated 2 and 7 days after 4-OHT treatment. Means +/- SEM are reported. **c.** and **f.** Average neurite length per neuron, generated starting from MEFs plated, respectively, 2 and 7 days after 4-OHT treatment, calculated with Neuritetracer. Means +/- SEM are reported. Representative experiments are reported. **d.** Cell mortality reported as the ratio between the sum of DAPI calculated 1 day after doxycycline administration at the chosen concentration (DAPI start) and the sum of DAPI of the corresponding chamber at 13 days (DAPI end) of MEFs plated 2 (on the left) and 7 (on the right) days (2d and 7d) after 4-OHT treatment. Means +/- SEM are reported. *** $p<0.0001$; ** $p<0.001$; *$p<0.01$; ns not significant $p>0.01$.

Moreover if knock-out MEFs were left in culture one additional week before transduction with BAM factors, mortality rate was not affected (Figure 15d on the right and 16e) as also demonstrated by the comparison of absolute counts of experimental condition and control at the same time points of FACS analysis (Figure 16f).

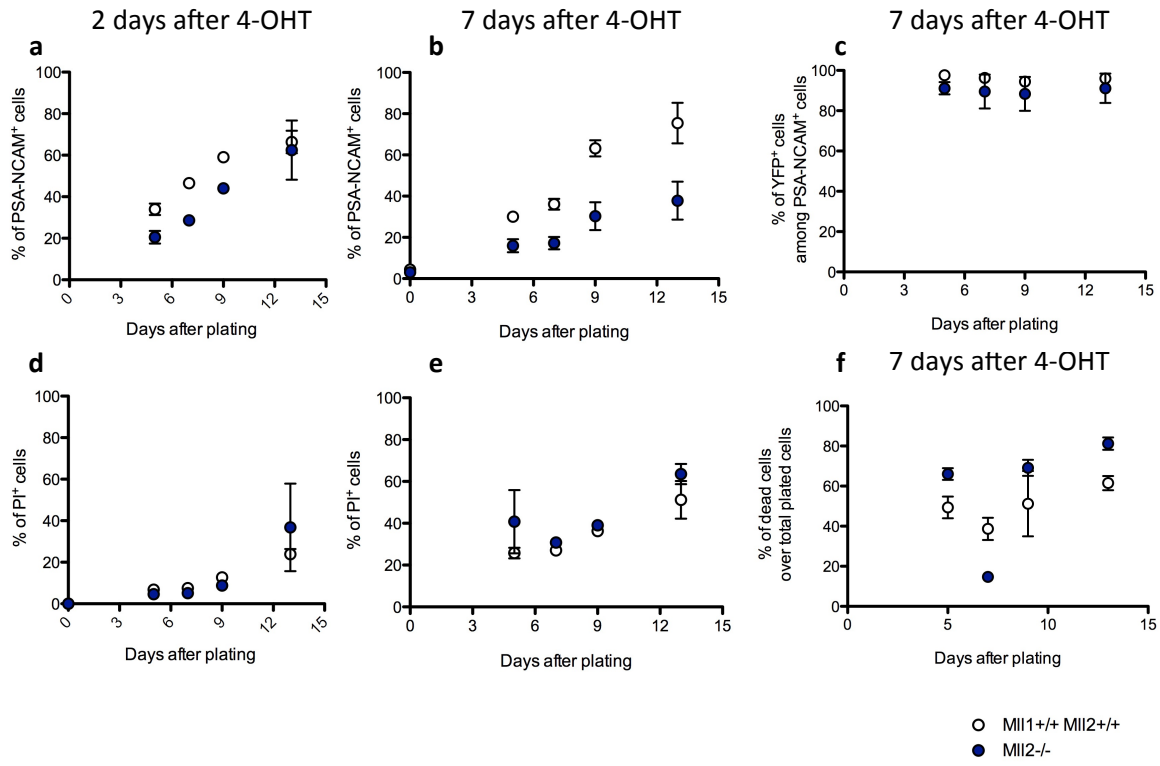


Figure 16: Transdifferentiation in the absence of MLL2: FACS experiments

a. Percentage of PSA-NCAM$^+$ cells 5, 7, 9 and 13 days after plating, assayed with FACS analysis. MEFs were plated for transdifferentiation 2 days after the end of 4-OHT treatment. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=1 at day 0, 7 and 9, n=4 at day 5 and 13; *Mll2*$^{-/-}$ n=1 at day 0, 7 and 9, n=3 at day 5 and 13. **b.** The percentage of PSA-NCAM$^+$ cells 5, 7, 9 and 13 days after plating, assayed with FACS analysis, is reported for one representative experiment. MEFs were plated for transdifferentiation 7 days after the end of 4-OHT treatment. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=1; *Mll2*$^{-/-}$ n=2. **c.** Percentage of YFP$^+$ cells in the PSA-NCAM$^+$ population assayed with FACS analysis 5, 7, 9 and 13 days after plating. MEFs were plated for transdifferentiation 7 days after the end of 4-OHT treatment. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=2; *Mll2*$^{-/-}$ n=2. **d.** Mortality rate as percentage of PI$^+$ cells 5, 7, 9, 13 and 21 days after plating, assayed with FACS analysis. MEFs were plated for transdifferentiation 2 days after the end of 4-OHT treatment. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=1 at day 0, 7 and 9, n=4 at day 5 and 13; *Mll2*$^{-/-}$ n=1 at day 0, 7 and 9, n=3 at day 5 and 13. **e.** Mortality rate as percentage of PI$^+$ cells 5, 7, 9, 13 and 21 days after plating, assayed with FACS analysis, in one representative experiment. MEFs were plated for transdifferentiation 7 days after the end of 4-OHT treatment. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=1; *Mll2*$^{-/-}$ n=2. **f.** The total number of cells was calculated 5, 7, 9 and 13 days after plating with the Countess Automated Cell Counter. The percentage of dead cells was calculated over the initial number of plated MEFs. One representative experiment is reported with means +/- SEM. MEFs were plated for transdifferentiation 7 days after the end of 4-OHT treatment. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=1; *Mll2*$^{-/-}$ n=2.

In the ScanR experiments where MEFs were immediately plated after 4-OHT treatment, transdifferentiation efficiency did not decrease in any of the 2 experiments (Figure 17a on the left). When the efficiency rate was calculated by FACS, using the same experimental timeline of ScanR experiments, an initial decrease in PSA-NCAM^{+} cells was observed, which was later overcome at 13 days (the time point analysed by ScanR) (Figure 16a). Indeed, the efficiency of transdifferentiation was higher in *Mll2*$^{-/-}$ than *Mll2*$^{+/+,}$ such as in ScanR experiments (Figure 17b on the left), but knock-out cells still proliferated after doxycycline administration (Figure 17c).

If MEFs were instead plated one week after 4-OHT treatment, we observed a lower transdifferentiation efficiency of *Mll2*$^{-/-}$ MEFs than *Mll2*$^{+/+}$ both in ScanR and FACS analyses (Figures 17a on the right and 17b on the right). The initial lower percentage of PSA-NCAM^{+} *Mll2*$^{-/-}$ iNs was maintained throughout the entire transdifferentiation process and was not overcome even 13 days after plating (Figure 16b). Moreover we can exclude that *Mll2*$^{-/-}$ cells were counterselected, since at 13 days the majority of iNs in the experimental and control conditions were YFP^{+}, which is expressed only upon Cre-mediated removal of the stop cassette upstream to the *yfp* gene (Figure 16c). Hence YFP positivity can be exploited as an indirect surrogate marker to assess the percentage of recombined cells throughout transdifferentiation.

One possible explanation for the observed discrepancy among the results of the two experimental set-ups is the following: in the case of MEFs immediately plated after 4-OHT treatment there could be a residual level of MLL2 protein and of H3K4me3 at its gene targets, respectively degraded (as confirmed by western blot) and erased/diluted if waiting for one additional week. To test this hypothesis the western blot for MLL2 has to be performed immediately after 4-OHT treatment.

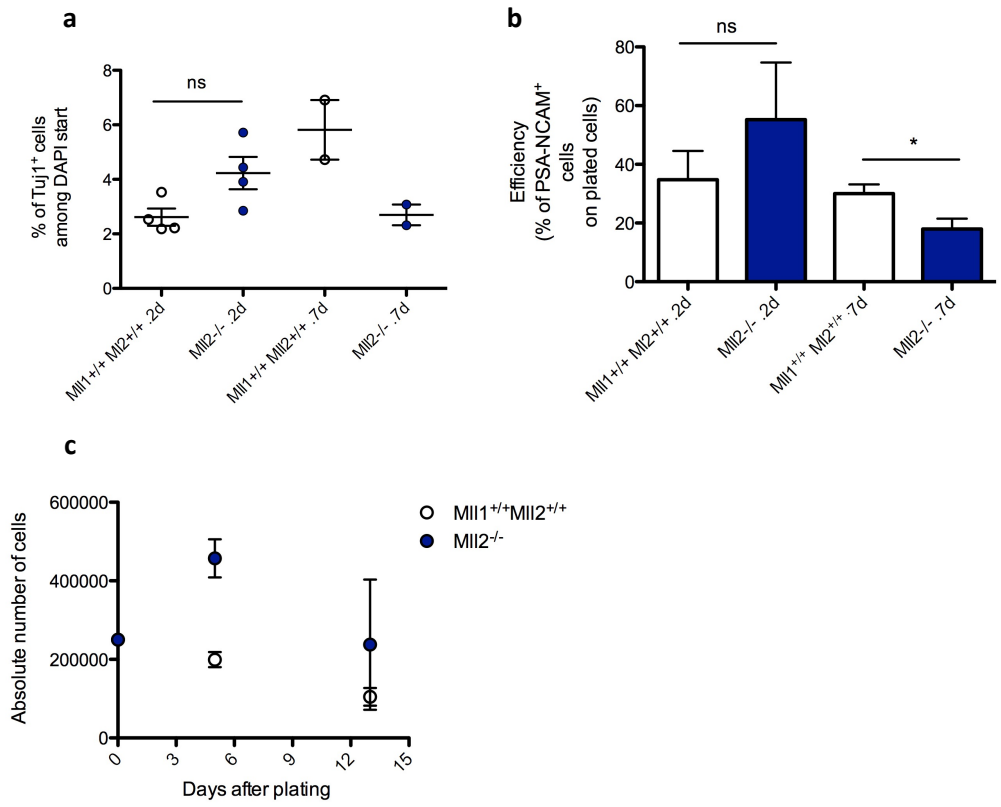


Figure 17: Transdifferentiation efficiency in the absence of MLL2

Efficiency of transdifferentiation in ScanR and FACS experiments. **a.** The number of Tuj1^{+} cells for the chosen concentration was calculated and related to the number of DAPI quantified 1 day after doxycycline administration. On the left the experiments with MEFs plated 2 days after 4-OHT treatment (2d) and on the right with MEFs plated 7 days after 4-OHT treatment (7d). Means +/- SEM are reported.**b.** Efficiency of transdifferentiation calculated as the percentage of PSA-NCAM^{+} iNs cells at 13 days with respect to the number of plated cells. On the left the experiments with MEFs plated 2 days after 4-OHT treatment (2d) (*Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=4; *Mll2*$^{-/-}$ n=3) and on the right with MEFs plated 7 days after 4-OHT treatment (7d) (*Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=9; *Mll2*$^{-/-}$ n=7). Means +/- SEM are reported.*** p<0.0001; ** p<0.001; *p<0.01; ns not significant p>0,01. **c.** Absolute cell number, evaluated through the Countess Automated Cell Counter, of *Mll2*$^{-/-}$ (n=2) and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ (n=3), 5 and 13 days after plating for transduction. MEFs were left in normal medium 2 days after 4-OHT treatment, before they were plated for transdifferentiation. Means +/- SEM are reported.

In 3 out of 3 experiments, independently of whether MEFs were plated 2 or 7 days after 4-OHT treatment and with different levels of affection, neurite elongation was impaired in Mll2$^{-/-}$ iNs with respect to control (Figures 15c, 15f and 18a). Despite this the *Mll2*$^{-/-}$ iNs,

in a preliminary experiment, were positive for MAP2b and NeuN, more mature neuronal markers (Figures 18 b and c).

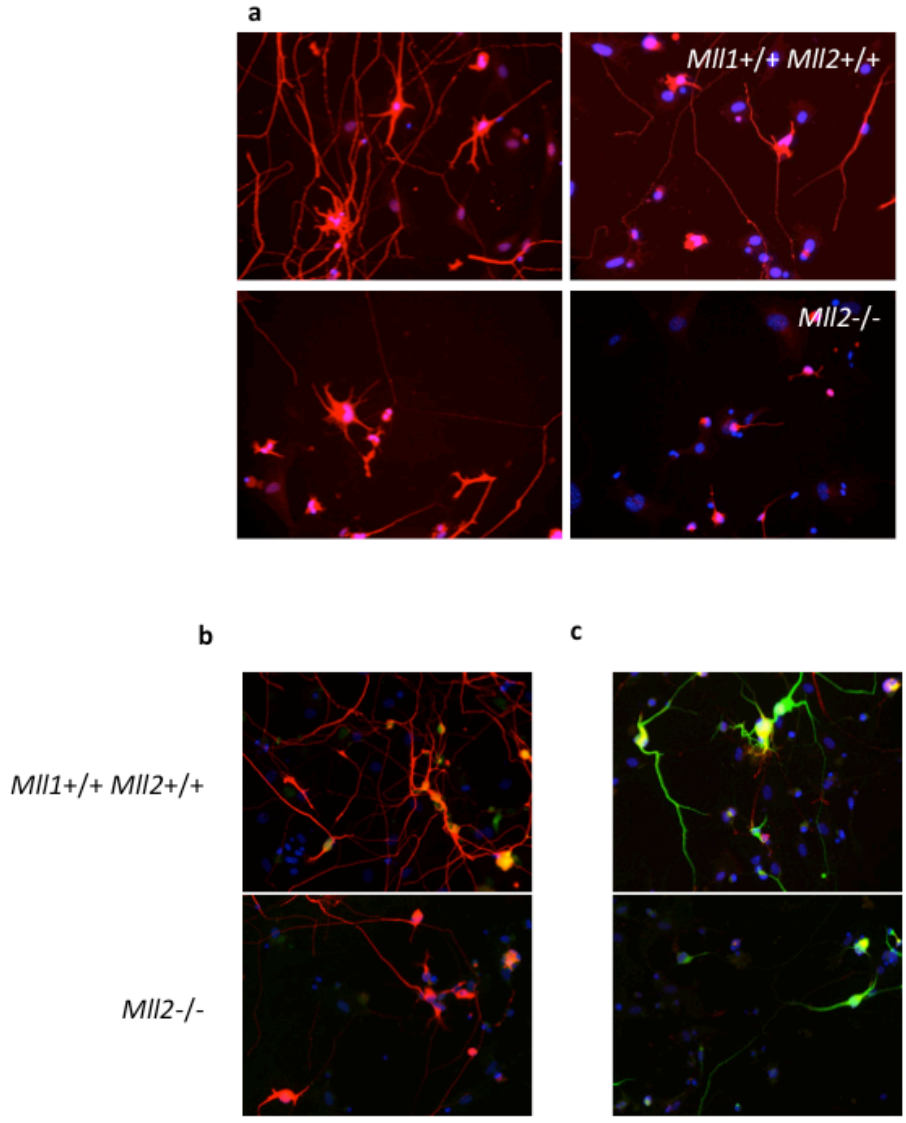


Figure 18: The absence of *Mll2* impairs iNs morphology

a. Representative images of *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ (upper panels) and *Mll2*$^{-/-}$ (lower panels) iNs (DAPI is in blue and Tuj1 in red) at 13 days. **b.** Representative images of *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ (upper panel) and *Mll2*$^{-/-}$ (lower panel) iNs (DAPI is in blue, Tuj1 in red, NeuN in green) at 13 days. **c.** Representative images of *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ (upper panel) and *Mll2*$^{-/-}$ (lower panel) iNs (DAPI is in blue, Doublecortin in red, MAP2b in green) at 13 days.

7.4 Lack of both MLL1 and MLL2 severely impairs transdifferentiation

Since MLL1 and MLL2 are two homolog proteins, to better characterize specific functions not visible in the single knock-out because of the compensation of the other methylase, we also studied the transdifferentiation of *Mll1*$^{-/-}$*Mll2*$^{-/-}$ MEFs. This strategy allowed the investigation of the overall role of the MLL1-MLL2 COMPASS-like complexes during the direct conversion MEFs-to-iNs. As in the case of transdifferentiation of *Mll1*$^{-/-}$ and *Mll2*$^{-/-}$ MEFs, also for the double knock-out I performed both the assessment of the transdifferentiation efficiency and the morphological analysis of *Mll1*$^{-/-}$*Mll2*$^{-/-}$ iNs through ScanR and FACS. In particular ScanR-based experiments were carried out on MEFs plated immediately after 4-OHT treatment, while the FACS experiment on MEFs plated one week later, to ensure the degradation of the two methylases and the erasure of the H3K4me3 mark, deposited prior to Cre activity induction.

During 4-OHT treatment, no difference in cell death was detected between *Mll1*$^{-/-}$*Mll2*$^{-/-}$ and control MEFs (Figures 19a and 19b). However upon infection with BAM factors and their induction with doxycycline, cell vitality massively dropped in the double knock-out MEFs both in ScanR and FACS experiments (Figures 20a, 20d and 20e).

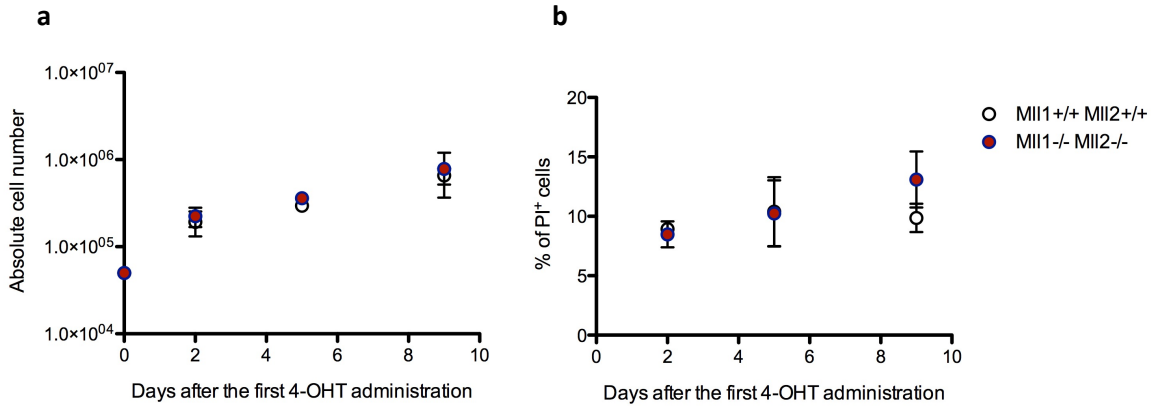


Figure 19: MEFs vitality upon deletion of both *Mll1* and *Mll2*

a. The total number of cells was calculated 2, 5 and 9 days after the first 4-OHT administration with the Countess Automated Cell Counter. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=5; *Mll1*$^{-/-}$*Mll2*$^{-/-}$ n=3. **b.** Mortality rate as percentage of PI^{+} cells 2, 5 and 9 days after the first 4-OHT administration. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=5; *Mll1*$^{-/-}$*Mll2*$^{-/-}$ n=3.

The initial decrease in the percentage of iNs generated was maintained until the last time point analysed (13 days) both in FACS and ScanR (Figures 20c and 20f).

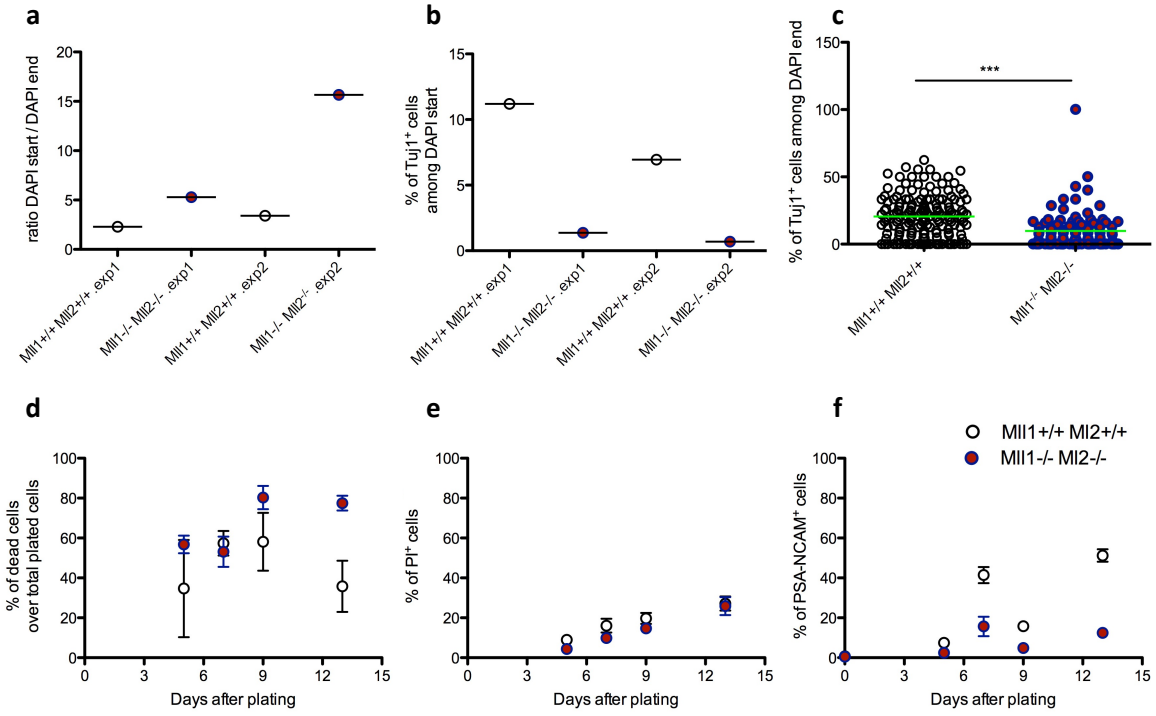


Figure 20: Transdifferentiation in the absence of both MLL1 and MLL2

a. Cell mortality reported as the ratio between the sum of DAPI calculated 1 day after doxycycline administration at the chosen concentration (DAPI start) and the sum of DAPI of the corresponding chamber at 13 days (DAPI end) in 2 independent experiments (exp1 on the left and exp2 on the right). **b.** Efficiency of transdifferentiation of ScanR experiments (exp1 on the left and exp2 on the right). The number of Tuj1^{+} cells for the chosen concentration was calculated and related to the number of DAPI quantified 1 day after doxycycline administration. **c.** Percentage of Tuj1^{+} cells among total DAPI in 100 images acquired for the chosen concentration at 13 days. One representative experiment is reported showing also means +/- SEM. **d.** The total number of cells was calculated 5, 7, 9 and 13 days after plating with the Countess Automated Cell Counter. The percentage of dead cells was calculated over the initial number of plated MEFs. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=3; *Mll1*$^{-/-}$*Mll2*$^{-/-}$ n=3. **e.** Mortality rate as percentage of PI^{+} cells 5, 7, 9 and 13 days after plating, assayed with FACS analysis. Means +/- SEM are reported. *Mll1*$^{+/+}$

Mll2$^{+/+}$ n=3; *Mll1*$^{-/-}$*Mll2*$^{-/-}$ n=3. **f.** Percentage of PSA-NCAM^{+} cells 5, 7, 9 and 13 days after plating, assayed with FACS analysis. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=3; *Mll1*$^{-/-}$*Mll2*$^{-/-}$ n=3. *** p<0.0001; ** p<0.001; *p<0.01; ns not significant p>0.01.

The transdifferentiation efficiency in *Mll1*$^{-/-}$*Mll2*$^{-/-}$ resulted very low both in ScanR (2 out of 2 independent experiments) (Figure 20b) and in FACS experiments (*Mll1*$^{-/-}$*Mll2*$^{-/-}$ n=3 batches of MEFs derived from different embryos; *Mll1*$^{+/+}$*Mll2*$^{+/+}$ n=3 batches of MEFs derived from different embryos) (Figure 21c). Moreover the few iNs generated presented very short neurites (Figures 21a and 21b).

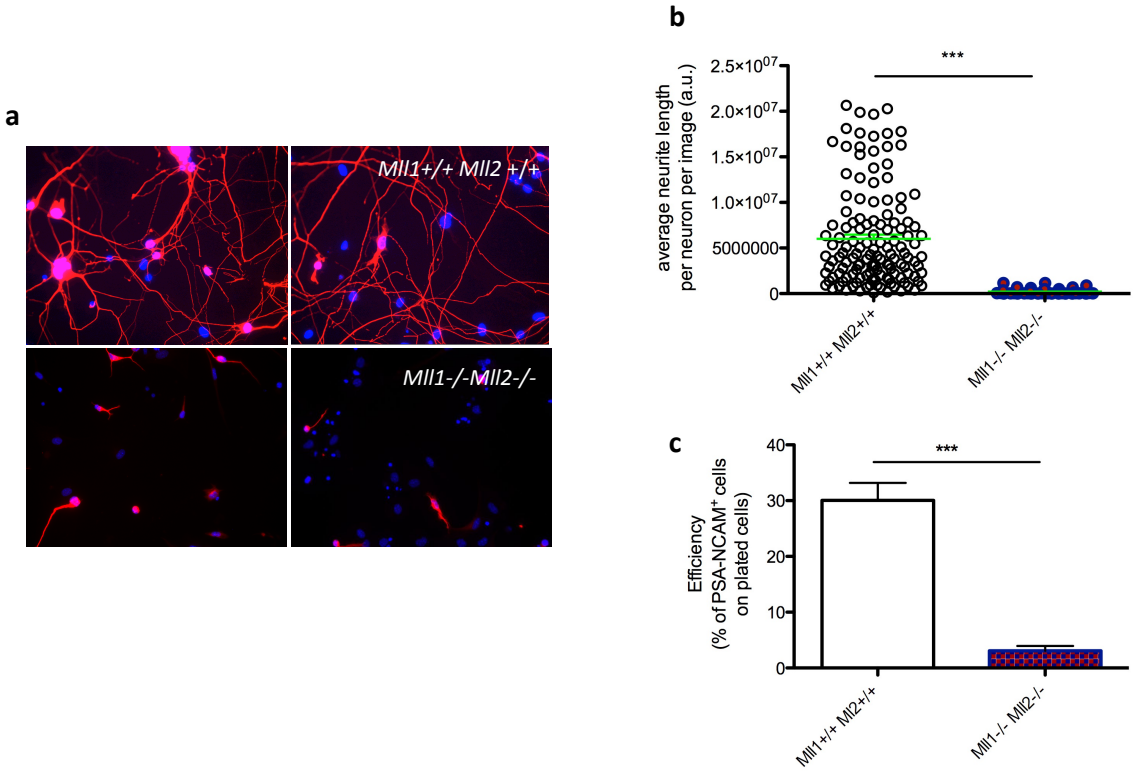


Figure 21: The absence of *Mll1* and *Mll2* highly affects transdifferentiation

a. Representative images of *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ (upper panels) and *Mll1*$^{-/-}$ *Mll2*$^{-/-}$ (lower panels) iNs (DAPI is in blue and Tuj1 in red) at 13 days. **b**. Average neurite length per neuron calculated with Neuritetracer. Means +/- SEM are reported. **c**. Efficiency of transdifferentiation calculated as the percentage of PSA-NCAM^{+} iNs cells at 13 days with respect to the number of plated cells. Means +/- SEM are reported. *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ n=9; *Mll1*$^{-/-}$ *Mll2*$^{-/-}$ n=3. *** p<0.0001; ** p<0.001; *p<0.01; ns not significant p>0.01.

7.5 Comparison of transdifferentiation efficiency across the different genotypes

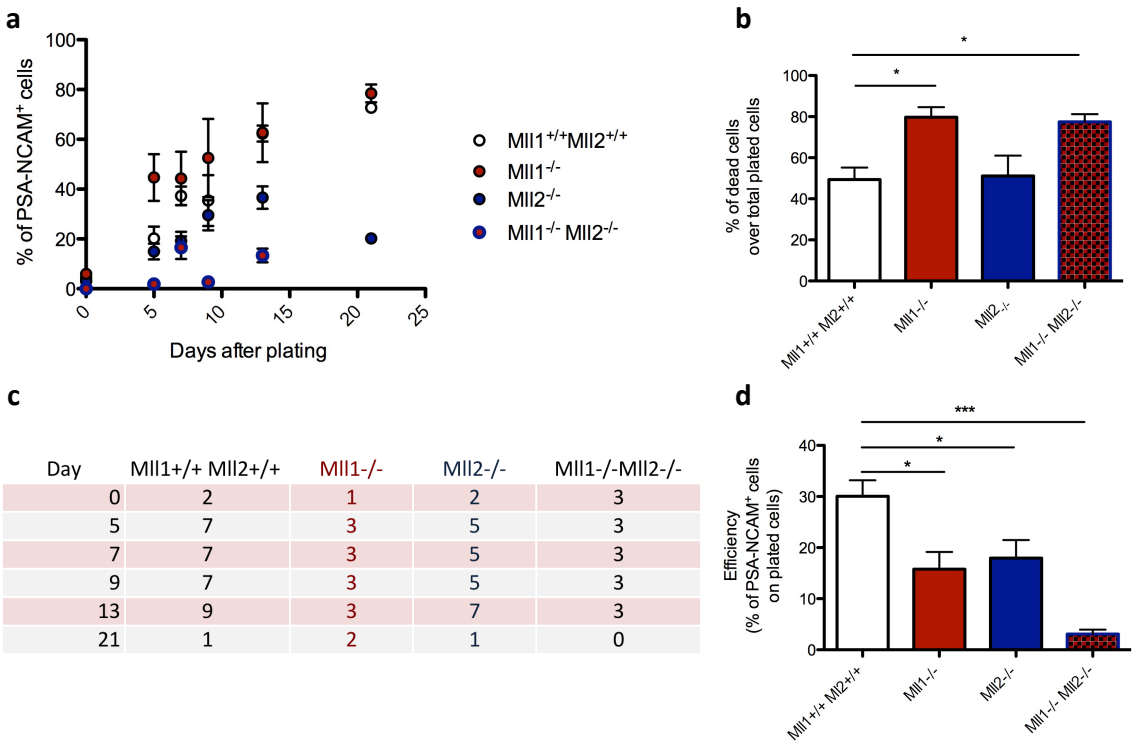
If we pooled the transdifferentiation analysis of *Mll1*$^{-/-}$, *Mll2*$^{-/-}$, *Mll1*$^{-/-}$*Mll2*$^{-/-}$ and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ MEFs, in FACS experiments *Mll1*$^{-/-}$ showed a reduction in transdifferentiation efficiency (Figure 22d), accompanied, however, by a mortality rate higher than *Mll1*$^{+/+}$*Mll2*$^{+/+}$ transdifferentiating MEFs (Figure 22b). In the case of ScanR-experiments, instead, *Mll1*$^{-/-}$ MEFs immediately plated presented the same mortality rate and transdifferentiation efficiency of *Mll1*$^{+/+}$*Mll2*$^{+/+}$ and *Mll1*$^{+/-}$ transdifferentiating MEFs (Figure 23a). While this excludes that *Mll1* is haploinsufficient for cell fate reassignment, it also suggests that one additional week in culture before plating the MEFs for transdifferentiation might be necessary to eliminate any residual MLL1 protein and hence expose the full impact of its depletion. Therefore the choice of plating the MEFs for transdifferentiation 2 or 7 days after 4-OHT treatment could explain the differences in cell mortality and in transdifferentiation efficiency between the two experimental strategies. Indeed, by FACS (in which MEFs were plated 7 days after 4-OHT), the efficiency resulted reduced for *Mll1*$^{-/-}$ MEFs (Figure 22d). Alternatively, we hypothesised that the prolonged absence of MLL1 could severely affect cell cycle/cell viability already at the MEFs state. Hence the observed defects could be the consequence of the highly compromised starting cells and not of the role of MLL1 during the transdifferentiation process *per se* (Figure 22b). Therefore, the reduction we observed in transdifferentiation efficiency should be only the consequence of lower cell viability in *Mll1*$^{-/-}$ MEFs.

Both in ScanR and FACS analysis, instead, the mortality rate of *Mll2*$^{-/-}$ transdifferentiating MEFs was the same as in the control ones (Figures 22b and 23a). Hence the observed reduction in transdifferentiation efficiency is indicative of the role of MLL2 specifically during transdifferentiation (Figures 22d and 23b).

Mll1$^{-/-}$ *Mll2*$^{-/-}$ MEFs presented the highest cell death rate (Figures 22b and 23b) and the lowest transdifferentiation efficiency (Figures 22d and 23a). In particular, in the FACS analysis, the percentage of dead cells was the same of the one of *Mll1*$^{-/-}$ transdifferentiated MEFs (Figure 22b), but the transdifferentiation efficiency was lower (Figure 22d). This underlines that MLL2 is the main MLL1-MLL2- COMPASS-like member involved in cell conversion and that it does not play a role in cell viability.

Finally, pooling all FACS experiments together, we observed that throughout transdifferentiation the *Mll1*$^{-/-}$ *Mll2*$^{-/-}$ condition had the lowest percentage of PSA-NCAM^{+} cells of all; this percentage was slightly higher in *Mll2*$^{-/-}$, while *Mll1*$^{-/-}$ and *Mll1*$^{+/+}$ *Mll2*$^{+/+}$ possessed similarly high percentages of iNs (Figure 22a). Moreover the defective morphology of *Mll2*$^{-/-}$ iNs was maintained up to day 21 (Figure 23c).

The comparison of transdifferentiation efficiency across different genotypes confirms that, while the absence of *Mll1* only impairs cell viability, the knock-out of *Mll2* impinges on the efficiency of direct cell conversion. MLL2 is therefore the main MLL1-MLL2 COMPASS-like member recruited during transdifferentiation. Indeed, the double knock-out, despite having the same cell death rate of *Mll1*$^{-/-}$ transdifferentiating MEFs, showed the lowest transdifferentiation efficiency, further underpinning a role for H3K4 MLL2-deposited trimethylation during transdifferentiation.



Day	Mll1+/+ Mll2+/+	Mll1-/-	Mll2-/-	Mll1-/-Mll2-/-
0	2	1	2	3
5	7	3	5	3
7	7	3	5	3
9	7	3	5	3
13	9	3	7	3
21	1	2	1	0

Figure 22: Comparison of transdifferentiation efficiency in all conditions: FACS analysis

a. The percentage of PSA-NCAM^{+} cells 5, 7, 9 and 13 days after plating, assayed with FACS analysis, is reported for *Mll1*$^{+/+}$*Mll2*$^{+/+}$, *Mll1*$^{-/-}$, *Mll2*$^{-/-}$ and *Mll1*$^{-/-}$*Mll2*$^{-/-}$ transdifferentiating MEFs. Means +/- SEM are reported. **b.** The total number of cells was calculated 13 days after plating with the Countess Automated Cell Counter. The percentage of dead cells 13 days was calculated over the initial number of plated MEFs. Means +/- SEM are reported. **c**. Numbers of *Mll1*$^{+/+}$ *Mll2*$^{+/+}$, *Mll1*$^{-/-}$, *Mll2*$^{-/-}$ and *Mll1*$^{-/-}$*Mll2*$^{-/-}$ samples analysed for time point. The number of samples analysed in b. and d. corresponds to the day 13th. **d**. Efficiency of transdifferentiation calculated as the percentage of PSA-NCAM^{+} iNs cells at 13 days with respect to the number of plated cells. Means +/- SEM are reported. *** p<0.0001; ** p<0.001; *p<0.01; ns not significant p>0.01.

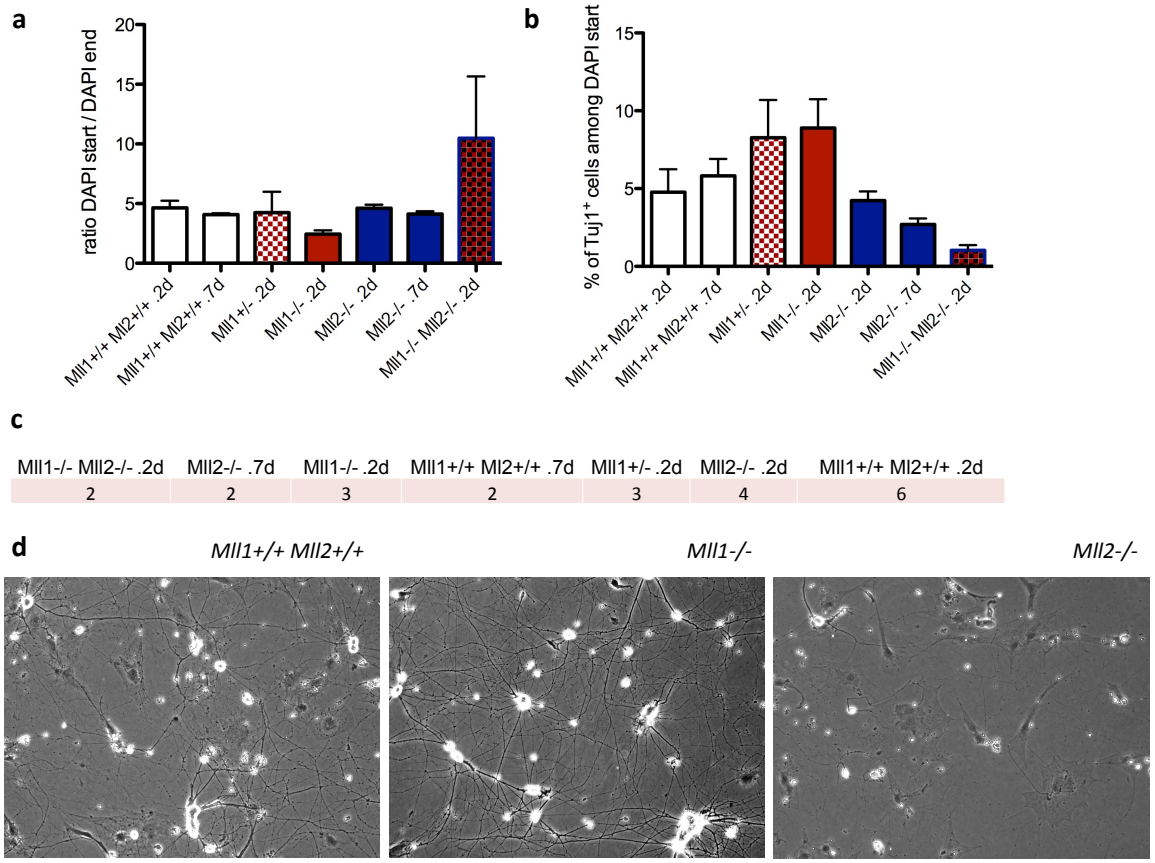


Figure 23: Comparison of transdifferentiation efficiency in all conditions: ScanR analysis

a. Cell mortality reported as the ratio between the sum of DAPI calculated 1 day after doxycycline administration at the chosen concentration (DAPI start) and the sum of DAPI of the corresponding chamber at 13 days (DAPI end). Means +/- SEM are reported. 2d: MEFs plated 2 days after 4-OHT treatment; 7d: MEFs plated 7 days after 4-OHT treatment. **b.** Efficiency of transdifferentiation in ScanR experiments. The number of $Tuj1^{+}$ cells for the chosen concentration was calculated and related to the number of DAPI quantified 1 day after doxycycline administration. Means +/- SEM are reported. 2d: MEFs plated 2 days after 4-OHT treatment; 7d: MEFs plated 7 days after 4-OHT treatment. **c.** Numbers of $Mll1^{+/+}$ $Mll2^{+/+}$, $Mll1^{-/-}$, $Mll2^{-/-}$ and $Mll1^{-/-}Mll2^{-/-}$ samples analysed in graphs b. and d. 2d: MEFs plated 2 days after 4-OHT treatment; 7d: MEFs plated 7 days after 4-OHT treatment. **d.** Representative images of 21 days $Mll1^{+/+}Mll2^{+/+}$ (left panel), $Mll1^{-/-}$, (middle panel) and $Mll2^{-/-}$ (right panel) iNs acquired at the bright field microscope.

7.6 Transcriptomic and epigenomic analysis set-up

To dissect how MLL2 modulates the epigenome and trascriptome resetting in order to enable transdifferentiation, a comparative transcriptomic and epigenomic analysis between

the experimental and control conditions was performed. We established the 5th and the 13th day after BAM factors transduction as the time points where to execute, respectively Chromatin Immunoprecipitation coupled to deep sequencing (ChIP-seq) for H3K4me3 and MENIN (5th day) and ChIP-seq for H3K4me3 and RNA-sequencing (RNA-seq) (13th day) (Figure 24). We selected the 5th day of transdifferentiation because at this stage MEFs have been cultured already 1 day in neuronal medium and 3 days have elapsed since the induction of the BAM factors. Hence transdifferentiating MEFs should be at the peak of the cell conversion process and therefore of epigenome resetting. Accordingly, at day 5 transdifferentiating MEFs, infected with *Ascl1* only, still do not express Tau-EGFP^{+}, but they are actively changing toward neuronal identity[108]. Only ~10% of *Mll2*$^{-/-}$ cells is PSA-NCAM^{+} at day 5 (Figure 22a) and we assumed that all the infected MEFs are equally competent to transdifferentiate at this stage[108]. Hence we performed the ChIP on the entire population of transdifferentiating MEFs.

As previously described, transdifferentiation does not have 100% efficiency. Therefore, at 13 days, iNs have to be separated from non-transdifferentiated MEFs for RNA-seq and ChIP-seq analyses. Moreover neurons are fragile cells for which sorting represents a major stress. Hence we performed an extensive set-up to allow the retrieval of good quantity/quality RNA and chromatin from pure populations of iNs.

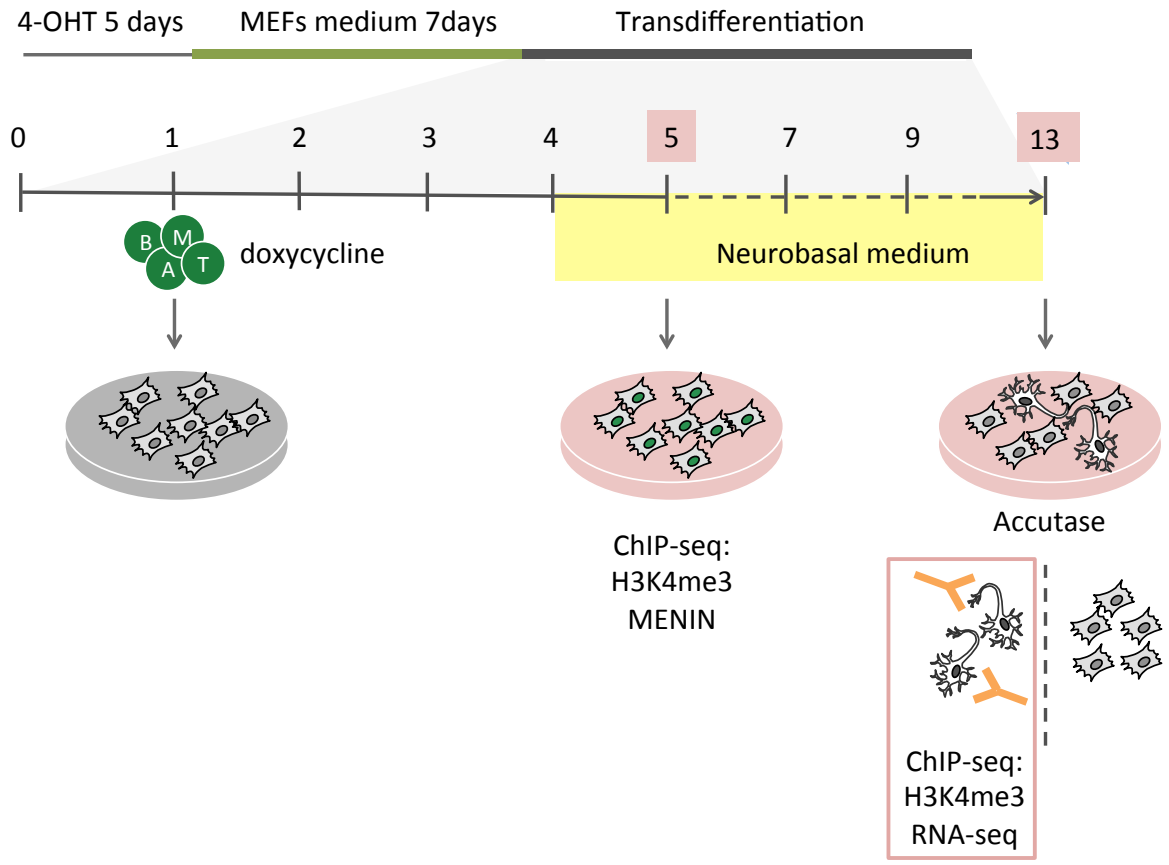


Figure 24: Experimental outline for the dissection of MLL2 role

After leaving 4-OHT-treated MEFs 7 days in medium not supplemented with 4-OHT, ChIP-seq was performed 5 and 13 days after plating. These time points were chosen because at 5 days cells are equi-competent to become iNs and are actively resetting their epigenomes. At day 13, instead, iNs were separated from non-transdifferentiated MEFs by FACS sorting and ChIP-seq was executed. In particular, at day 5 I performed the ChIP-seq for H3K4me3 and MENIN and at day 13 the ChIP-seq for H3K4me3.

I did a first pilot sorting for PSA-NCAM^{+} cells 7 days after plating. At this time point neurites are not elongated and therefore iNs should be less fragile. The goal was then to replate the pure population of sorted iNs and to extract RNA 7 days later (13 days after the MEFs plating for transdifferentiation, the time point used for all the analyses). However, almost all sorted cells died few days after plating.

Afterwards an extensive screening was performed to choose the detaching reagent less stressful for iNs. The final protocol provided sorting at 13 days after plating and the use of Accutase as detaching reagent. Accutase was selected because it allows, in 2-3 minutes, to

have all the adherent cells detached with relatively low mortality. To retrieve at least 100 ng of RNA and enough chromatin (*i.e.*, 100 µg), 10^7 cells were plated for transdifferentiation.

The gating strategy is reported in Figure 25a. The samples from which RNA was extracted and chromatin cross-linked had a purity of around 90% (Figure 25b). In the case of RNA-seq, after the collection of the positive fraction of each sample, RNA was immediately extracted with RNeasy micro kit (to obtain RNA with the level of purity required for RNA-seq) and stored at -80 °C. Extracted RNA presented both an RNA integrity number (RIN) close to 10 and a good rRNA Ratio (28s/18s) (Figure 25c). In the case of ChIP, as soon as the positive fraction was collected, iNs were immediately fixed.

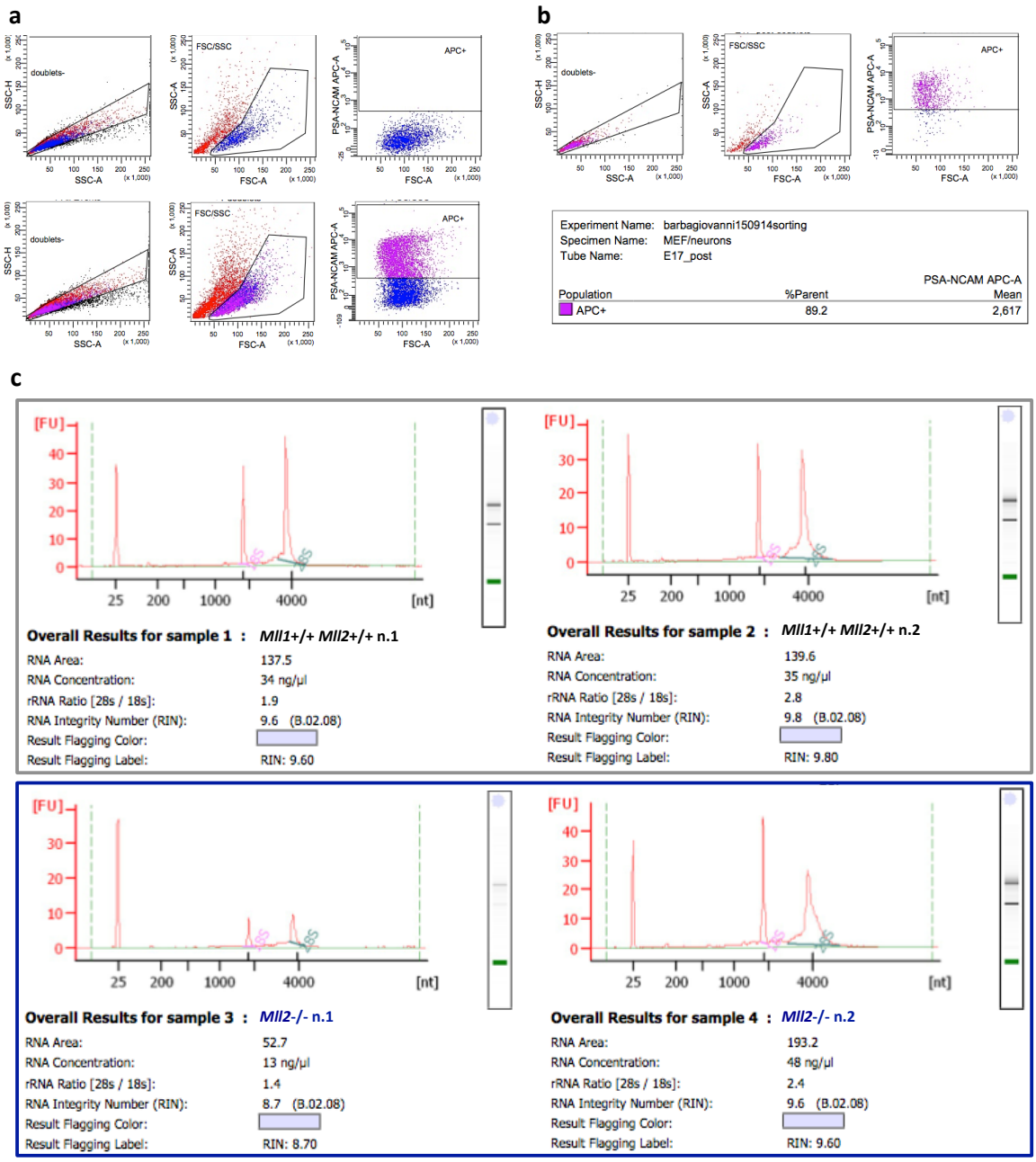


Figure 25: Sorting of PSA-NCAM^{+} fraction and RNA extraction for RNA-seq

a. Gating strategy used for sorting: doublets/ physical parameters/ PSA-NCAM positive cells. Unstained sample (upper panel); stained sample (lower panel). **b.** Analysis of purity post sorting**. c.** RNA analysis at Bioanalyzer. Two representative *Mll1*$^{+/+}$*Mll2*$^{+/+}$ (upper panels) and two *Mll2*$^{-/-}$ (lower panels) RNA samples, extracted from different embryos, are shown.

7.7 RNA-seq analysis

RNA-seq was performed on 5 *Mll2*$^{+/+}$ (ctrl_1, ctrl_3, ctrl_6, ctrl_I15, ctrl_I17) and 5 *Mll2*$^{-/-}$ (KO_11, KO_17, KO_F16, KO_F11, KO_E10) PSA-NCAM^{+} iNs derived from 10 different embryos, sorted at day 13. In particular, in order to discriminate sex-linked differences, we included 3 males and 2 females *Mll2*$^{+/+}$ transdifferentiating MEFs and 4 males and 1 female *Mll2*$^{-/-}$ MEFs. We sequenced each sorted sample with a coverage of 120 millions of reads at a read length of 100 bp, paired end. Since the H3K4me3 has been associated with alternative splicing[142-145], with this setup we could determine also if some genes were differentially spliced in the absence of *Mll2*.

The unsupervised clustering of the transcriptomes showed a clear distinction between *Mll2*$^{-/-}$ and control iNs samples (Figure 26a). Hence, although this analysis comprises two different RNA-seq rounds (*i.e.*, two different infections and sortings), iNs did not show signs of a batch effect.

With a false discovery rate (FDR) < 0.01 and a fold change (FC) > 1.5, 1828 genes resulted differentially expressed between *Mll2*$^{+/+}$ and *Mll2*$^{-/-}$ iNs and the majority of them were down-regulated in the latter, consistently with a gene activator function of MLL2 (Figure 26b).

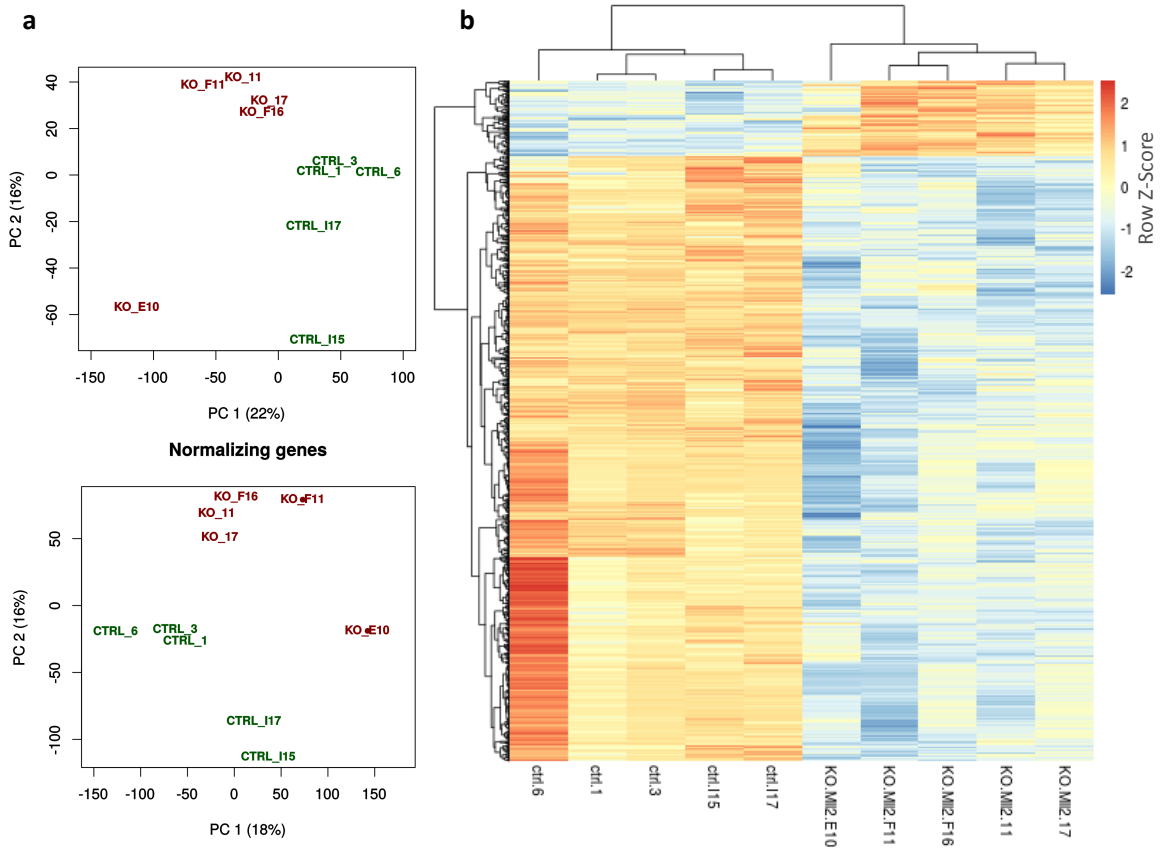


Figure 26: RNA-seq analysis of *Mll2*$^{+/+}$ and *Mll2*$^{-/-}$ iNs

a. Principal component analysis (PCA) of *Mll2*$^{+/+}$ and *Mll2*$^{-/-}$ iNs before (upper graph) and after (lower graph) genes normalization. **b.** Heat map of the differentially expressed genes between *Mll2*$^{+/+}$ and *Mll2*$^{-/-}$ iNs with a FDR < 0.01 and a FC > 1.5. The colour scale (-2 blue to +2 red) represents Z-Score. *Mll2*$^{+/+}$: ctrl_1, ctrl_3, ctrl_6, ctrl_I15, ctrl_I17; *Mll2*$^{-/-}$: KO_11, KO_17, KO_F16, KO_F11, KO_E10.

Moreover *Magohb*, the best characterize MLL2-exclusive target so far[64], was down-regulated in *Mll2*$^{-/-}$ iNs, confirming the absence of a counterselection during the 14 days of transdifferentiation (Figure 27c). The *B-cell lymphoma 2* (*Bcl2)* was down-regulated in *Mll2*$^{-/-}$ iNs as in ESCs[84], but the *Bcl-2 associated X protein* (*Bax)* and the *Bcl-2 associated agonist of cell death* (*Bad)* had similar level of expression between knock-out and control iNs (Figure 27e).

In *Mll2*$^{-/-}$ iNs *Mll2* was not down-regulated, both in RNA-seq (Figure 27a on the left) and in real time quantitative PCR (RT-qPCR) (Figures 27b and 27d), pointing to the lack of its

degradation by nonsense-mediated decay. Moreover, *Mll1* and *menin* were not up-regulated to compensate for the absence of *Mll2* (Figure 27a).

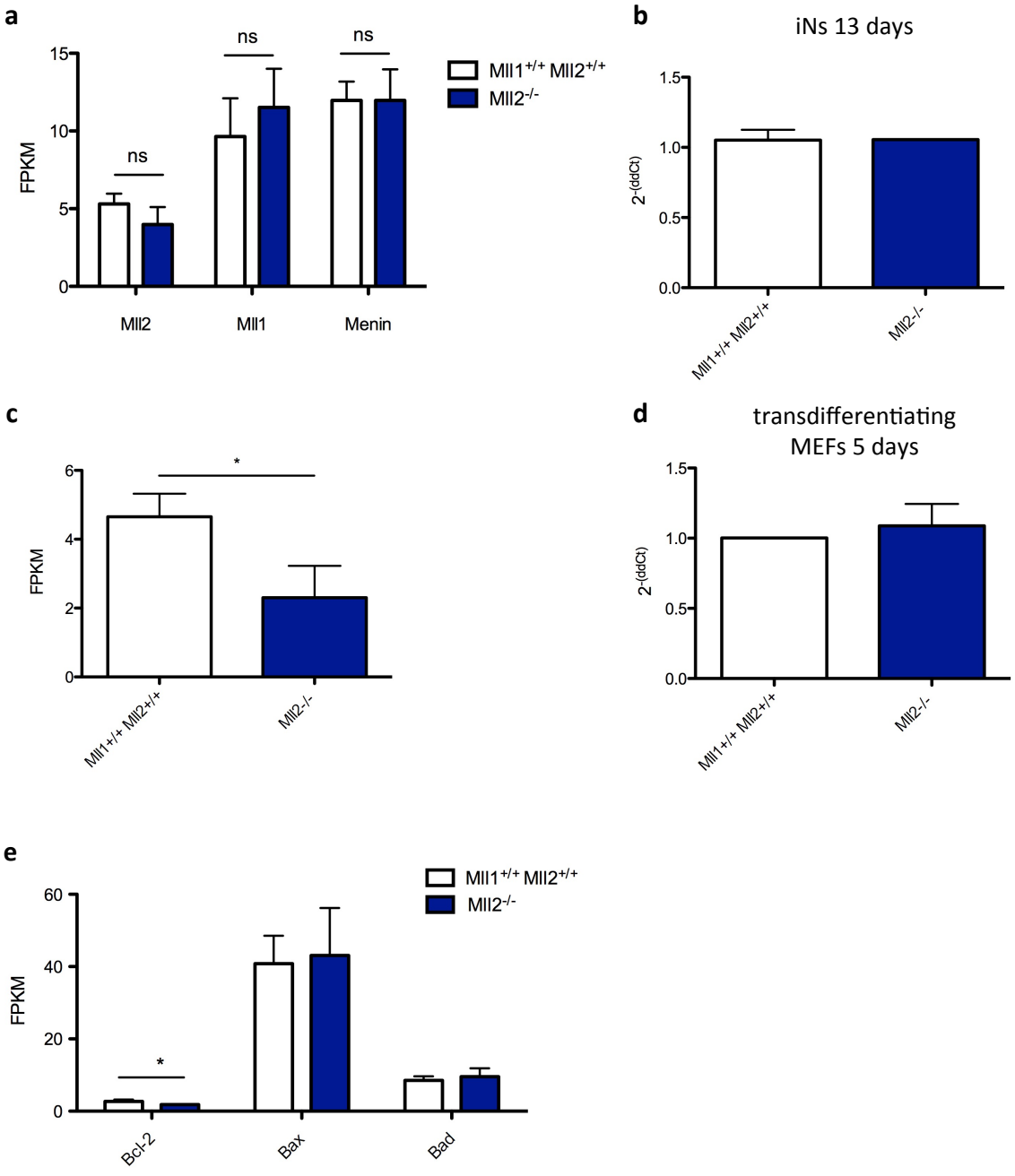


Figure 27: RNA level of MLL1-MLL2 COMPASS-like subunits, of *Magohb* and of apoptosis related genes

a. Fragments Per Kilobase of exon per Million fragments mapped (FPKM) of *Mll2*, *Mll1* and *menin* in RNA-seq analysis. Means and Standard Deviation (SD) are reported. **b.** *Mll2* RNA level of sorted iNs at 13 days, assayed by RT-qPCR. Means and SD are reported. $Mll1^{+/+}Mll2^{+/+}$ n=2; $Mll2^{-/-}$ n=1. **c.** FPKM of *Magohb* in RNA-seq analysis. Means and SD are reported. **d.** *Mll2* RNA

level in transdifferentiating MEFs at 5 days assayed by RT-qPCR. Means + SD are reported. *Mll1*$^{+/+}$*Mll2*$^{+/+}$ n=1; *Mll2*$^{-/-}$ n=2. **e.** FPKM of *Bcl2*, *Bax* and *Bad* in RNA-seq analysis. Means and SD are reported.

* FDR <0.01; ns not significant p>0.01.

Among the 10 genes, already reported by Wapinski and colleagues to be induced to guide neuronal transdifferentiation[106], 6 were repressed in *Mll2*$^{-/-}$ iNs in a statistically significant manner (Figure 28a).

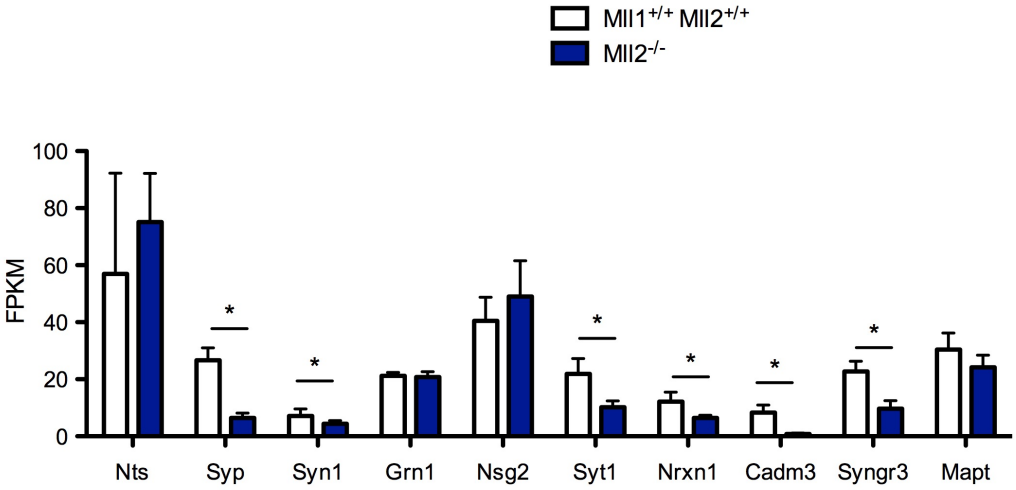


Figure 28: RNA level of neuronal induced genes

FPKM of genes reported to be induced during transdifferentiation by Wapinski and colleagues[106]. Means and SD are reported. * FDR <0.01.

Recently, the network of genes involved in transdifferentiation has begun to be elucidated[108]. In particular a subdivision has been proposed between the MEF subnetwork, the initiation subnetwork and the maturation subnetwork, based on the stage of cell fate conversion that the genes regulate (Figure 29a). When we analysed the behaviour of these genes in our RNA-seq analysis, only 3 out of ~30 genes belonging to the MEFs subnetwork (Figure 29c) and 3 out of ~25 genes belonging to the initiation subnetwork were misregulated (Figure 29e), while 7 out of 25 genes of the maturation subnetwork were down-regulated, and 1 up-regulated, in *Mll2*$^{-/-}$ iNs with respect to control (Figure

29g). In particular, the 3 genes involved in the MEFs subnetwork were down-regulated in *Mll2*$^{-/-}$ iNs, suggesting that MLL2 is, directly or indirectly, necessary for their up-regulation. On the contrary the 3 genes of the initiation subnetwork had an opposite trend from the one published by Wernig and colleagues. Indeed *Maf*, up-regulated at 13 days (Figure 29d on the left) had lower expression in *Mll2*$^{-/-}$ iNs than in controls (Figure 29e on the left), while *Tsdp2* and *Zfp238*, both down-regulated at 13 days (Figure 29d in the middle and on the right), were more highly expressed in *Mll2*$^{-/-}$ iNs than in controls (Figure 29e in the middle and on the right). This was the case also for *Insm1*, down-regulated at 13 days (Figure 29f). Therefore the analysis showed that the genes of the maturation subnetwork were the most affected in the absence of MLL2 and that the down-regulation of a repressor in *Mll2*$^{-/-}$ iNs could be the cause of the observed up-regulation of specific genes.

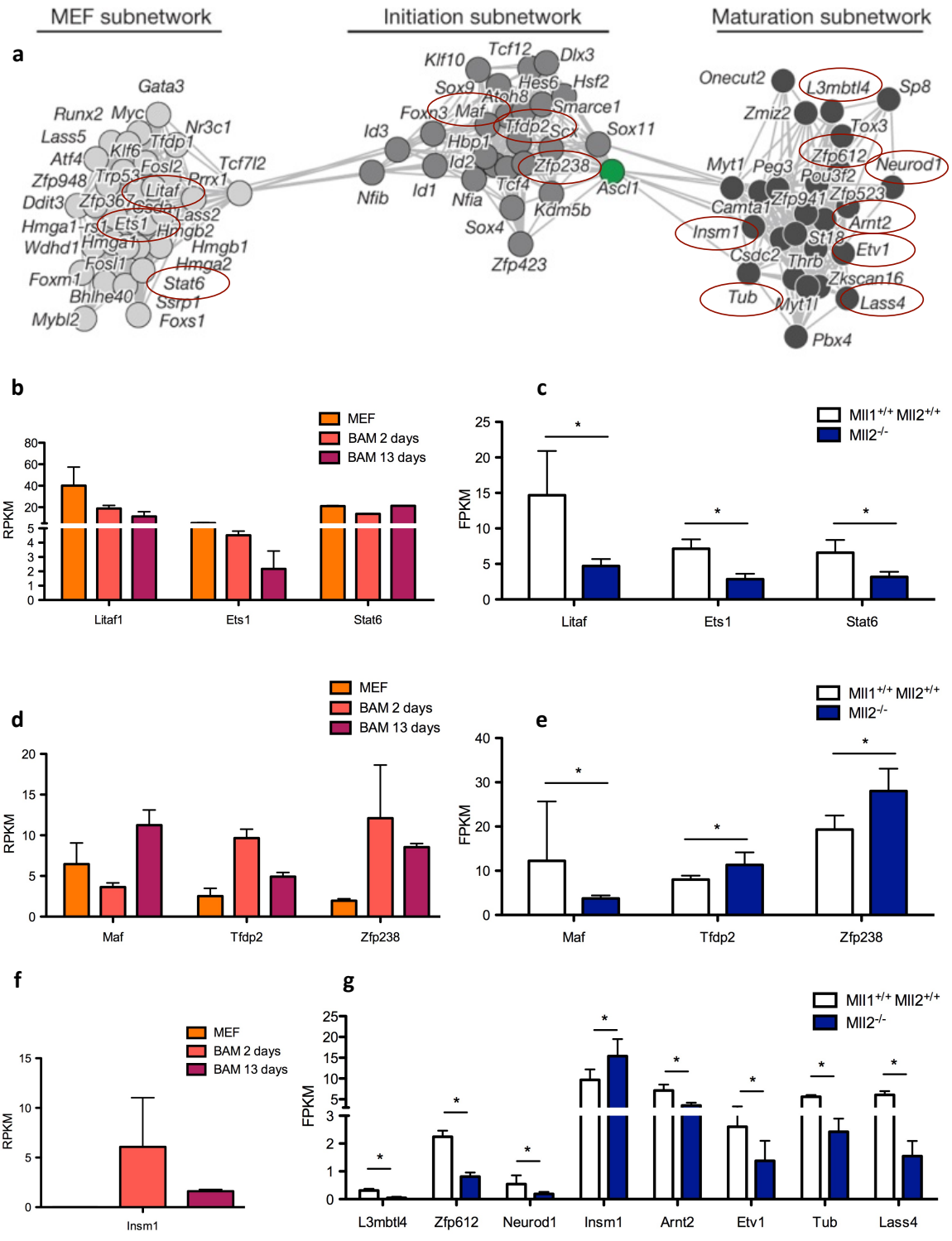


Figure 29: Misregulation of transdifferentiation subnetworks

a. Modified image from Treutlein et al., Nature 2016[108]. **b.** Reads per Kilobase per Million mapped reads (RPKM), during transdifferentiation, of the 3 genes of the MEFs subnetwork down-regulated in *Mll2*$^{-/-}$ iNs. Means and SD are reported. Source data: Wapinski et al., Cell 2013[106]. **c.** FPKM of the 3 genes of the MEFs subnetwork down-regulated in *Mll2*$^{-/-}$ iNs in our RNA-seq at 13 days. Means and SD are reported. **d.** RPKM of the 3 genes of the initiation subnetwork, misregulated in *Mll2*$^{-/-}$ iNs, during transdifferentiation. Means and SD are reported. Source data: Wapinski et al.,

Cell 2013[106]. **e.** FPKM of the 3 genes of the initiation subnetwork, misregulated in *Mll2*$^{-/-}$ iNs, in our RNA-seq at 13 days. Means and SD are reported. **f.** RPKM during transdifferentiation of the *Insm1,* the only misregulated gene of the maturation subnetwork, up-regulated in *Mll2*$^{-/-}$ iNs. Means and SD are reported. Source data: Wapinski et al., Cell 2013[106]. **g.** FPKM of the 8 genes of the maturation subnetwork misregulated in *Mll2*$^{-/-}$ iNs in our RNA-seq at 13 days. Means and SD are reported. * FDR <0.01.

When we analysed the top 30 most differentially expressed genes, which stood out for their statistical significance and/or for their FC, as expected, the majority of them were down-regulated in *Mll2*$^{-/-}$ iNs and related to neuronal function (Figure 30).

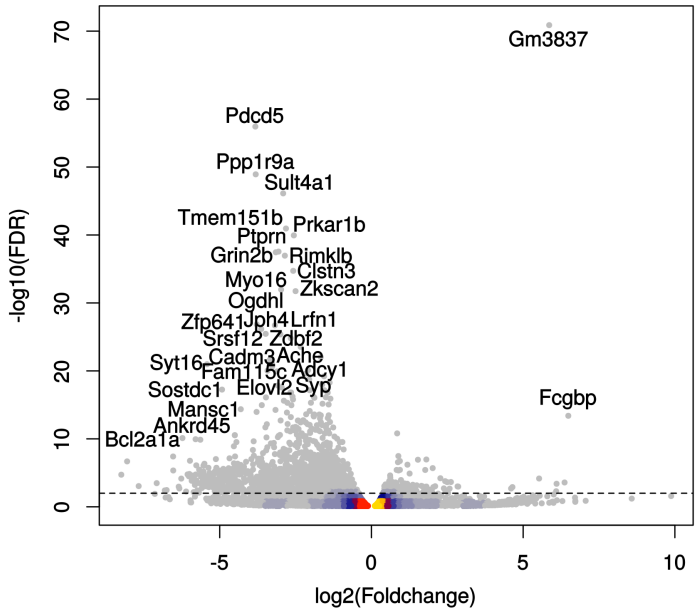


Figure 30: Top 30 differentially expressed genes

Volcano plot of the differentially expressed genes, in *Mll2*$^{-/-}$ with respect to control, with a FDR <0.01 and a FC > 0.5.

5 out of 30 genes promote physiological neurite extension, in accordance with the main defect observed in *Mll2*$^{-/-}$ iNs. The *Protein phosphatase 1 regulatory subunit 9A (Ppp1r9a)* or *Neurabin-I,* is a neuronal specific-F-actin binding protein important for neurite outgrowth and dendritic spine formation, hence crucial both at precocious and late stages of neuronal maturation. Its deletion impairs cortical and hippocampal neuron projection

elongation[146-148], in line with our results. Also *Myosin 16* (*Myo16)*, which belongs to Neuronal tYrosine-phosphorylated Adaptor for the PI 3-kinase (NYAP) phosphoproteins family, favours neuritogenesis through the interaction with the WASP-family verprolin homologous protein (WAVE) complex and the Phosphatidylinositol 3 (PI3) kinase[149,150]. *Calsyntenin 3* (*Clstn3*) is involved in both excitatory and inhibitory axon presynaptic organization. Mice *Clstn3*$^{-/-}$ present a reduction in synapse density[151]. Moreover both MYO16 and CLSTN3 interact with NEUREXIN-1[151,152], also down-regulated in *Mll2*$^{-/-}$ iNs (Figure 31b), fundamental for synapse differentiation and transmission. The *Leucine-rich repeat and fibronectin type III domain-containing protein 1* (*Lrfn1*), and to a lower extent the other family members *Lrfn3* and *Lrfn5*, are genes down-regulated in the knock-out (Figure 31c) (their trend during transdifferentiation is reported in Figure 31a), that by similarity of the domains they encode have been associated with neurite outgrowth[153]. Finally the type I-beta regulatory subunit of protein kinase A (*Prka1b*) has been linked to neurophilament phosphorylation, the main axon cytoskeleton components[154,155].

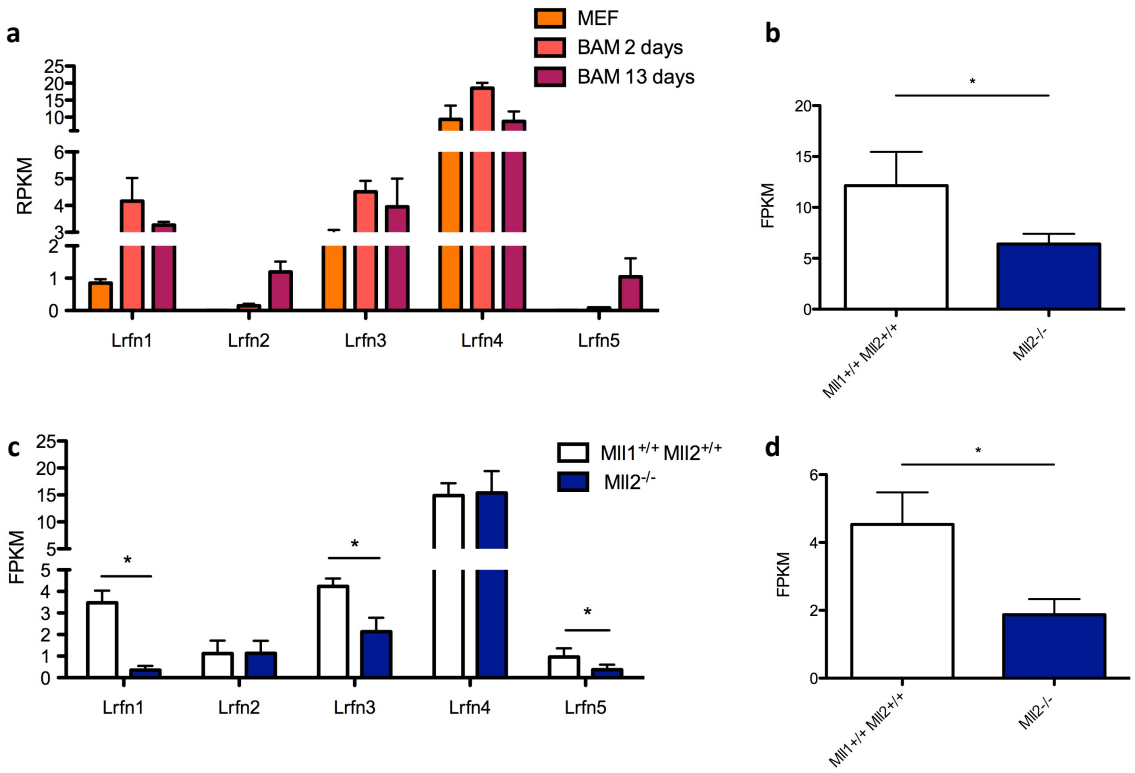


Figure 31: Differentially expressed genes involved in synapse and neurite formation

a. RPKM, during transdifferentiation, of the genes of *Lrfn* family. Means and SD are reported. Source data: Wapinski et al., Cell 2013[106]. **b.** FPKM of *Nrxn1* in our RNA-seq at 13 days. Means and SD are reported. **c.** FPKM of *Lrfn* family members in our RNA-seq at 13 days. Means and SD are reported. **d.** FPKM of *Jph3* in our RNA-seq at 13 days. Means and SD are reported. * FDR <0.01.

8 genes were linked to synapse formation and neurotransmission. JUNCTOPHILIN 4 (Jph4) and Jph3 (Figure 31d) are the two brain-specific proteins involved both in the formation of complexes between the plasma membrane and the endo-sarcoplarmatic reticulum (*i.e.*, subsurface cisternae) and in the synaptic plasticity[156]. *Synaptophysin* (*syp*) encodes for a synaptic vescicle transmembrane protein, important for synaptic vescicle endocytosis[157] and SYNAPTOBREVIN (*i.e.*, a synaptic vescicle transmembrane protein, important for synaptic vescicle exocytosis) retrieval[158]. Mice $syp^{-/-}$ do not show neurotransmission defects[159], but their learning and memory are altered[160]. The absence of a clear synaptic transmission phenotype is attributed to the compensation of other synaptic vescicles proteins: the SYNAPTOPORIN and the SYNAPTOGYRINS (syngr1-4). Interestingly, among *synaptogyrins*, only *syngr1* and *syngr3* are up-regulated as long as transdifferentiation process progresses[106] (Figure 32b) and they were both down-regulated in $Mll2^{-/-}$ iNs (Figure 32a).

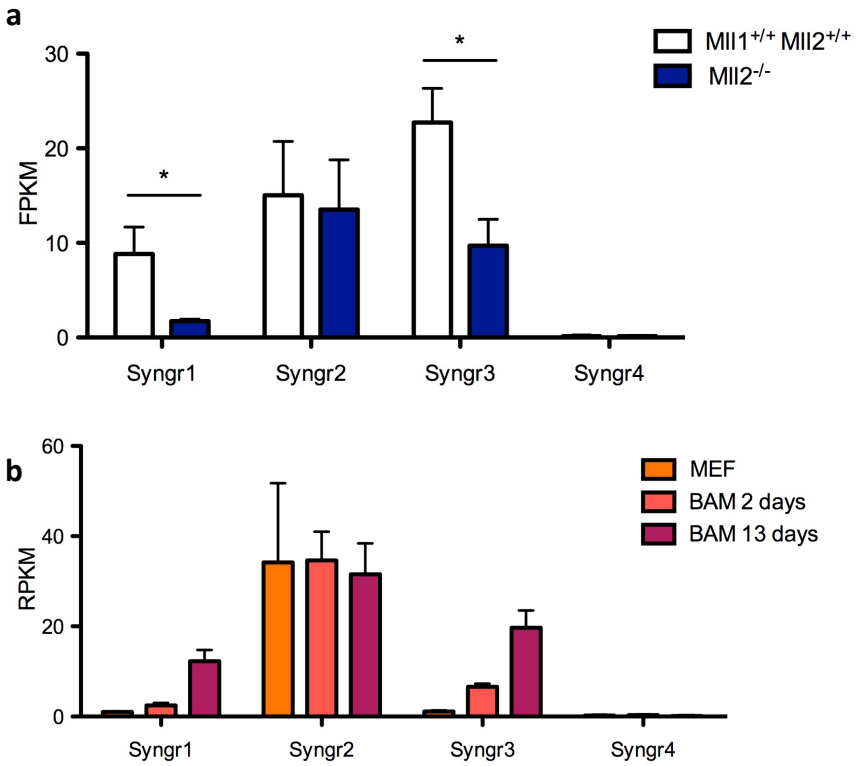


Figure 32: Synaptogyrins

a. FPKM of *Synaptogyrin* (*Syngr*) genes in our RNA-seq at 13 days. Means and SD are reported. **b.** RPKM, during transdifferentiation, of the genes of *Syngr* family. Means and SD are reported. Source data: Wapinski et al., Cell 2013[106].*FDR <0.01.

The *sulfotransferase family 4A member 1* (*Sult4a1*) is associated with schizophrenia, since it is located in a region often deleted in schizophrenic patients[161] and is believed to possess a role in neurotransmitter metabolism. The protein tyrosine phosphatase receptor type N (PTPRN) or insulinoma associated protein 2 (IA-2) is a membrane-associated vescicular protein that, despite being one of the main type 1 diabetes autoantigens, is involved in neurotransmitter release[162] and in the nervous system it is mainly localized in neurites[163]. *Cell adhesion molecule 3* (*Cadm3*) belongs to the immunoglobulin-like family and it is important for the synapse formation and also for myelinisation and axon bundles[164]. *Type-1 adenylyl cyclase* (*Adcy1*) is fundamental for memory formation, coupling Ca^{2+} release to cyclic AMP. Finally the glutamate ionotropic receptor NMDA type subunit 2B (*Grin2b*) encodes, as the acronym states, for the subunit with the agonist binding site of the NMDA

receptor, while the ACHETYLCOLINESTERASE (Ache) hydrolyzes the neurotransmitter acetylcolin.

The *programmed cell death 5* (*Pdcd5*) is a gene whose level increases during apoptosis[165]. Moreover the knock-down of *Pdcd5* reduces the apoptosis rate[166]. Hence its down-regulation in *Mll2*$^{-/-}$ iNs further confirms their lower cell death.

The majority of the other genes are poorly characterized in literature and few of them are without a clear relation with the observed phenotype.

Using as a guiding criterion the gene ontology (GO) analysis we chose for validation 14 genes linked to neuronal maturation and with a high Fragments Per Kilobase of exon per Million fragments mapped (FPKM), but after several trials only for 5 of them (Figures 28, 29e and 33a) I was able to find primers with a good efficiency of amplification. The residual precious RNA was spared to validate the targets coming out from the integrated analysis with the H3K4me3 ChIP-seq. In particular we analysed *Zfp238* and *Syt1,* fundamental during the establishment of transdifferentiation subnetworks (Figures 28 and 29d), *Ptprn*, one of the highest differentially expressed genes (Figure 30) and the *apolipoptrotein E (ApoE)*, that, besides having a role in Alzheimer disease, has been involved in neurite extension[167,168]. Also the *Kinesin family member 5A* (*Kif5a*) was chosen for its role in neurite outgrowth[169]. In all the analysed genes the reduction was confirmed but only 2 of them were statistically significant (Figures 33b-f). *Zfp238*, one of the main down-stream effectors of ASCL1[106] was confirmed to be higher in *Mll2*$^{-/-}$ iNs (Figure 33b). We also analysed its level 5, 9 and 13 days after MEFs were plated for transdifferentiation, in unsorted cells in two independent experiments observing that this *Ascl1*-effector was highly induced at 5 days in *Mll2*$^{+/+}$ cells and lowly repressed at 13 days in *Mll2*$^{-/-}$ cells (Figure 33h).

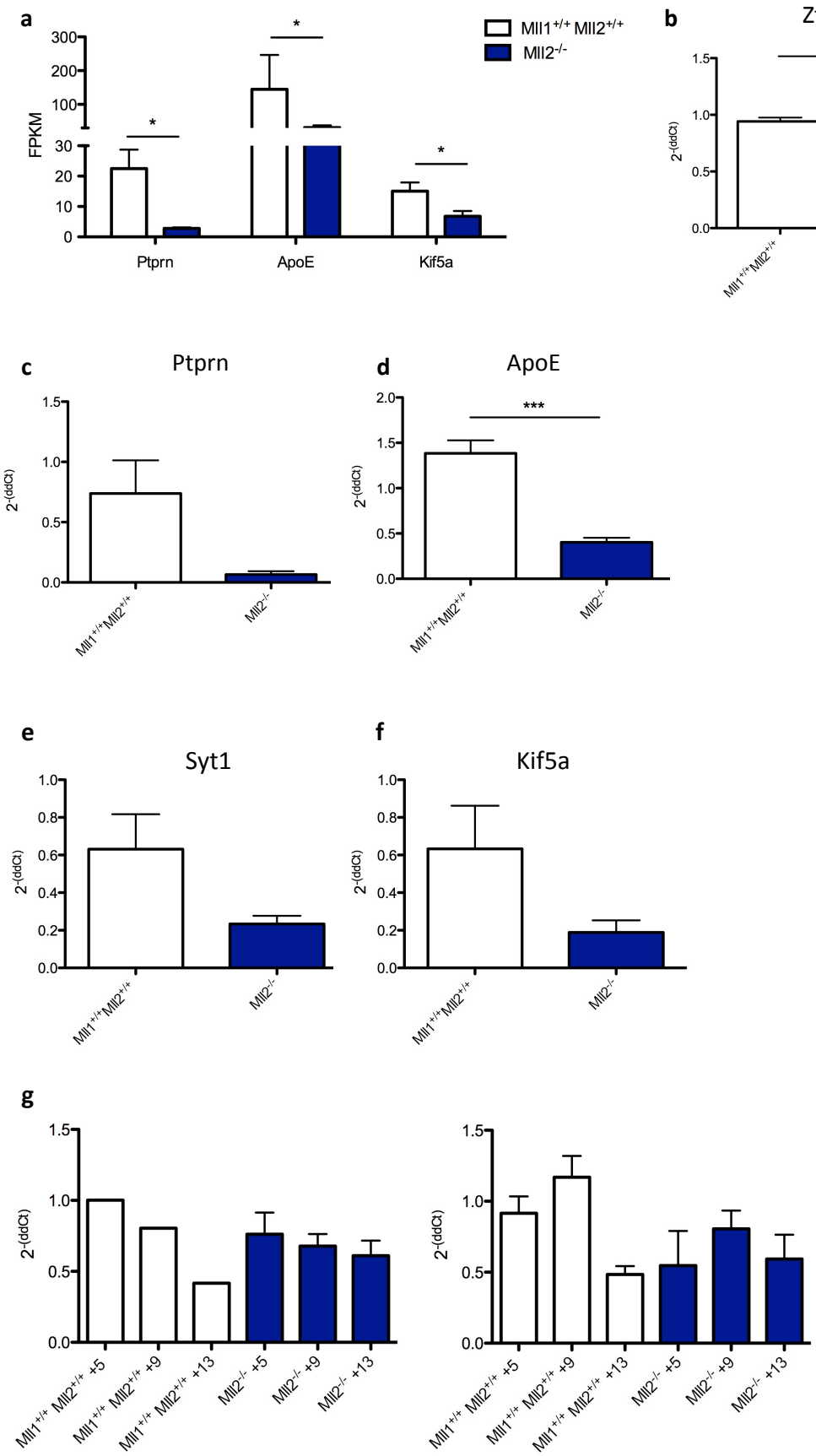


Figure 33: RNA-seq validation

a. FPKM of the genes chosen for validation. Means and SD are reported. *FDR <0.01. **b.** to **f.** RNA level of the genes chosen for validation assayed by qPCR. Means and SD are reported. ***

p<0.0001; ** p<0.001; *p<0.01. **g.** *Zfp238* RNA level in unsorted cells 5, 9 and 13 days after plating, assayed by RT-qPCR. Means and SD are reported. The two graphs represent 2 independent experiments. On the left $Mll1^{+/+}Mll2^{+/+}$ n=1, $Mll2^{-/-}$ n=2. On the right $Mll1^{+/+}Mll2^{+/+}$ n=2, $Mll2^{-/-}$ n=2.

Hence, in the absence of MLL2, the process of transdifferentiation starts, but the networks are misregulated, as confirmed by the level of *Zfp238* and of all the aforementioned genes. Moreover several differentially expressed genes are involved in neurite extension, the main defect observed in $Mll2^{-/-}$ iNs.

7.8 Differential splicing

Among the top 30 differentially expressed genes the *Serine/Arginine-Rich Splicing Factor 12* (*Srsf12*) was present, further suggesting that splicing might be affected (Figure 30). In particular SRSF12 induces distal 5' splicing sites and functions as a repressor of canonical Serine Arginine rich proteins[170].

We found 7 genes subjected to differential exons usage in $Mll2^{-/-}$ with respect to the control PSA-$NCAM^{+}$ iNs. By this analysis *Mll2* came out as differentially spliced at exon 2, confirming the lack of MLL2 in knock-out iNs and therefore the absence of any counterselection for $Mll2^{+/+}$ MEFs that could have survived through the end of 4-OHT treatment (Figure 34).

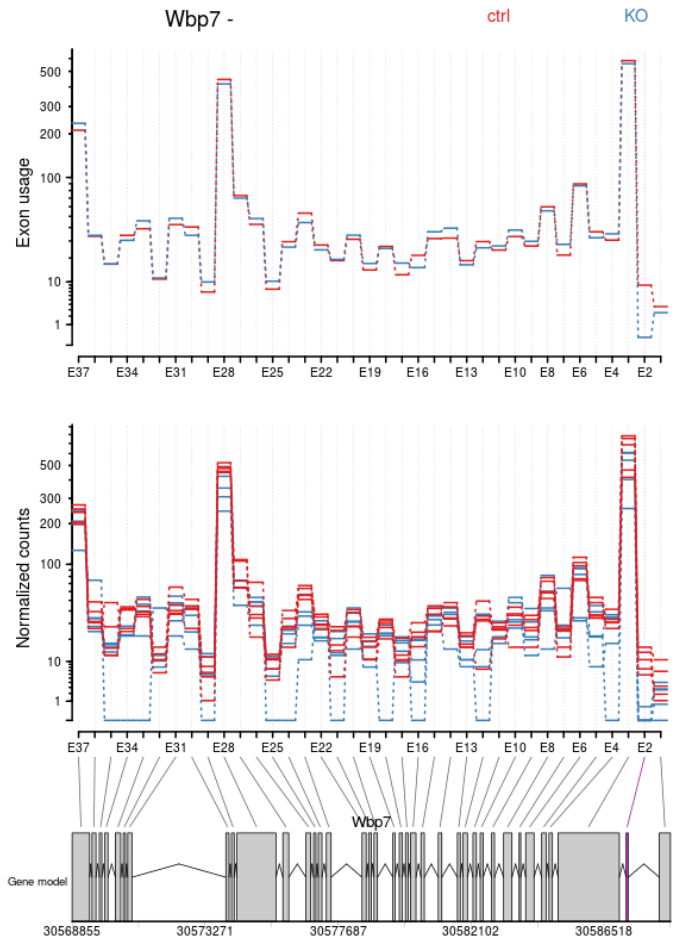


Figure 34: *Mll2* exon 2 deletion

In the top graph *Mll2* exon usage in $Mll2^{-/-}$ and $Mll1^{+/+}Mll2^{+/+}$ iNs. In the middle graph normalized read counts. In the bottom graph schematic of the differentially used exons. The statistically significant ones are in purple. $Mll2^{-/-}$ is in blue and $Mll1^{+/+}Mll2^{+/+}$ in red.

The *ArfGAP with GTPase domain, ankyrin repeat and PH domain 2* (*Agap2*) possesses two known isoforms (Figure 35d). The first isoform (*i.e.*, NM_001033263) is brain-specific and negatively associated with neuronal apoptosis through the interaction with Phosphatidylinositol 3 (PI3) kinase. Recently it was also demonstrated to be involved in neurite outgrowth[171]. The isoform 2 (*i.e.*, NM_001301014), instead, prevents apoptosis by interacting with Akt. In our datasets we detected only the isoform 1, that, moreover, is significantly down-regulated in $Mll2^{-/-}$ iNs (Figure 35b). The $Mll2^{-/-}$ *Agap2* transcript presented a lower usage of exon 7, encoding for part of the small GTPase domain, while the $Mll2^{+/+}$ *Agap2* transcript showed a lower usage of exon 19, containing part of the ankyrin-repeat containing domain, a protein-protein interaction domain (Figures 35a and

c).

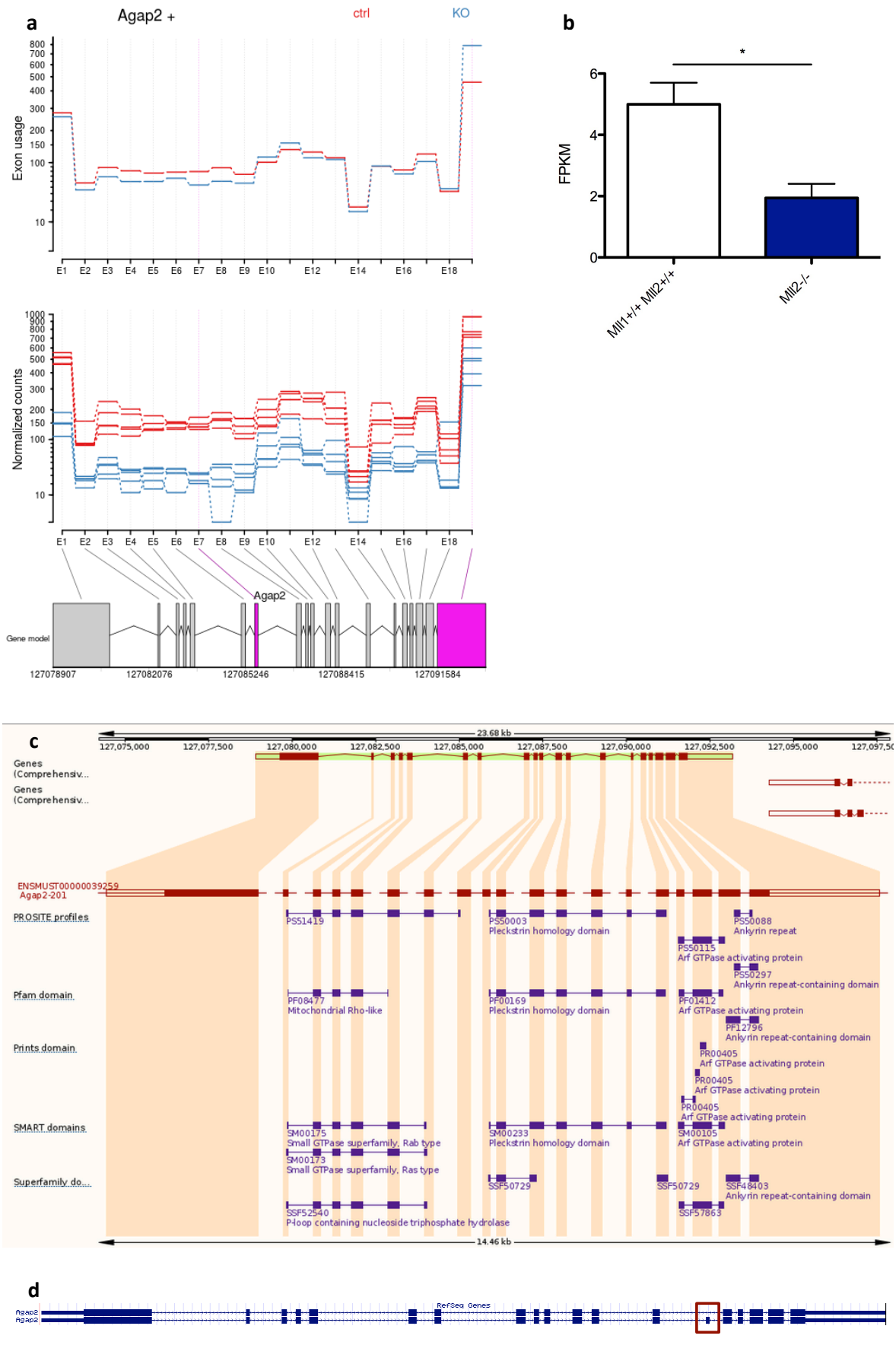

a
Agap2 +
ctrl
KO
Exon usage
E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E12 E14 E16 E18
Normalized counts
Agap2
Gene model
127078907
127082076
127085246
127088415
127091584
b
FPKM
Mll1+/+ Mll2+/+
Mll2-/-
c
23.68 kb
Genes (Comprehensiv...
ENSMUST00000039259 Agap2-201
PROSITE profiles
PS51419
PS50003 Pleckstrin homology domain
PS50088 Ankyrin repeat
PS50115 Arf GTPase activating protein
PS50297 Ankyrin repeat-containing domain
Pfam domain
PF08477 Mitochondrial Rho-like
PF00169 Pleckstrin homology domain
PF01412 Arf GTPase activating protein
PF12796 Ankyrin repeat-containing domain
Prints domain
PR00405 Arf GTPase activating protein
SMART domains
SM00175 Small GTPase superfamily, Rab type
SM00173 Small GTPase superfamily, Ras type
SM00233 Pleckstrin homology domain
SM00105 Arf GTPase activating protein
Superfamily do...
SSF50729
SSF48403 Ankyrin repeat-containing domain
SSF52540 P-loop containing nucleoside triphosphate hydrolase
SSF57863
14.46 kb
d
RefSeq Genes
Agap2
Agap2

Figure 35: *Agap2* differential exons usage

a. In the top graph *Agap2* exon usage in *Mll2*$^{-/-}$ and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ iNs. In the middle graph normalized read counts. In the bottom graph schematic of the differentially used exons. The statistically significant ones are in purple. *Mll2*$^{-/-}$ is in blue and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ in red. **b.** FPKM of *Agap2* in our RNA-seq. Means and SD are reported. *t test with a FDR <0.01. **c.** Ensembl schematic of the correspondence between exons and protein domains. **d.** *Agap2* Refseq genes from UCSC Genome browser. Assembly Mouse December 2011 GRCm38/mm10. In the purple rectangle the alternative spliced exon between the two isoforms.

Also *Grin2b* and *Pdcd5*, 2 of the top 30 differentially expressed genes (Figure 30), were subjected to alternative exon usage in the knock-out with respect to the control (*i.e.*, respectively exons 2, 3 and 4 and exons 3, 4 and 6) (Figures 36a and 36b respectively). Both genes have only one isoform annotated (respectively NM_008171 and NM_019746) (Figures 36c and 36d). In the case of *Grin2b* the exon 4 belongs to a known domain, the extracellular binding one (Figure 36e), while we were not able to determine the correspondence of the other exons with specific domains.

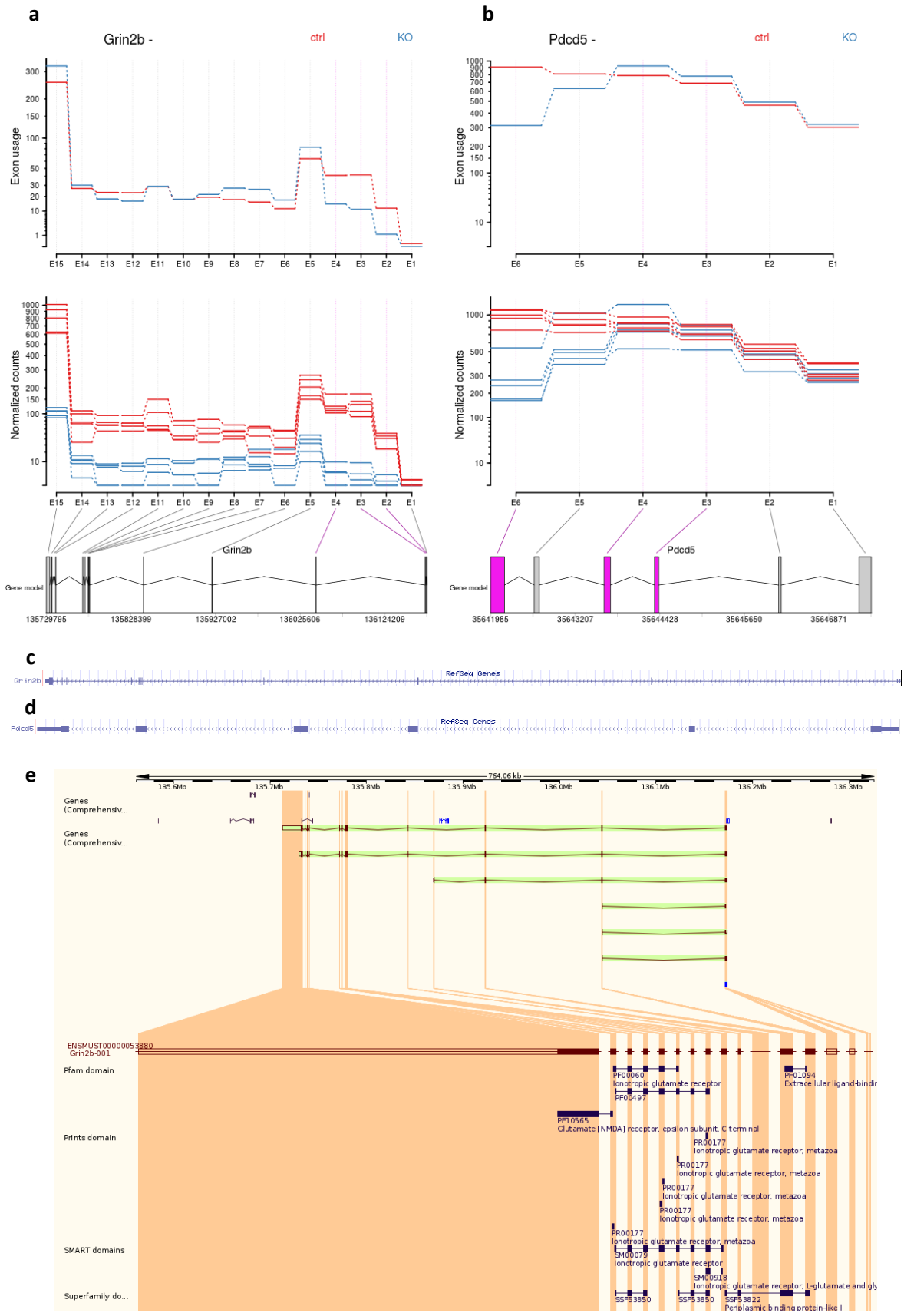
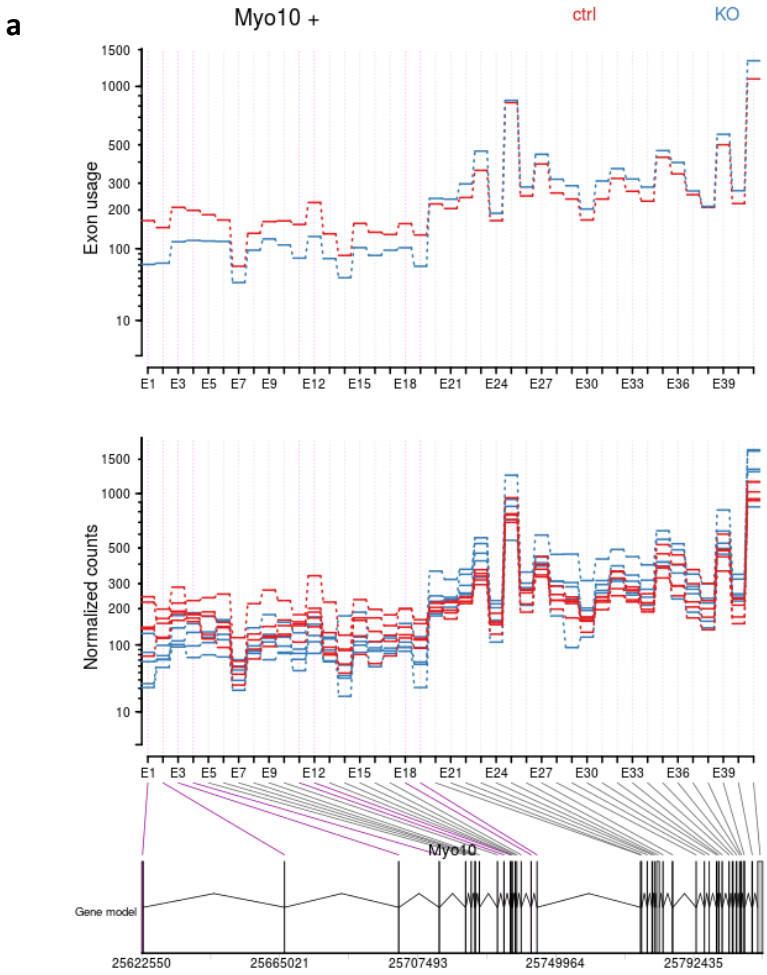


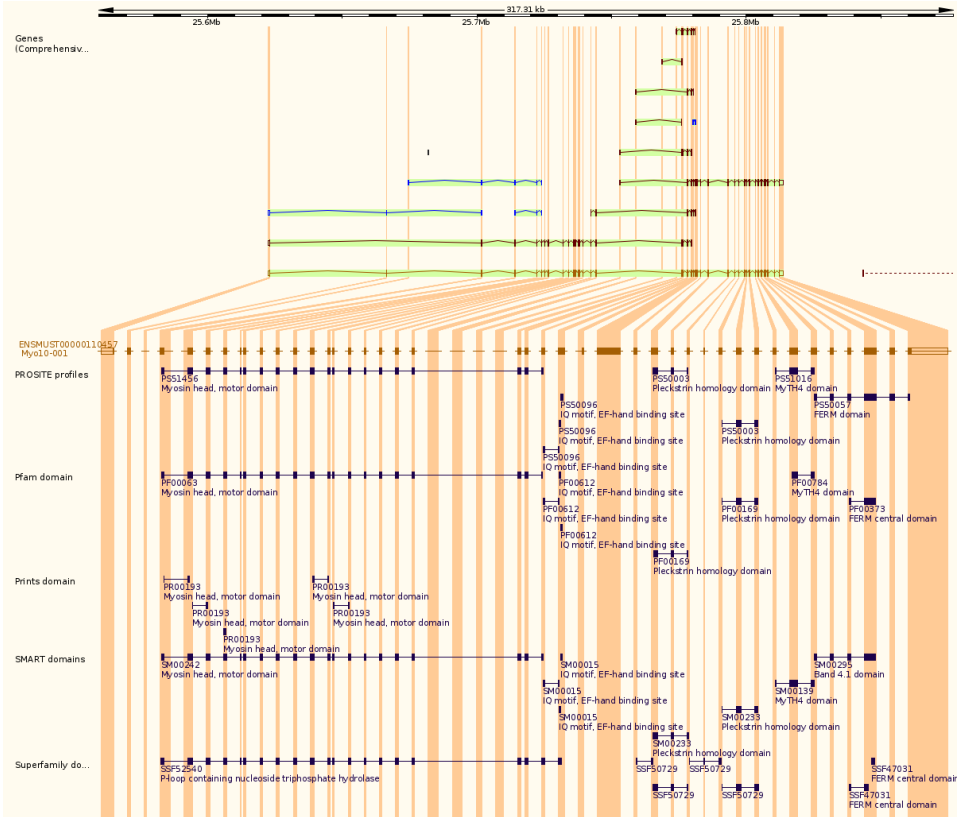
Figure 36: *Grin2b* and *Pdcd5* differential exons usage

a. and **b.** In the top graph *Grin2b* (a.) and *Pdcd5* (b.) exon usage in *Mll2*$^{-/-}$ and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ iNs. In the middle graph normalized read counts. In the bottom graph schematic of the differentially used exons. The statistically significant ones are in purple. *Mll2*$^{-/-}$ is in blue and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ in

red. **c.** and **d.** *Grin2b* and *Pdcd5,* respectively, Refseq genes from UCSC Genome browser. Assembly Mouse December 2011 GRCm38/mm10. **e.** Ensembl schematic of the correspondence between exons and protein domains of *Grin2b* gene.

Myosin 10 (*Myo10*) has only one validated isoform (NM_019472) (Figure 37c) and it has been recently associated with neuronal migration during development thanks to the interaction of its FERM domain with N-cadherin[172]. This domain remained unaffected, since the differential used exons were the 1, 2, 3, 4, 11, 12, 18, 19, all belonging to the myosin head motor domain (Figure 37a and 37b).


a
Myo10 +
ctrl
KO
Exon usage
1500
1000
500
300
200
100
10
E1 E3 E5 E7 E9 E12 E15 E18 E21 E24 E27 E30 E33 E36 E39
Normalized counts
Myo10
Gene model
25622550
25665021
25707493
25749964
25792435


b
317.31 kb
25.6Mb
25.7Mb
25.8Mb
Genes (Comprehensiv...
ENSMUST00000110457
Myo10-001
PROSITE profiles
PS51456 Myosin head, motor domain
PS50003 Pleckstrin homology domain
PS51016 MyTH4 domain
PS50057 FERM domain
PS50096 IQ motif, EF-hand binding site
Pfam domain
PF00063 Myosin head, motor domain
PF00612 IQ motif, EF-hand binding site
PF00784 MyTH4 domain
PF00169 Pleckstrin homology domain
PF00373 FERM central domain
Prints domain
PR00193 Myosin head, motor domain
SMART domains
SM00242 Myosin head, motor domain
SM00015 IQ motif, EF-hand binding site
SM00295 Band 4.1 domain
SM00139 MyTH4 domain
SM00233 Pleckstrin homology domain
Superfamily do...
SSF52540 P-loop containing nucleoside triphosphate hydrolase
SSF50729
SSF47031 FERM central domain

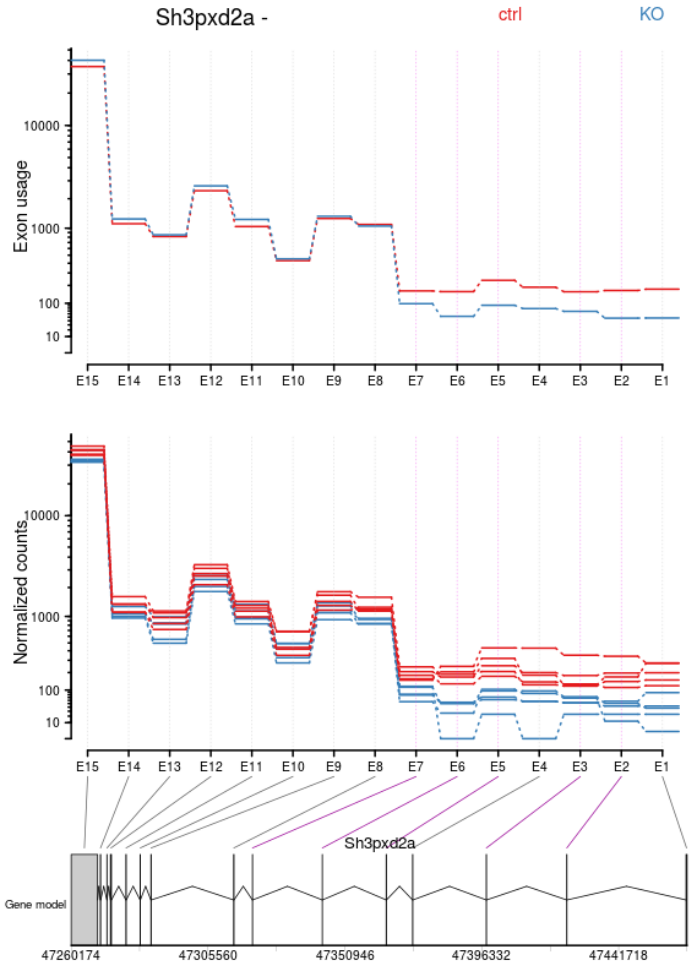

c
Myo10
RefSeq Genes

Figure 37: *Myo10* differential exons usage

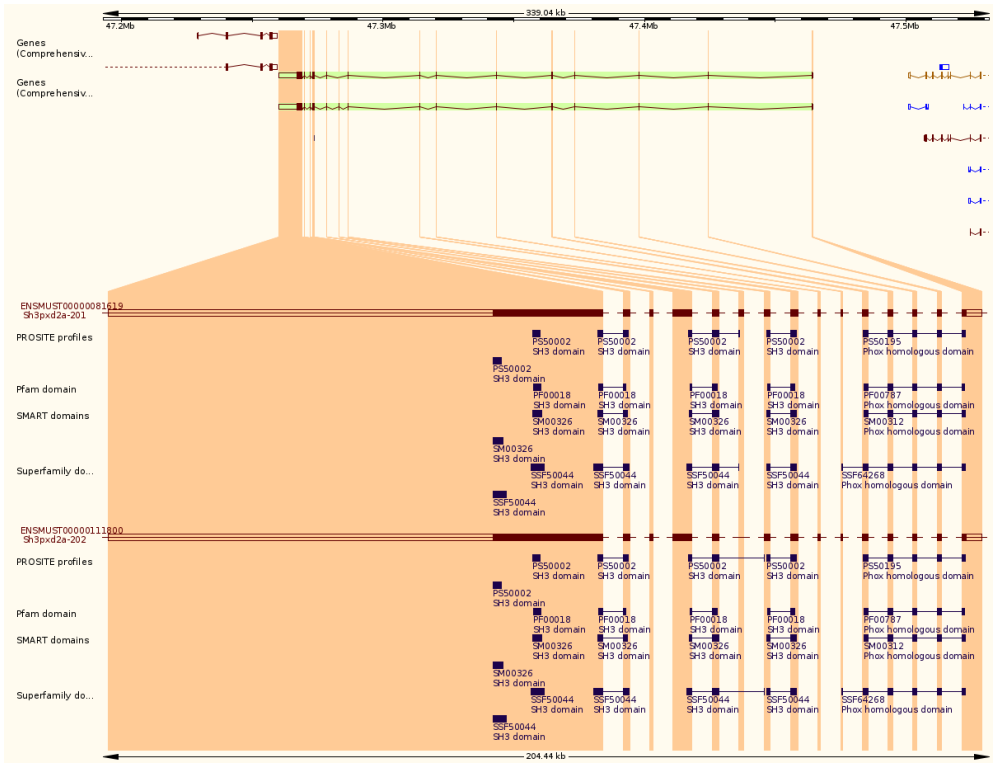
a. In the top graph *Myo10* exon usage in *Mll2*$^{-/-}$ and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ iNs. In the middle graph normalized read counts. In the bottom graph schematic of the differentially used exons. The statistically significant ones are in purple. *Mll2*$^{-/-}$ is in blue and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ in red. **b.** Ensembl schematic of the correspondence between exons and protein domains of *Myo10* gene. **c.** *Myo10* Refseq gene from UCSC Genome browser. Assembly Mouse December 2011 GRCm38/mm10.

The *SH3 and PX domain-containing protein 2A* (*sh3pxd2a*) gene encodes for a protein necessary for matrix digestion and therefore, for axon extension[173]. The differential used exons were 2, 3, 5, encoding for the phox homologous domain, which confers affinity to PI-3,4-bisphosphate, and exons 6 and 7, encoding for part of the SH3 domain (Figures 38a and 38b). In particular exon 6 is the spliced exon between the two validated isoforms (*i.e.*, NM_008018.4 long isoform 1; NM_001164717 short isoform 2) (Figure 38c). The isoform 1, which has been associated with a higher invasion rate of metastatic cells[174], was more expressed than the isoform 2 in our datasets, with no differences between the knock-out and the control (Figure 38d). Recently a new isoform was described, which loses the first 5 exons and has an alternative start codon upstream of exon 6, which is quite similar to the *Mll2*$^{-/-}$ *sh3pxd2a*[175], but with exon 6 and 7.

a


Sh3pxd2a -
ctrl
KO
Exon usage
10000
1000
100
10
E15 E14 E13 E12 E11 E10 E9 E8 E7 E6 E5 E4 E3 E2 E1
Normalized counts
10000
1000
100
10
E15 E14 E13 E12 E11 E10 E9 E8 E7 E6 E5 E4 E3 E2 E1
Sh3pxd2a
Gene model
47260174
47305560
47350946
47396332
47441718

b


339.04 kb
47.2Mb
47.3Mb
47.4Mb
47.5Mb
Genes (Comprehensiv...
Genes (Comprehensiv...
ENSMUST00000081619 Sh3pxd2a-201
PROSITE profiles
Pfam domain
SMART domains
Superfamily do...
ENSMUST00000111800 Sh3pxd2a-202
PROSITE profiles
Pfam domain
SMART domains
Superfamily do...
204.44 kb

c


Sh3pxd2a
Sh3pxd2a
RefSeq Genes

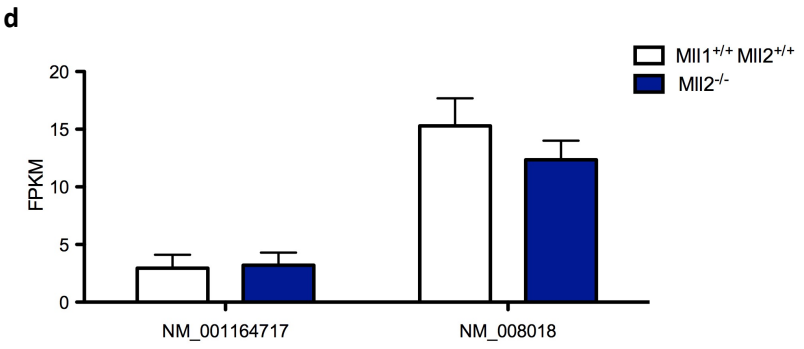


Figure 38: *Sh3pdx2a* differential exons usage

a. In the top graph *Sh3pdx2a* exon usage in *Mll2*$^{-/-}$ and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ iNs. In the middle graph normalized read counts. In the bottom graph schematic of the differentially used exons. The statistically significant ones are in purple. *Mll2*$^{-/-}$ is in blue and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ in red. **b.** Ensembl schematic of the correspondence between exons and protein domains of *Sh3pdx2a* gene. **c.** *Sh3pdx2a* Refseq genes from UCSC Genome browser. Assembly Mouse December 2011 GRCm38/mm10. In the purple rectangle the alternative spliced exon between the two isoforms. **d.** FPKM of the two *Sh3pdx2a* isoforms in our RNA-seq. Means and SD are reported.

The *SH3/ankyrin domain gene 3 (Shank3)* is a gene involved in post-synaptic organization and dendritic spine maturation. Its microdeletion and point mutation have been associated with autism. *Shank3* has a finely tuned post-transcriptional regulation and its alternative splicing, modulating the protein domains composition, could regulate its function and localization[176]. Despite this, only one isoform has been validated (*i.e.,* NM_021423) (Figure 39e). In our RNA-seq *Shank3* was down-regulated in *Mll2*$^{-/-}$ iNs (Figure 39b) that showed also a lower usage of exons 4-10 (ankyrin repeat-containing domain), 13-16 (PSD-95/Discs large/ZO-1 PDZ domain), 19, 21 and 22 (Pro and Sterile Alpha Motif (SAM) domains) (Figures 39a, 39c and 39d). The truncated form of *Shank3* (*i.e.*, without SAM and Pro domains) has been shown to preferentially localize in the nucleus and on the basis of domains composition, when overexpressed, it either increases or decreases the spine density and length[176]. Our isoform has not been described in literature, but we can hypothesize that, since *shank3* was both down-regulated and with only the SH3 domain, the post-synaptic compartment in *Mll2*$^{-/-}$ iNs could also be altered.

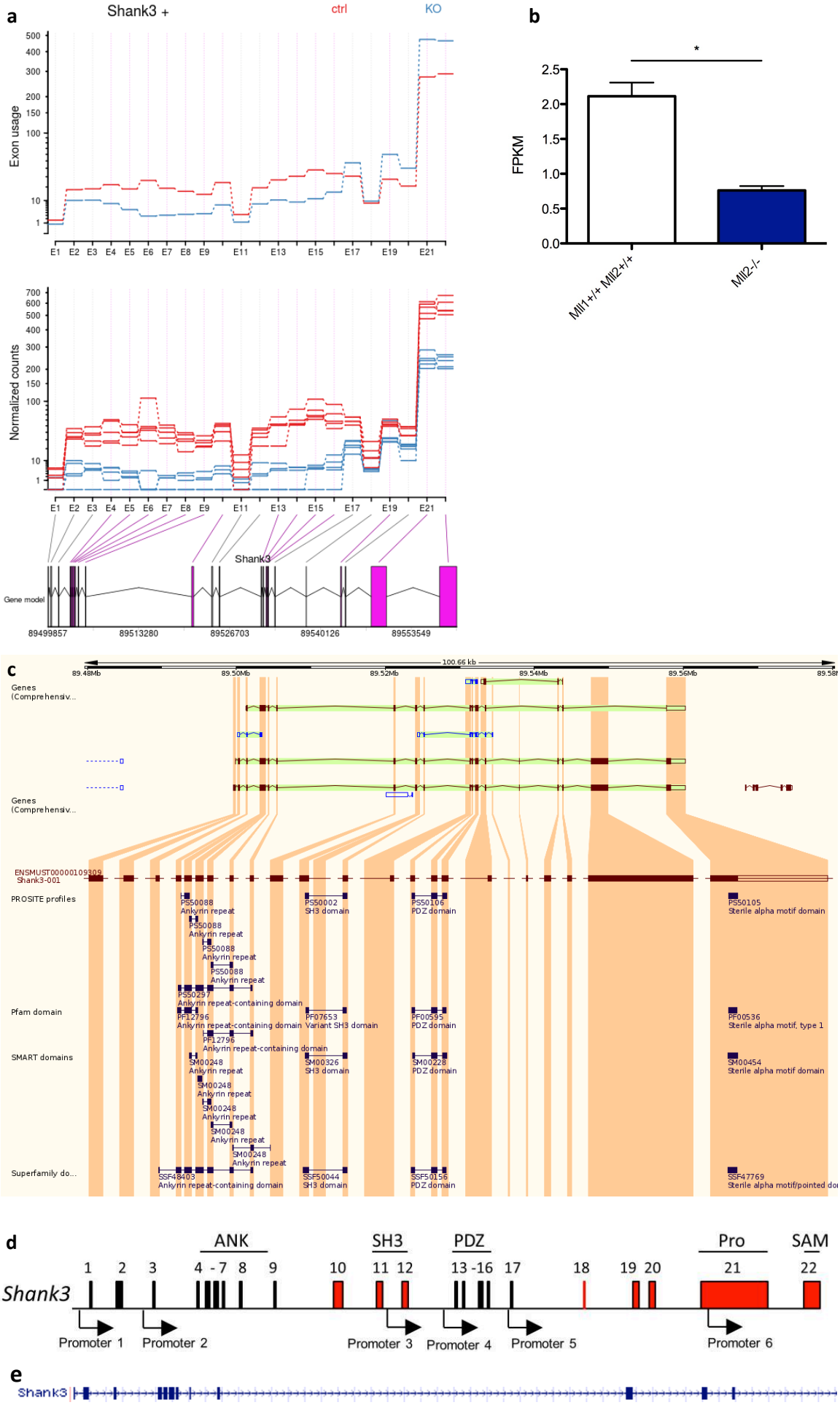

a
Shank3 +
ctrl
KO
Exon usage
Normalized counts
Shank3
Gene model
b
FPKM
Mll1+/+ Mll2+/+
Mll2-/-
c
Genes
ENSMUST00000109309
Shank3-001
PROSITE profiles
Pfam domain
SMART domains
Superfamily do...
d
Shank3
ANK
SH3
PDZ
Pro
SAM
Promoter 1
Promoter 2
Promoter 3
Promoter 4
Promoter 5
Promoter 6
e
Shank3

Figure 39: *Shank3* differential exons usage

a. In the top graph *Shank3* exon usage in *Mll2*$^{-/-}$ and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ iNs. In the middle graph normalized read counts. In the bottom graph schematic of the differentially used exons. The statistically significant ones are in purple. *Mll2*$^{-/-}$ is in blue and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ in red. **b.** FPKM of *Shank3* in our RNA-seq. Means and SD are reported. * FDR<0.01 **c.** Ensembl schematic of the correspondence between exons and protein domains of *Shank3* gene. **d.** Schematic of SHANK3 protein domains and promoters modified from Wang et al., Molecular Autism 2014[176]. In red the alternative spliced exons. **e.** *Shank3* Refseq gene from UCSC Genome browser. Assembly Mouse December 2011 GRCm38/mm10. In the purple rectangle the alternative spliced exon between the two isoforms.

The *X-linked inhibitor of apoptosis proteins* (*XIAP*) anti-apoptotic function is counteracted by the *X-linked IAP-associated factor-1* (*XAF1*), which shows a fine-tuned regulation, for example, during motoneurons apoptosis along development[177]. Two isoforms of *Xaf1* have been validated: a full length isoform (*i.e.*, isoform 1 NM_001037713) and a shorter one, without the exons 3 and 4 (*i.e.*, isoform 2 NM_001291153) (Figure 40b). Only the isoform 1 was detected in our datasets. However *Mll2*$^{-/-}$ *Xaf1* transcripts presented a lower usage of exon 4 and 5, preserving the exon 6, which was instead lost in *Mll2*$^{+/+}$ *Xaf1* transcripts (Figure 40a). The protein N-terminus part, which contains a zinc finger domain fundamental for the interaction with XIAP, was therefore maintained in both variants. It has been shown that a truncated form of *Xaf1* functions as dominant negative[178], however in our case, the exon 6 was preserved. Hence further investigations should be needed.

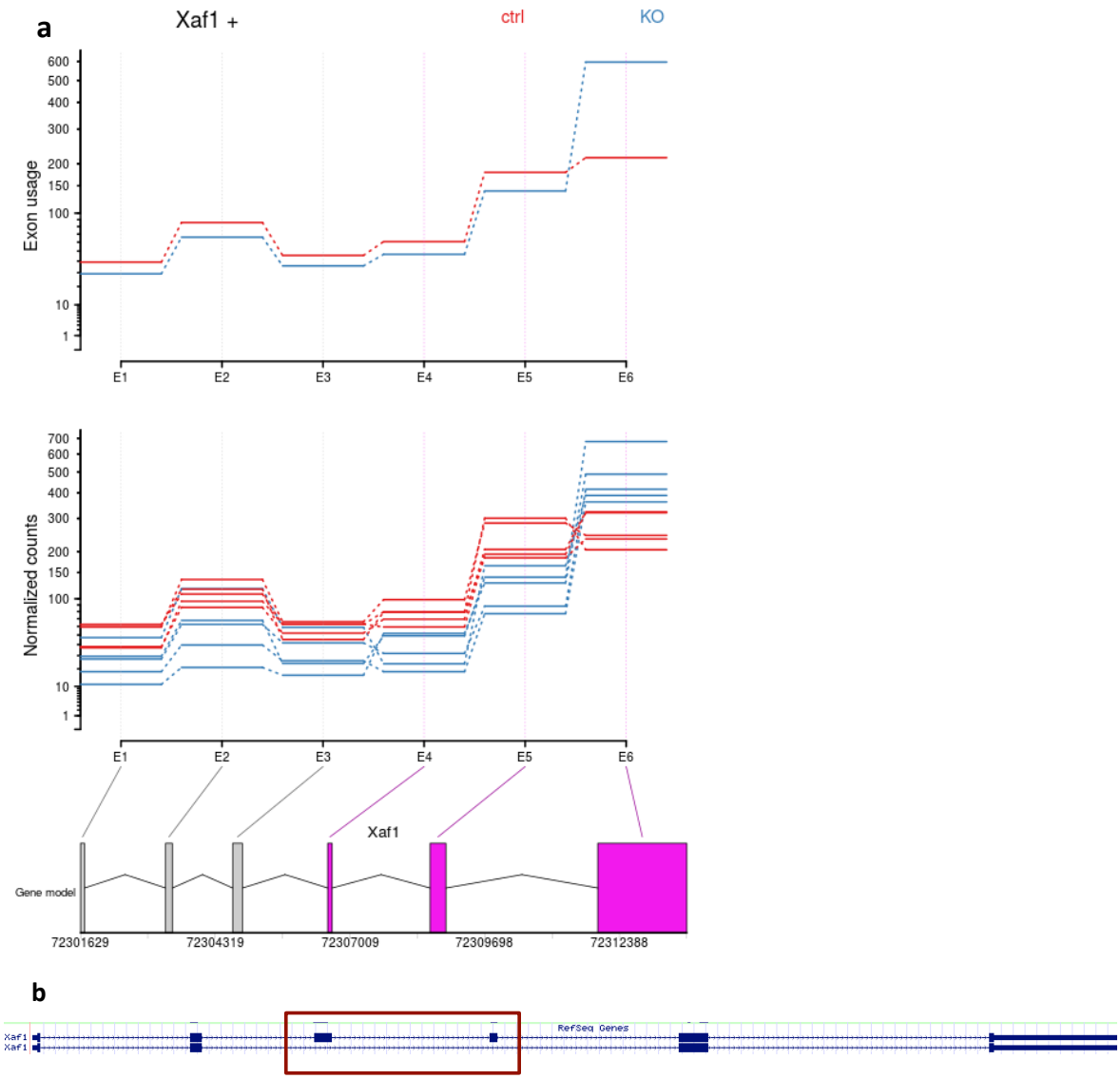


Figure 40: Xaf1 differential exons usage

a. In the top graph *Xaf1* exon usage in *Mll2*$^{-/-}$ and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ iNs. In the middle graph normalized read counts. In the bottom graph schematic of the differentially used exons. The statistically significant ones are in purple. *Mll2*$^{-/-}$ is in blue and *Mll1*$^{+/+}$*Mll2*$^{+/+}$ in red. **b.** *Xaf1* Refseq genes from UCSC Genome browser. Assembly Mouse December 2011 GRCm38/mm10. In the purple rectangle the alternative spliced exons between the two isoforms. In red the alternative spliced exons.

7.9 Identification of MLL2 direct targets

7.9.1 ESCs with EGFP-tagged MLL2

Since the antibody for MLL2 is not exploitable for ChIP, our first strategy was to import ESCs, through a collaboration with Francis Stewart's laboratory, with both copies of the endogenous *Mll2* exons 2 flanked by *loxP* sites and harbouring a BAC (that recapitulates the endogenous regulation of *Mll2*) carrying an EGFP tagged version of MLL2[76]. In particular EGFP has been inserted at the N-terminus of the taspase cleavage site (Figure 41a). This does not affect MLL2 activity and upon exon 2 deletion the MLL2 encoded by the BAC can rescue the defects due to the absence of the endogenous *Mll2*[76].

Therefore we needed a protocol to differentiate ESCs into either generic fibroblasts or, even better, into MEFs, the starting population of all our transdifferentiation experiments (Figure 41b).

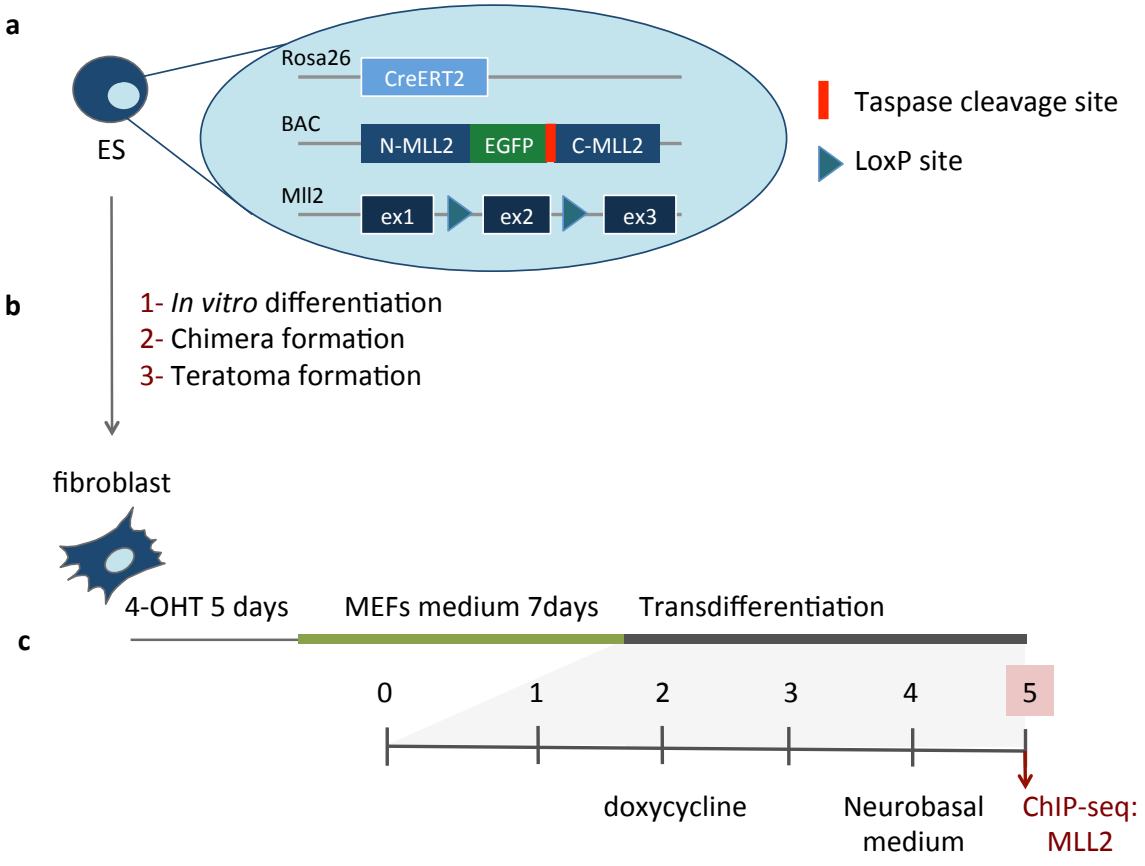


Figure 41: Identification of the MLL2 direct targets

a. Schematic of the imported EGFP-MLL2 ESCs**. b.** Approaches for the ESCs differentiation into fibroblasts. **c.** Once fibroblasts will be obtained they will be treated with 4-OHT and ChIP-seq will be performed at day 5 of transdifferentiation.

We could not find in literature mESCs *in vitro* differentiation protocols, most likely because mouse fibroblasts represent a highly accessible cell-type. Hence, we adapted a protocol used for human ESCs differentiation that entails the formation of embryoid bodies and several passaging steps of cells in trypsin[139]. Unfortunately, at the end of the protocol we were not able to obtain fibroblasts like-cells. Indeed both among control and tagged differentiated ESCs, there were cells positive for pluripotency markers and other uncharacterized cells (Figures 42a and 42b).

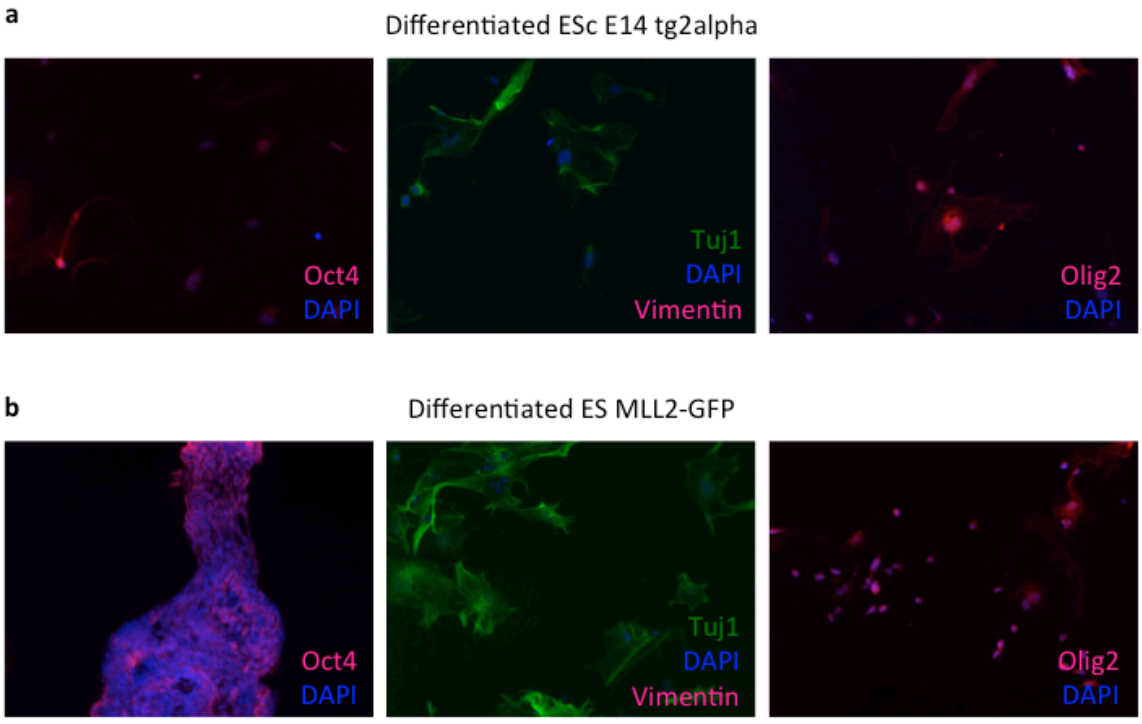


Figure 42: Mixed population by *in vitro* ESCs differentiation

Representative images of differentiated control ESCs (E14TG2alpha) (**a.**) and ESCs MLL2-EGFP tagged (**b.**) tested for the presence of stem cells (left panels, DAPI is in blue and Oct4 in red), of neurons and fibroblasts (middle panels, DAPI is in blue, Vimentin in red and Tuj1 in green) and of oligodendrocytes (right panels, DAPI is in blue and Olig2 in red).

Subsequently we tried two alternative approaches: i) we injected the tagged ESCs into blastocysts, with the purpose of deriving MEFs from chimeric embryos; and ii) we cultured ESC-derived teratoma cells to allow the ESCs first to differentiate *in vivo* into the three different germ layers and only later, when explanted, to be positively selected for fibroblasts.

We performed four rounds of blastocyst injections, then implanted in 8 pseudo-pregnant females, observing a very high rate of abortion (we recovered only 1 or 2 embryos per female with respect to the 4-5 commonly retrieved) (Figure 43a). This should not be attributed to the overexpression of *Mll2* because ESCs carry only 1 additional copy of the gene (Figure 43c). We assayed the rate of chimerism through the TaqMan assay for the *Mll2* exon 2, because the BAC encoded-EGFP is not visible at the microscope. When we analysed the recovered embryos they did not show more than 2 copies of *Mll2* (Figure 43d). Also after culturing the $EGFP^{+}$ MEFs (Figure 43b) with G418, exploiting the selection cassette present in the BAC, we could not detect any increase in the exon 2 copy number (Figure 43e). Therefore, the concentration of G418 we used was probably too low to select only the BAC-engineered MEFs. When I increased the level of G418 in the medium, most of the cells died, confirming the low level of chimerism. The few MEFs still alive, instead, senesced becoming unusable for high-throughput studies, which envisage high numbers of cells. Moreover, since senescence *per se* could affect transdifferentiation, these MEFs were useless.

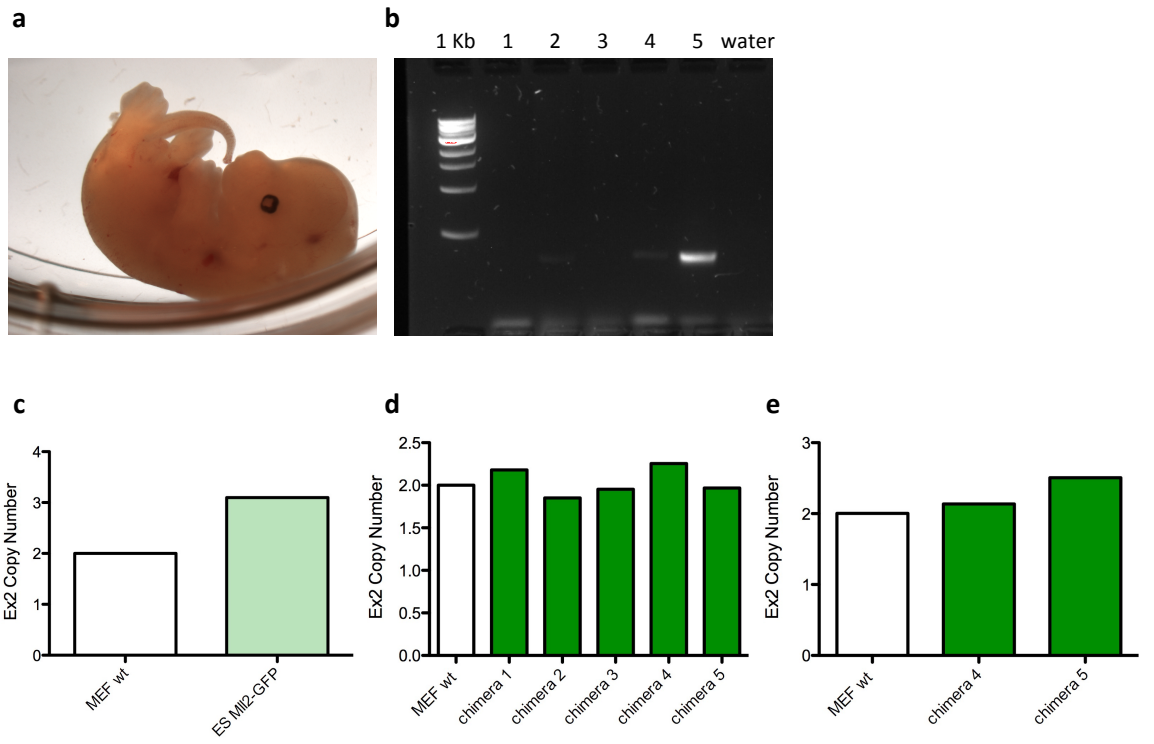
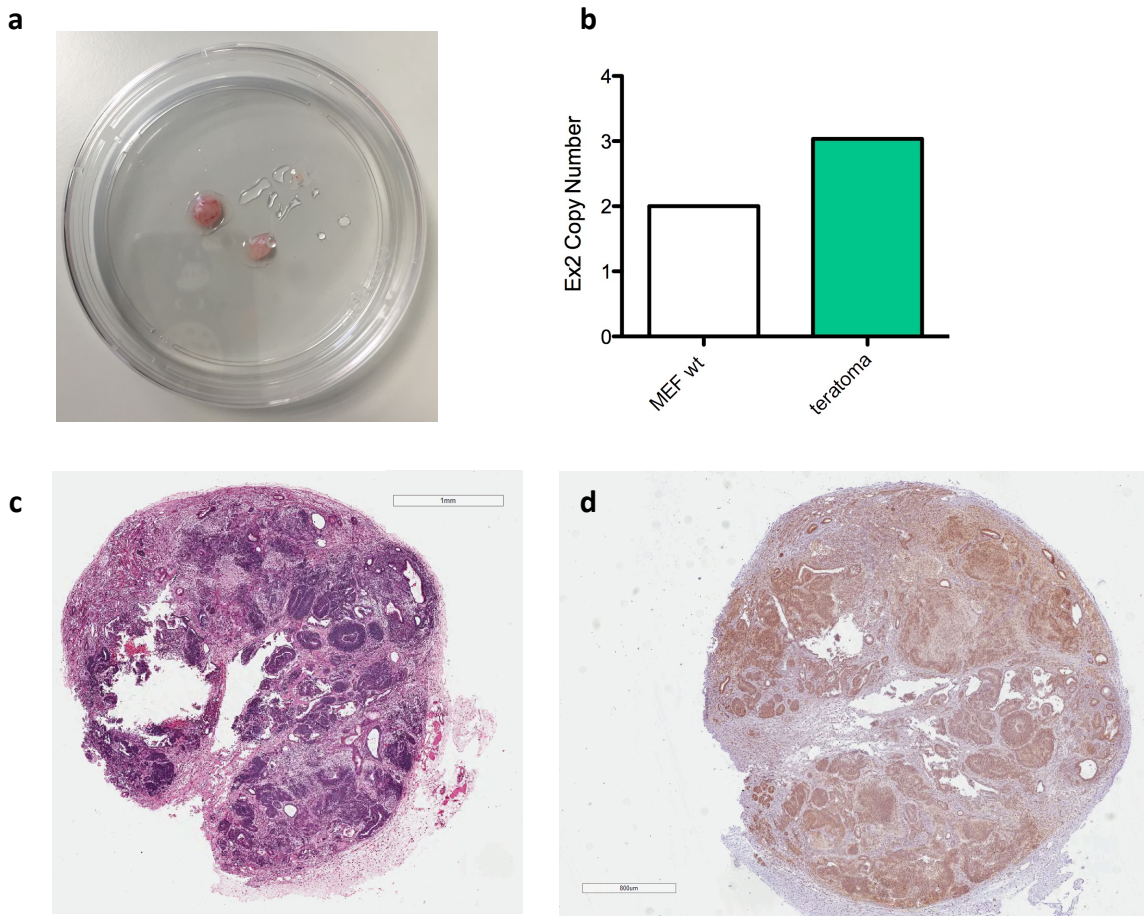


Figure 43: Chimeras formation

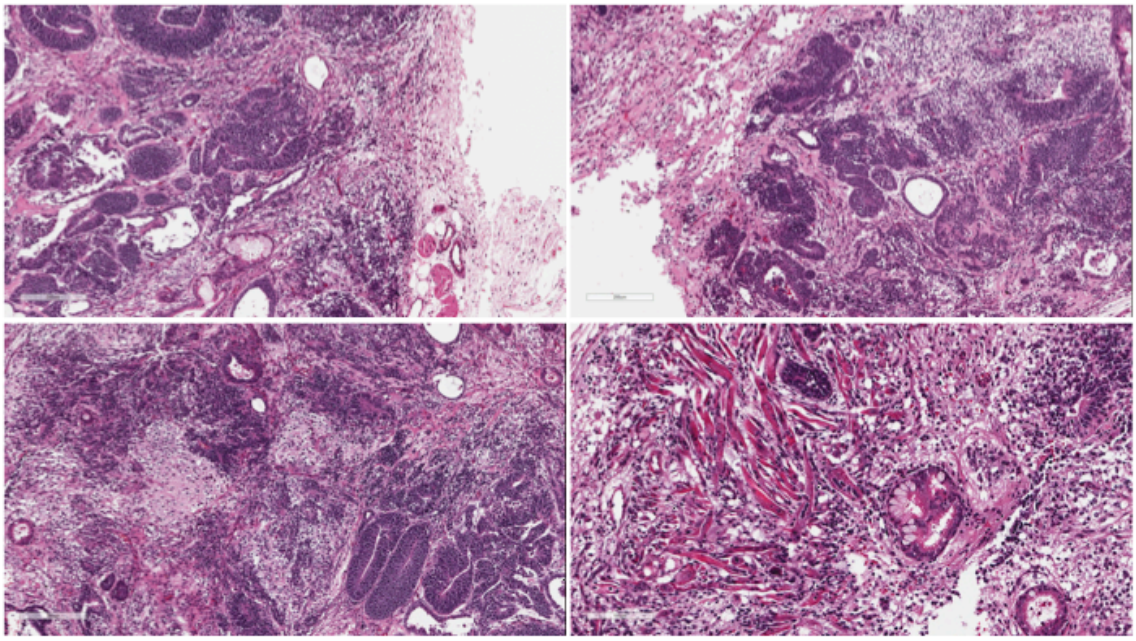
a. Representative picture of one retrieved embryo (left panel) **b.** Genotyping of EGFP of chimeric MEFs. **c.** Exon 2 *Mll2* copy number of ESCs MLL2-EGFP tagged assayed by TaqMan. **d.** Exon 2 *Mll2* copy number of chimeric MEFs at passage 0 assayed by TaqMan. **e.** Exon 2 *Mll2* copy number of chimeric MEFs number 4 and 5 after 2 weeks of G418 selection, assayed by TaqMan. Wt: wild type.

In the case of teratoma formation, we injected bilaterally and subcutaneously in 2 females NOD/SCID IL2Rγ$^{-/-}$ 10^6 ESCs previously maintained in culture with G418. After less than one month we collected 2 teratoma masses (Figure 44a). Part of them was sectioned and stained with haematoxylin/eosin to assess the trilineages specification that was indeed confirmed (Figures 44c and 44e), but some transformed areas were identified (Figure 44e). We also performed an immunohistochemistry staining for EGFP (Figure 44d). Interestingly we observed a widespread localization of EGFP (and therefore, indirectly, of MLL2) that was both nuclear and cytoplasmic (Figure 44f). We then cultured the teratoma-derived-cells in the MEFs medium, containing serum, to positively select only fibroblasts. Teratoma cells presented the same *Mll2* exon 2 copy number as the ESCs from which they derived (Figure 44b), but after 11 passages with trypsin we detected the appearance of

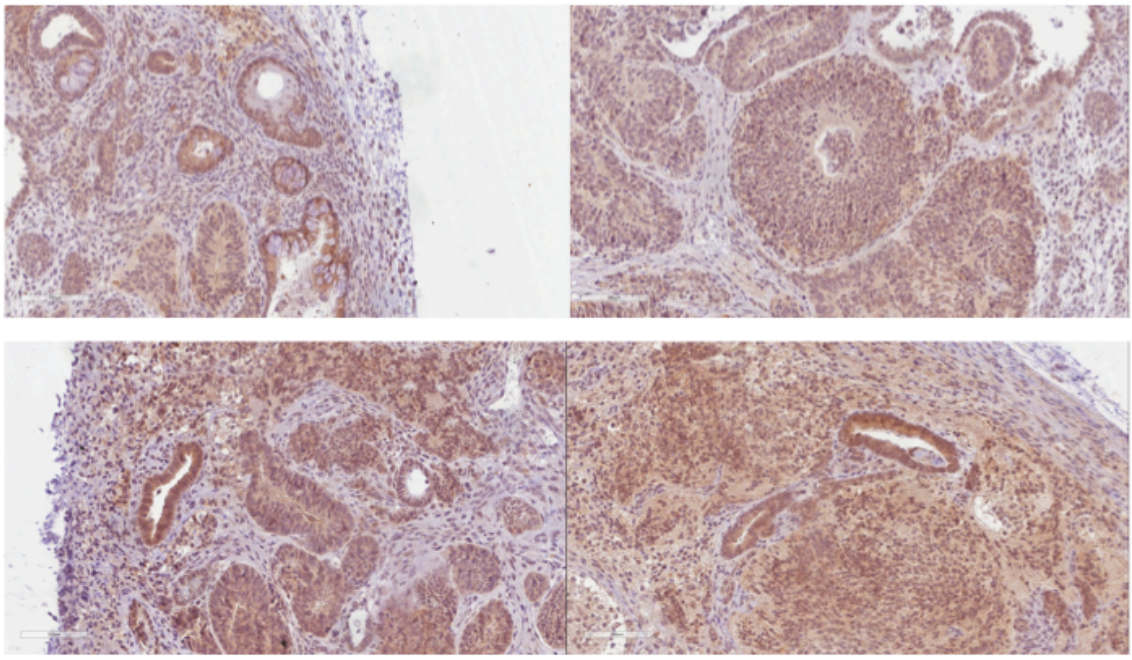
colonies (Figure 44g) that expressed pluripotency markers (Figure 44h), pointing to a probable transformation of the cells. Indeed cells appeared to have lost the contact inhibition and were growing faster with respect to the first passages.



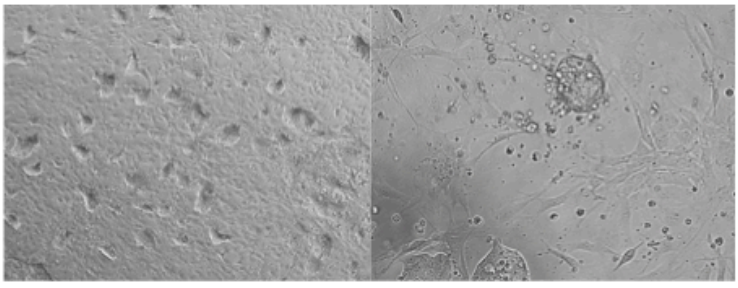
e



f



g



h

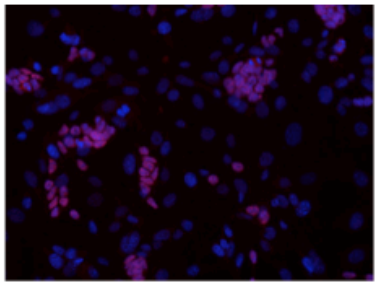


Figure 44: Teratoma formation

a. Teratoma masses explanted. **b.** Exon 2 *Mll2* copy number of teratoma cells and wild type fibroblasts assayed by TaqMan. **c**. and **e.** Haematoxylin/eosin staining of teratoma respectively in a low and a high magnification. **d**. and **f** Immunohistochemistry staining for EGFP of teratoma respectively in a low and a high magnification. **g.** Representative images of teratoma cell culture at passage 11 in a low (on the left) and a high (on the right) magnification. **h**. Representative image of teratoma cell culture stained for Oct4 (DAPI is in blue and Oct4 in red).

7.9.2 Identification of the MLL2 direct targets through MENIN

Since MLL2-EGFP tagged ESCs were not exploitable to assess MLL2 targets during transdifferentiation, we undertook a different strategy.

MENIN is the common subunit of only MLL1 and MLL2 TrxG proteins and it is fundamental for the positioning at their specific targets[43,55-57]. Moreover there are ChIP-grade commercially available anti-MENIN antibodies. Therefore we envisaged MENIN ChIP-seq, 5 days after cells were plated for transdifferentiation, in $Mll1^{+/+}Mll2^{+/+}$ and $Mll2^{-/-}$ transdifferentiating MEFs as alternative approach.

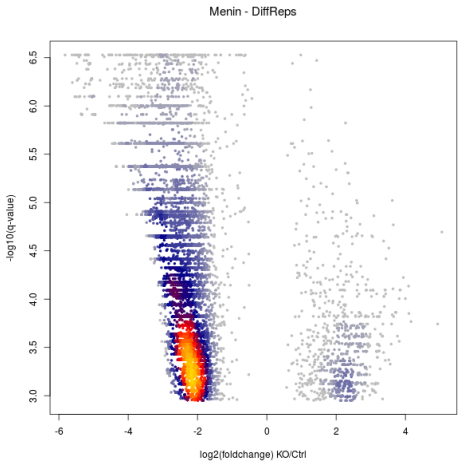
Through the immunoprecipitation of MENIN in $Mll1^{+/+}Mll2^{+/+}$ samples we could theoretically obtained an overview of the targets of both methylases at the peak of cell conversion. On the contrary, by performing MENIN ChIP-seq in $Mll2^{-/-}$ cells, we could identify the MLL2 specific targets, that are not compensated by MLL1 by comparing the ChIP peaks that are lost or reduced in the knock-out with respect to the control.

Since MENIN ChIP-seq was not very sensitive at the level of peak-calling, with a large variability in terms of number of peaks per sample, we decided to proceed with a quantitative, peak-call-agnostic method for the identification of differentially-enriched regions. This analysis identified a very large number of sites with reduced enrichment in the knock-out, along with a small number of up-regulated sites, most likely representing compensation by *Mll1* (Figure 45a). These sites overlapped with the promoters of, respectively, 839 and 213 genes. When we performed the GO analysis for the top most

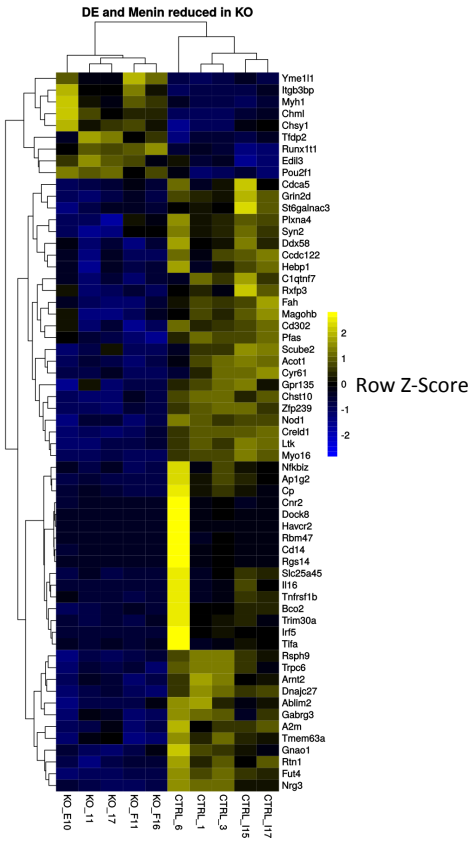
specific biological processes, we observed that the genes with a reduction in MENIN binding in *Mll2*$^{-/-}$ iNs were related to the basic metabolism and function of the cell (Figure 45d). On the contrary the genes that acquired MENIN binding in *Mll2*$^{-/-}$ iNs, were mostly connected to the acquisition of a different cell fate (Figure 45e). The subset of those genes that was differentially-expressed at day 13 is shown in Figure 45b.

In particular, *Magohb,* the best characterized direct MLL2 target so far[64] was both differentially MENIN-bound and lower expressed in the knock-out. In addition, relying on interactions observed in Hi-C data from mouse neural progenitors, we identified 98 additional target genes likely to be in contact with a region with reduced MENIN binding in the knock-out, including however only 3 genes that were differentially-expressed (shown in Figure 45c).

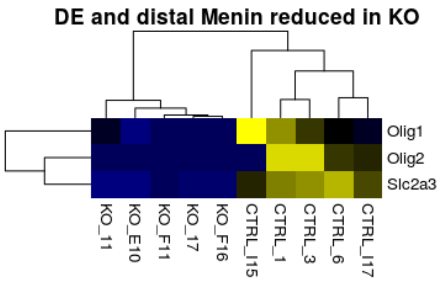
a


Menin - DiffReps
-log10(q-value)
log2(foldchange) KO/Ctrl

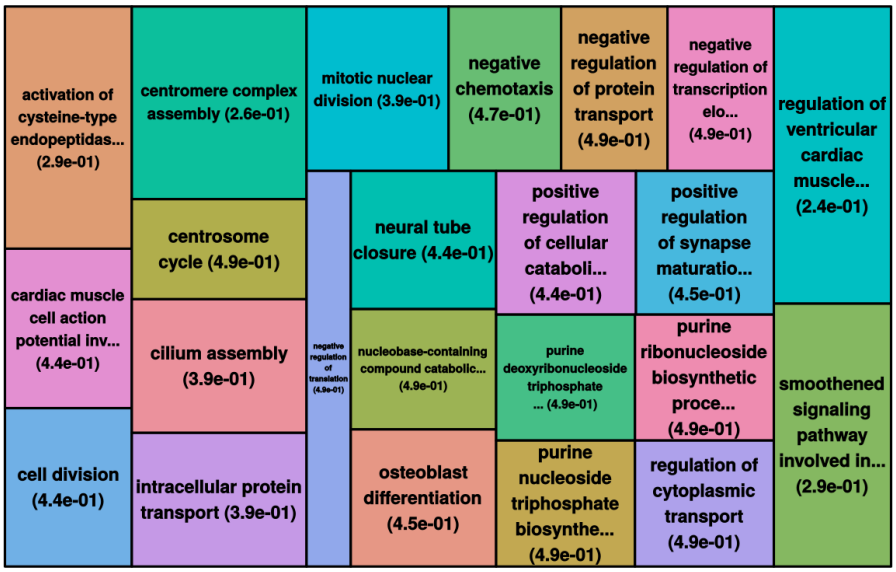
b


DE and Menin reduced in KO
Yme1l1
Itgb3bp
Myh1
Chml
Chsy1
Tfdp2
Runx1t1
Edil3
Pou2f1
Cdca5
Grin2d
St6galnac3
Plxna4
Syn2
Ddx58
Ccdc122
Hebp1
C1qtnf7
Rxfp3
Fah
Magohb
Cd302
Pfas
Scube2
Acot1
Cyr61
Gpr135
Chst10
Zfp239
Ncd1
Creld1
Ltk
Myo16
Nfkbiz
Ap1g2
Cp
Cnr2
Dock8
Havcr2
Rbm47
Cd14
Rgs14
Slc25a45
Il16
Tnfrsf1b
Bco2
Trim30a
Irf5
Tlla
Rsph9
Trpc6
Arnt2
Dnajc27
Ablim2
Gabrg3
A2m
Tmem63a
Gnao1
Rtn1
Fut4
Nrg3
KO_E10
KO_11
KO_17
KO_F11
KO_F16
CTRL_6
CTRL_1
CTRL_3
CTRL_I15
CTRL_I17
Row Z-Score

c


DE and distal Menin reduced in KO
Olig1
Olig2
Slc2a3
KO_11
KO_E10
KO_F11
KO_17
KO_F16
CTRL_I15
CTRL_1
CTRL_3
CTRL_6
CTRL_I17

d


activation of cysteine-type endopeptidas... (2.9e-01)
centromere complex assembly (2.6e-01)
mitotic nuclear division (3.9e-01)
negative chemotaxis (4.7e-01)
negative regulation of protein transport (4.9e-01)
negative regulation of transcription elo... (4.9e-01)
regulation of ventricular cardiac muscle... (2.4e-01)
centrosome cycle (4.9e-01)
neural tube closure (4.4e-01)
positive regulation of cellular cataboli... (4.4e-01)
positive regulation of synapse maturatio... (4.5e-01)
cardiac muscle cell action potential inv... (4.4e-01)
cilium assembly (3.9e-01)
negative regulation of translation (4.9e-01)
nucleobase-containing compound catabolic... (4.9e-01)
purine deoxyribonucleoside triphosphate ... (4.9e-01)
purine ribonucleoside biosynthetic proce... (4.9e-01)
smoothened signaling pathway involved in... (2.9e-01)
cell division (4.4e-01)
intracellular protein transport (3.9e-01)
osteoblast differentiation (4.5e-01)
purine nucleoside triphosphate biosynthe... (4.9e-01)
regulation of cytoplasmic transport (4.9e-01)

e

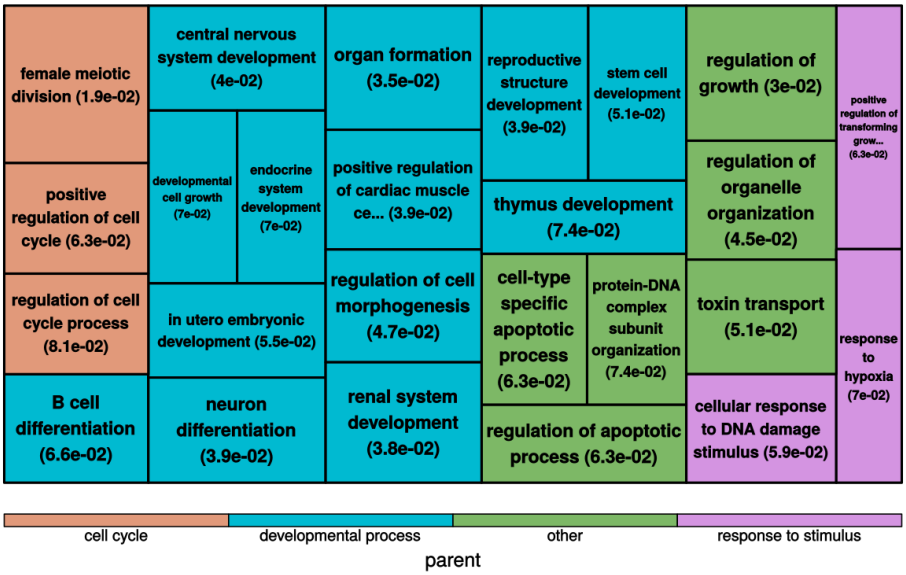


Figure 45: MENIN ChIP-seq

a. Volcano plot of the differentially bound MENIN regions, in *Mll2*$^{-/-}$ with respect to control. **b.** Heatmap of the differentially bound MENIN genes (proximal) that were also differentially expressed (FDR <0.01 and a FC > 0.5). The colour scale (-2 blue to +2 yellow) represents Z-Score. **c.** Heatmap of the differentially bound MENIN genes (distal) that were also differentially expressed (FDR <0.01 and a FC > 0.5). **d.** and **e.** Quilts of the top most enriched Biological Process, respectively for the genes with a lower and a higher MENIN binding in *Mll2*$^{-/-}$ transdifferentiating MEFs.

7.9.3 H3K4 MLL2-deposited methylation during transdifferentiation

To identify the fundamental targets that need to be methylated to drive transdifferentiation, we performed H3K4me3 ChIP-seq 5 and 13 days after *Mll1*$^{+/+}$*Mll2*$^{+/+}$ and *Mll2*$^{-/-}$ MEFs were plated for transdifferentiation. In particular, in the 5 days H3K4me3 ChIP-seq, we included also one *Mll1*$^{-/-}$ transdifferentiating MEFs sample.

The list of samples analysed is reported in Table 7.

Mll2+/+ 5 days	Mll2-/- 5 days	Mll1-/- 5 days
L6_K4	M6_K4	Mll1_K4
L7_K4	M7_K4	

Mll2+/+ 13 days	Mll2-/- 13 days
L6_13K4	M6_13K4
L7_13K4	M7_13K4

Table 7: List of samples analysed by H3K4me3 ChIP-seq

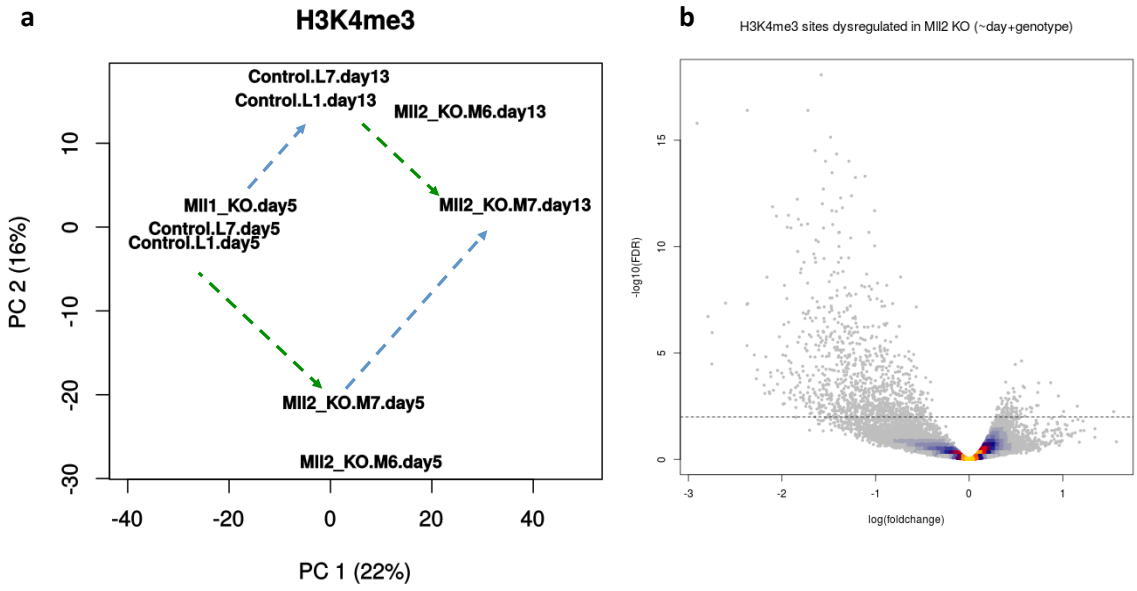
L6 and L7 transdifferentiating MEFs derived from different embryos, such as M6 and M7.

First we performed a Principal Component Analysis (PCA) on the logarithmic normalized read counts across the union of H3K4me3 sites (Figure 46a). Interestingly, and as expected by the previously demonstrated lack of impact of the absence of MLL1 on transdifferentiation, *Mll1*$^{-/-}$ transdifferentiating MEFs clustered together with the controls (*i.e.,* L6 and L7). Hence, either MLL1 is dispensable during MEF-to-iNs cell conversion or MLL2 can compensate for its absence, trimethylating MLL1-targets. This was in line with the published role of MLL1 and MLL2 during retinoic acid-driven ESCs differentiation into neurons[76]. Indeed it was demonstrated that it is MLL2 the main H3K4 trimethylase at bivalent promoters and that in *Mll1*$^{-/-}$ ESCs, MLL2 can compensate, trimethylating MLL1 targets. On the other side, contrarily from what has been shown for SVZ NSC differentiation[123], MLL1 during transdifferentiation is dispensable for the neuronal specification. Noteworthy a reduction in H3K4me3 at specific targets was not observed in *Mll1*$^{-/-}$ SVZ NSC pointing, also in this case, to a possible compensation by MLL2.

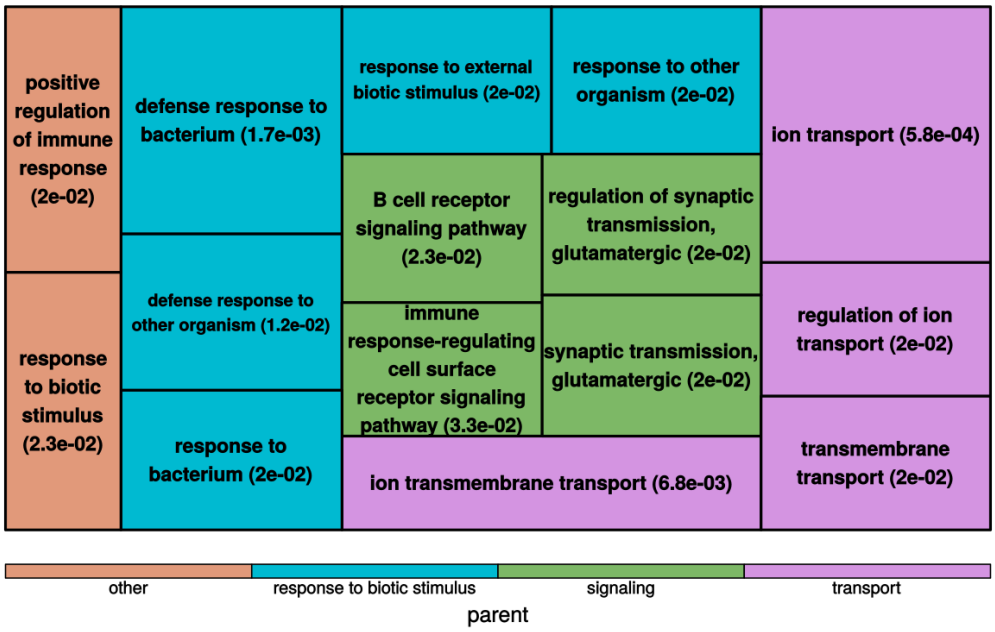
Afterwards, if two lines were drawn between the *Mll2*$^{-/-}$ and the controls and between the day 5 and the day 13, they would be almost orthogonal. Moreover the displacement between controls and knock-out was parallel between day 5 and day 13 (Figure 46a, green arrows), as the displacement from day 5 to day 13 between controls and knock-out (Figure 46a, light blue arrows). This suggested that the global mis-deposition of H3K4me3 in *Mll2*$^{-/-}$ transdifferentiating MEFs was maintained, and mostly unchanged in *Mll2*$^{-/-}$ iNs.

For this reason, first we executed an analysis considering all samples together and adjusting for the differences linked to time points. Among 15405 sites marked by H3K4me3 in at least one sample, 1089 were statistically differentially marked in *Mll2*$^{-/-}$ and controls (FDR<0.01) (Figure 46b). In particular 994 sites, mapping at TSS of 545 genes, showed a decreased H3K4 trimethylation, while 95 sites, mapping at the TSS of 82 genes, showed an increase in H3K4 trimethylation. Among the 82 genes only *Runt-related transcription factor 1 translocated to 1* (*Runx1t1*) was differentially expressed and, accordingly to its increased H3K4 trimethylation, was up-regulated in *Mll2*$^{-/-}$ with respect to controls (Figure 46e). *Runx1t1* is a transcriptional co-repressor that recently has been implicated in RGs differentiation. However, if overexpressed it favours neuronal differentiation[179].

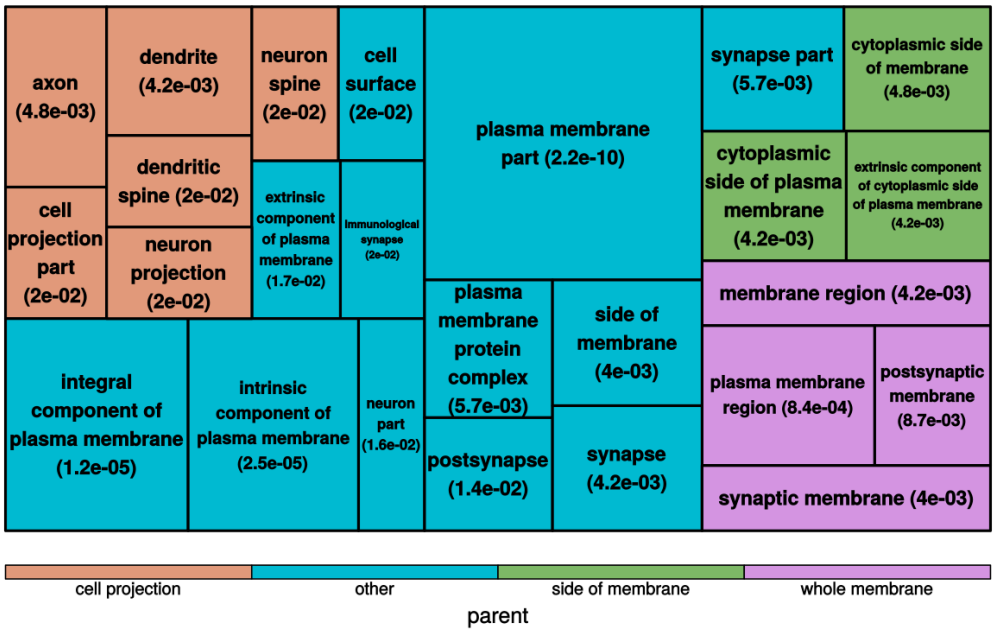
Of the 545 genes that showed a decrease in H3K4 trimethylation, 184 were differentially expressed at FDR<0.05 and 143 at FDR<0.01. In particular if we performed GO enrichments analysis on them, we observed, among the biological process, synaptic transmission and membrane transport (Figure 46c). Among the cellular components we could find especially membranes, synapses and dendrite, axon and neuron projections (Figure 46d).



c


positive regulation of immune response (2e-02)
defense response to bacterium (1.7e-03)
response to external biotic stimulus (2e-02)
response to other organism (2e-02)
ion transport (5.8e-04)
B cell receptor signaling pathway (2.3e-02)
regulation of synaptic transmission, glutamatergic (2e-02)
defense response to other organism (1.2e-02)
regulation of ion transport (2e-02)
response to biotic stimulus (2.3e-02)
immune response-regulating cell surface receptor signaling pathway (3.3e-02)
synaptic transmission, glutamatergic (2e-02)
response to bacterium (2e-02)
transmembrane transport (2e-02)
ion transmembrane transport (6.8e-03)
other
response to biotic stimulus
signaling
transport
parent

d


axon (4.8e-03)
dendrite (4.2e-03)
neuron spine (2e-02)
cell surface (2e-02)
plasma membrane part (2.2e-10)
synapse part (5.7e-03)
cytoplasmic side of membrane (4.8e-03)
dendritic spine (2e-02)
extrinsic component of plasma membrane (1.7e-02)
Immunological synapse (2e-02)
cytoplasmic side of plasma membrane (4.2e-03)
extrinsic component of cytoplasmic side of plasma membrane (4.2e-03)
cell projection part (2e-02)
neuron projection (2e-02)
membrane region (4.2e-03)
plasma membrane protein complex (5.7e-03)
side of membrane (4e-03)
integral component of plasma membrane (1.2e-05)
intrinsic component of plasma membrane (2.5e-05)
neuron part (1.6e-02)
plasma membrane region (8.4e-04)
postsynaptic membrane (8.7e-03)
postsynapse (1.4e-02)
synapse (4.2e-03)
synaptic membrane (4e-03)
cell projection
other
side of membrane
whole membrane
parent

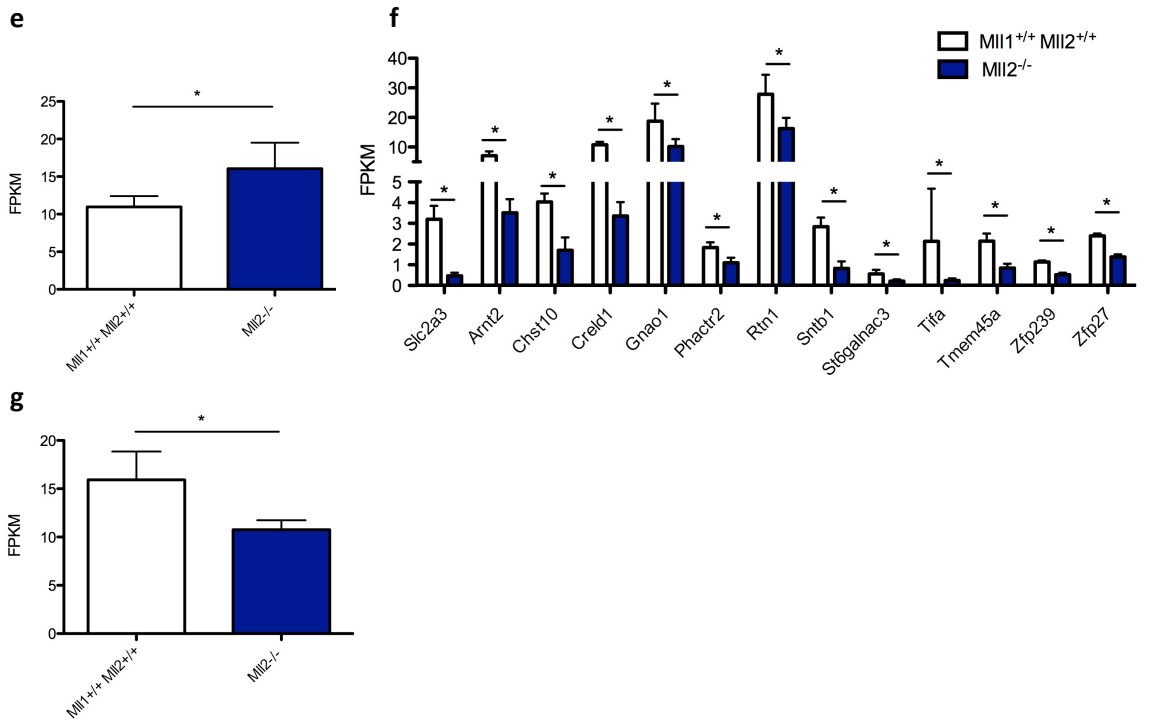


Figure 46: H3K4me3 ChIP-seq

a. PCA on the logarithmic normalized read counts across the union of H3K4me3 sites. Dashed green arrows represent the displacement between controls and knock-out at day 5 and day 13. Dashed light blue arrows represent the displacement between day 5 and day 13 in knock-out and controls. **b.** Volcano plot of the differentially H3K4 trimethylated regions. **c.** Quilt of the gene ontology enrichment analysis, regarding the biological processes, of the genes both differentially H3K4 trimethylated and expressed (FDR<0.01 and a FC > 0.5). **d.** Quilt of the gene ontology enrichment analysis, regarding the cellular components, of the genes both differentially H3K4 trimethylated and expressed (FDR<0.01 and a FC > 0.5). **e. f.** and **g.** FPKM, in our RNA-seq, respectively of *Runx1t1,* of the genes that showed a decrease in H3K4me3 and were differentially expressed and of *Tuba4a*. Means and SD are reported * FDR<0.01

Of the 143 genes, 13, according to MENIN ChIP-seq, can be considered direct targets of MLL2: *Slc2a3*, *Arnt2*, *Chst10*, *Creld1*, *Gnao1*, *Phactr2*, *Rtn1*, *Sntb1*, *St6galnac3*, *Tifa*, *Tmem45a*, *Zfp239* and *Zfp27*. Moreover, in line to their reduced H3K4 trimethylation, they were down-regulated in *Mll2*$^{-/-}$ with respect to controls (Figure 46f).

The *solute carrier family 2 member 3* (*Slc2a3*) or *Glut-3* is a brain specific glucose transporter, while the *guanine nucleotide binding protein, alpha O* (*Gnao1*) is a brain-specific G protein coupled receptor associated to epilepsy[180].

The *Aryl hydrocarbon receptor nuclear translocator 2* (*Arnt2*) has been implicated in zebrafish and mouse brain development. In particular it has been shown to be bivalently marked in ESCs and to be induced, upon demethylation of H3K27, during retinoic acid induced-ESC differentiation[181].

The *carbohydrate sulfotransferase 10* (*Chst10*) is an enzyme, which synthesised the neuronal specific sulfoglucuronyl carbohydrate, involved, among the others, also in neurite outgrowth[182].

The *phosphatase and actin regulator 2* (*Phactr2*) is a brain specific protein induced upon injury, during neurogenesis[183].

The *reticulon 1* (*Rtn1*) has three isoforms: *Rtn1-a*, *Rtn1-c* and *Rtn1-c1*. In our datasets only *Rtn1-a* and *Rtn1-c* were expressed and *Rtn1-a* more than *Rtn1-c*. Isoform a has been associated with neuronal excitation: it encodes for an endoplasmic reticulum membrane protein coupled with the ryanodine receptor 2, which modulates calcium oscillation[184]. Isoform c, instead, modulates histone deacetylases activity[185].

Zfp239 and *Zfp27* are poorly characterized, but because of their ability to bind DNA, and RNA in case of Zfp239[186], they are two interesting candidates as MLL2 mediators.

All the other genes are poorly described in literature or without a clear connection with neuronal differentiation.

Finally, we looked also at the differentially expressed genes that were bound by MENIN in the control and without any peak in the knock-out and differentially H3K4me3. The only gene which came out from this analysis was the *tubulin alpha 4 a* (*tuba4a*) (Figure 46g), a specific tubulin with its highest expression in brain and whose mutation is associated to familial amyotrophic lateral sclerosis[187].

Therefore MLL2 emerges from these data as having a key role for the modulation of genes implied in the acquisition of neuronal functions and metabolism. However few genes came out from the integrative analysis of the H3K4me3 ChIP-seq, the MENIN ChIP-seq and the RNA-seq, probably because the RNA-seq and MENIN ChIP-seq were performed on

different time points. Therefore through the RNA-seq at 13 days we could observe only the downstream effectors of MLL2 and not its direct targets.

Next, we thus looked at the genes differentially bound by MENIN and differentially H3K4 trimethylated, without considering their expression level (Figure 47a). Indeed, although changes in MENIN and H3K4me3 enrichments did not always lead to the corresponding change in gene expression, the majority of the genes showed reduced MENIN binding and had a lower level of H3K4me3 in *Mll2*$^{-/-}$ cells. These genes could be considered as the direct targets of MLL2, not compensated by MLL1.

Among the most interesting ones there was *Zfp277*, a transcriptional repressor that through the interaction with the PRC1 complex component BMI1 (up-regulated in *Mll2*$^{-/-}$ iNs), is involved in the *Ink4a/Arf* locus silencing[188]. Of note, *Bmi1* itself and *Chromobox 7* (*Cbx7)* were also differentially-expressed in *Mll2*$^{-/-}$ iNs (Figure 47b), and the differentially-expressed genes were enriched for targets of PRC2 (~2-fold enrichment, FDR ~4e-75) in mESCs, suggesting that the Polycomb axis of regulation might be responsible for part of the dysregulation caused by *Mll2* depletion during transdifferentiation. In addition, the *patched homolog 1* (*Ptch1*), the receptor of the sonic hedgehog (sh) ligand, is involved in the Sonic Hedgehog pathway that is fundamental, *in vivo*, for axon guidance[189]. Moreover the *dedicator of cytokinesis 4* (*Dock4*), a guanine nucleotide exchange factor for Rac1, has been shown to regulate axon-dendrite polarity and dendrites arborisation[190,191]. A truncated mutation of *Dock4* affects neurite elongation[192]. Interestingly, also *Mll1*$^{-/-}$ transdifferentiating MEFs showed a reduced H3K4 trimethylation at day 5, pointing to the necessity of the presence of both methylases for its trimethylation. The *neuralized homolog 1 a* (*Neurl1a*) is an E3 ubiquitin ligases that in hippocampal neurons activates CPEB3, whose in turn induces *glutamate A1* (*GluA1)* and *GluA2* (*GluA2* was down-regulated in our dataset Figure 47c), important for synapse plasticity[193]. The *Drosophila* homolog is fundamental for neurogenesis. Indeed, in the fly Notch signal receiving cells, during lateral

inhibition process, neuralized ubiquitinates the Delta ligand causing its internalization and consequently its degradation[194,195].

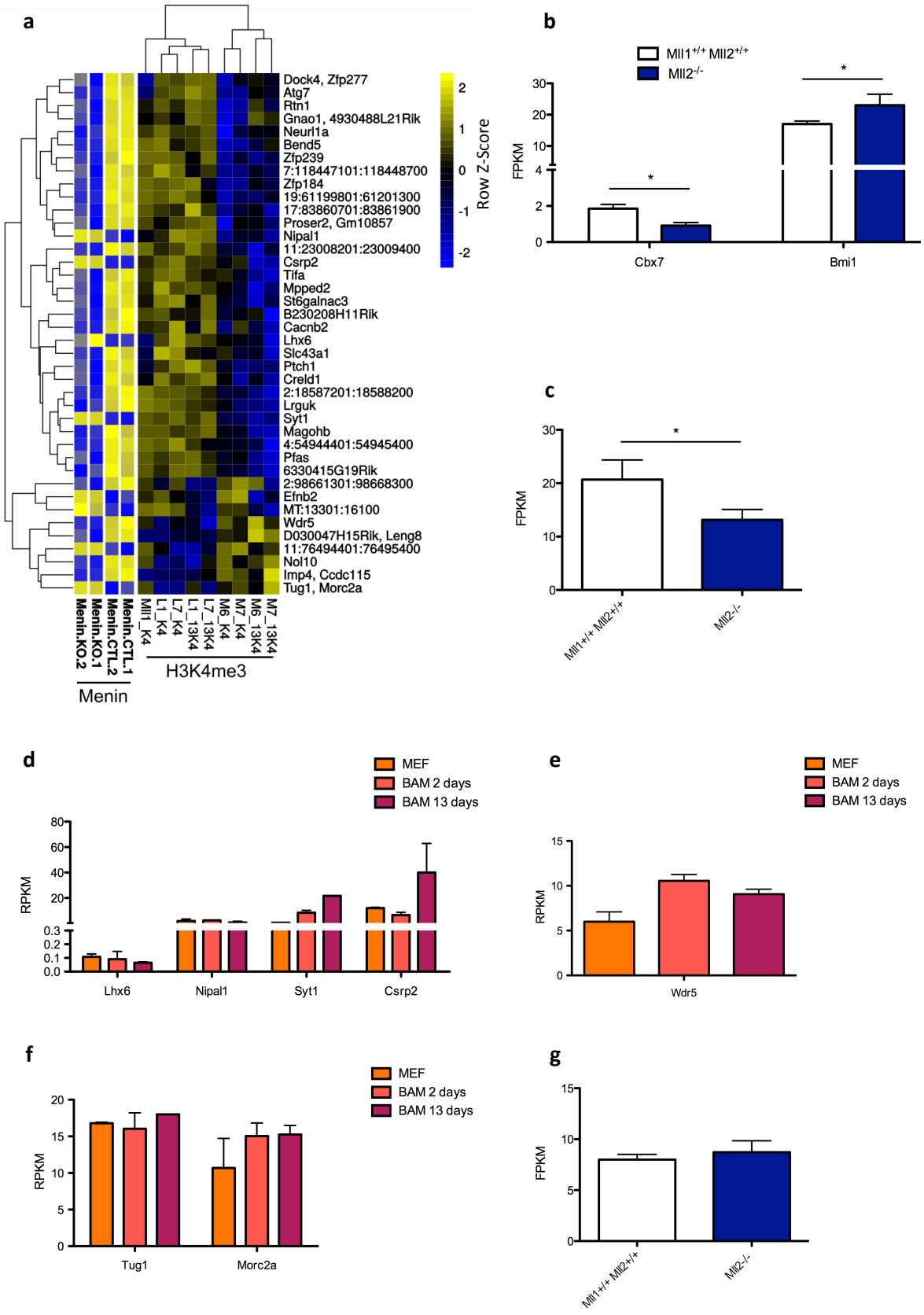
A second category of genes was not bound by MENIN, but H3K4 trimethylated in *Mll2*$^{+/+}$ cells and bound by MENIN, but not H3K4 trimethylated in *Mll2*$^{-/-}$ cells. We subdivided these genes among i) the ones that are stably expressed throughout transdifferentiation (*Nipal1* and *Lhx6*), that therefore are already trimethylated at the time point analysed, ii) the ones that are activated earlier (2 days after BAM transduction (*Syt1*) in which, most likely the trimethylation, and therefore the MLL2 binding, occurred earlier than the 5 days-time point analysed and iii) the one whose expression is independent from H3K4me3 (*Csrp2*) (Figure 47d). The hypothesis for this category of genes is that either the H3K4me3 was deposited by MENIN-containing complexes earlier with respect to the time point analysed, and therefore MENIN is no more detectable through ChIP, or that the H3K4me3 deposition at these genes was performed by another TrxG member. In *Mll2*$^{-/-}$ cells a mechanism of compensation probably led to MENIN mis-binding, which however failed in H3K4 trimethylation, not present even at 13 days. Alternatively MENIN in the knock-out could function as a repressor, interacting, for example, with the Suppressor of variegation 3-9 homolog 1 (Suv39h1) for the deposition of H3K9 methylation[196]. However, among them, only *Syt1* was down-regulated in our dataset, a gene fundamental during transdifferentiation, further suggesting a more probable mislocalization of MENIN (Figure 28).

A third category was constituted by the genes bound by MENIN, but not H3K4 trimethylated in *Mll2*$^{+/+}$ cells and not bound by MENIN, but H3K4 trimethylated in *Mll2*$^{-/-}$ cells (*Nol10*, *Imp4*, *Ccdc115*, *Wdr5* and *Leng8*). Surprisingly, for this class of genes the MLL2-MENIN binding seemed to prevent the H3K4me3 deposition, that was instead acquired in *Mll2*$^{-/-}$ iNs, where the MENIN binding was absent. Interestingly one of them was the common component of COMPASS complexes[50,51] *Wdr5* that, besides having a stable expression along transdifferentiation (Figure 47e), in *Mll2*$^{-/-}$ cells acquired H3K4

trimethylation, probably also for a compensatory mechanism, but failed to be more expressed in the knock-out with respect to the control iNs (Figure 47g).

The taurine upregulated 1 (Tug1) and *the MORC family CW-type finger protein 2A* (*Morc2a),* instead, were bound by MENIN and H3K4 trimethylated in *Mll2*$^{-/-}$ cells. This was probably the result of the mislocalization of the trithorax COMPASS-like complexes in absence of MLL2. The two genes are stable during transdifferentiation (Figure 47f). In particular, *Tug1* is a long noncoding RNA induced by p53 that associates with PRC2 and regulates many cell cycle genes[197].

Finally the H3K4me3 mark was erased from *Ephrinb2* (*Efnb2)* (important for neuronal migration during development[198]) at 13 days independently from both the genotype and the MENIN binding, indicating that the H3K4me3 deposition and erasure on this gene was not associated to MLL2. *Efnb2* is up-regulated 2 days after BAM transfection and down-regulated at 13 days (Figure 47h). In *Mll2*$^{-/-}$ iNs it failed to be down-regulated as in *Mll2*$^{+/+}$ iNs, despite the erasure of H3K4me3 mark (Figure 47i).


a
Dock4, Zfp277
Atg7
Rtn1
Gnao1, 4930488L21Rik
Neurl1a
Bend5
Zfp239
7:118447101:118448700
Zfp184
19:61199801:61201300
17:83860701:83861900
Proser2, Gm10857
Nipal1
11:23008201:23009400
Csrp2
Tifa
Mpped2
St6galnac3
B230208H11Rik
Cacnb2
Lhx6
Slc43a1
Ptch1
Creld1
2:18587201:18588200
Lrguk
Syt1
Magohb
4:54944401:54945400
Pfas
6330415G19Rik
2:98661301:98668300
Efnb2
MT:13301:16100
Wdr5
D030047H15Rik, Leng8
11:76494401:76495400
Nol10
Imp4, Ccdc115
Tug1, Morc2a
Row Z-Score
2
1
0
-1
-2
Menin.KO.2
Menin.KO.1
Menin.CTL.2
Menin.CTL.1
Mll1_K4
L1_K4
L7_K4
L1_13K4
L7_13K4
M6_K4
M7_K4
M6_13K4
M7_13K4
H3K4me3
Menin
b
Mll1+/+ Mll2+/+
Mll2-/-
FPKM
30
20
10
4
2
0
*
*
Cbx7
Bmi1
c
FPKM
30
20
10
0
*
Mll1+/+ Mll2+/+
Mll2-/-
d
MEF
BAM 2 days
BAM 13 days
RPKM
80
60
40
20
0.3
0.2
0.1
0.0
Lhx6
Nipal1
Syt1
Csrp2
e
MEF
BAM 2 days
BAM 13 days
RPKM
15
10
5
0
Wdr5
f
MEF
BAM 2 days
BAM 13 days
RPKM
20
15
10
5
0
Tug1
Morc2a
g
FPKM
15
10
5
0
Mll1+/+ Mll2+/+
Mll2-/-

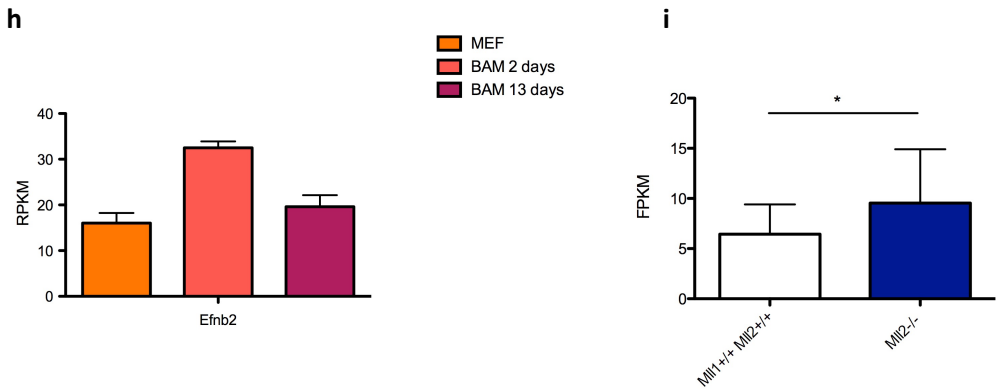


Figure 47: Analysis of the genes differentially H3K4me3 and MENIN bound

a. Heat map of the genes differentially H3K4 trimethylated and MENIN bound in *Mll2*$^{-/-}$ and controls. The colour scale (-2 blue to +2 yellow) represents Z-Score. **b. c. g.** and **i.** FPKM in our RNA-seq at 13 days of respectively Polycomb components, *GluA2*, *Wdr5* and *Efnb2*. Means and SD are reported. **d. e. f.** and **h.** RPKM, during transdifferentiation, of the genes differentially H3K4 trimethylated and MENIN bound. Means and SD are reported. Source data: Wapinski et al., Cell 2013[106]. * FDR<0.01

Moreover the majority of the genes belonging to the first category had a further decrease in H3K4 trimethylation at 13 days. Therefore we performed another analysis focused only on this time point.

112 genes showed a lower H3K4me3 enrichment in *Mll2*$^{-/-}$ iNs with respect to *Mll2*$^{+/+}$ on the same day. Among them, 40 were differentially expressed (Figure 48). As expected, all the genes that showed a further H3K4 demethylation at 13 days, in *Mll2*$^{-/-}$ transdifferentiating MEFs, were down-regulated in *Mll2*$^{-/-}$ iNs.

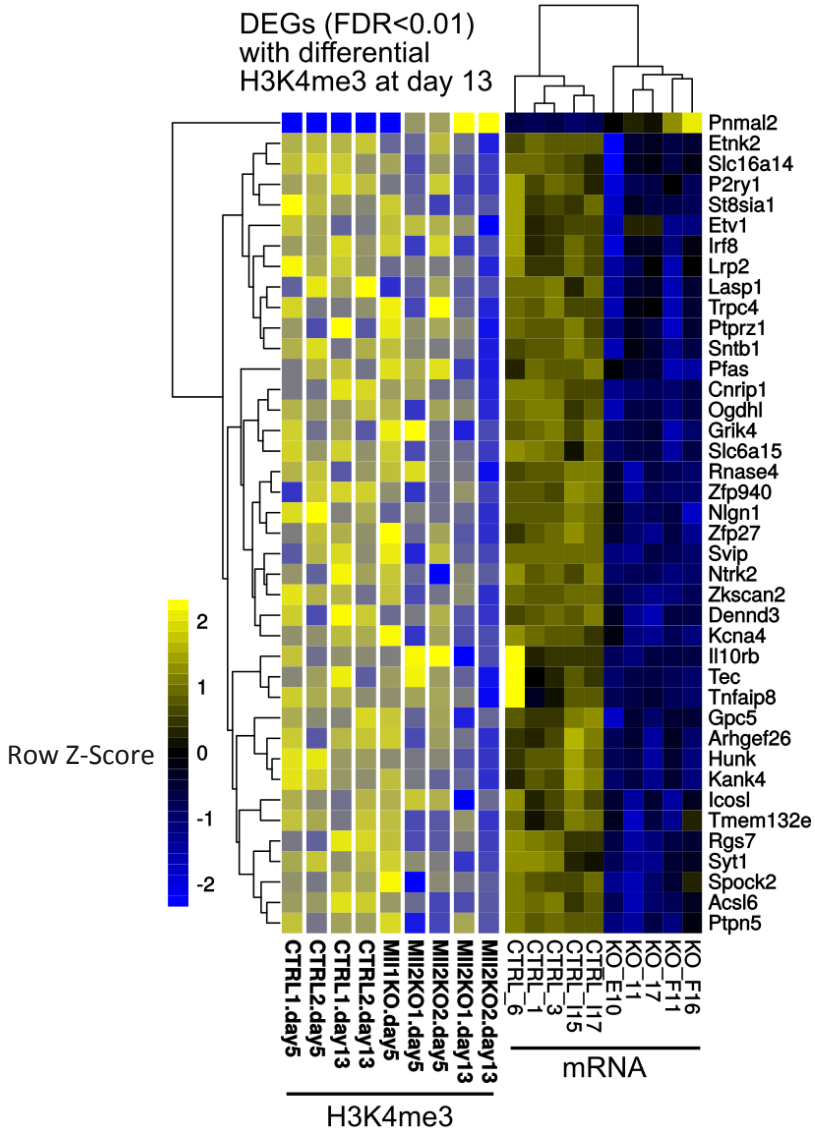


Figure 48: Analysis of the genes with reduced H3K4me3 at 13 days

Heat map of the genes that further lost H3K4 trimethylation at 13 days and their expression in RNA-seq at 13 days. The colour scale (-2 blue to +2 yellow) represents Z-Score.

8. DISCUSSION

In this thesis I defined the role of MLL1 and MLL2 during MEFs-to-iNs BAM-driven transdifferentiation. MLL1 and MLL2 are two H3K4 trimethylases, historically discovered for their function in maintaining the expression of *Hox* genes[51,55,77], important during both development and differentiation. Furthermore both enzymes were shown to exert fundamental roles during neuronal differentiation. Indeed ectodermal *in vitro* differentiation of *Mll2*$^{-/-}$ ESCs is impaired[84], while in *Mll1*$^{-/-}$ ESCs, MLL2 can compensate, trimethylating MLL1-targets genes[76]. The ablation of *Mll2* in excitatory forebrain neurons leads to memory defects, but does not impinge brain morphology[125]. The neuronal differentiation of *Mll1*$^{-/-}$ SVZ NCS is impaired and this is not due to a reduction in H3K4me3 mark deposition, but to the absence of the erasure of the H3K27me3 at important target genes[123]. On the contrary, the absence of MLL1 during *Danio rerio* development reduces neural progenitors proliferation, anticipating neuronal differentiation[124].

It is therefore conceivable that MLL1 and MLL2 are involved in the conspicuous epigenome resetting that allows the conversion between MEFs and iNs, two cell types so embryologically distant. Therefore I investigated how the absence of MLL1 and/or MLL2 affects transdifferention.

I took advantage of mouse models conditional knock-out for *Mll1*[128], *Mll2*[56] or their compound strain, in a tamoxifen inducible system. Interestingly, when I assessed the MLL2 level through western blot, despite the fact that in ESCs both the full length and the active taspase-cleaved MLL2 were present, in MEFs only the uncleaved protein was observed, which, as expected, disappeared upon 4-OHT treatment. This suggests that in MEFs, *Mll2* is translated, but it is inactive. When I evaluated the level of *Mll2* mRNA 5 and 13 days (both through RT-qPCR and through RNA-seq) after cells were plated for

transdifferentiation, I did not observe a decrease in the level of *Mll2* transcript due to nonsense-mediated decay, probably because MLL2 is maintained inactive in MEFs.

First I showed that MLL1 is dispensable for the MEFs-to-iNs direct conversion. The efficiency of transdifferentiation of *Mll1*$^{-/-}$ cells was reduced at FACS, but this was paralleled by their higher cell death (Figure 49b). Moreover the percentage of PSA-NCAM^{+} *Mll1*$^{-/-}$ cells was fully overlapping to the one of control, along the entire transdifferentiation process, suggesting that the proportion of dying cells was the same among PSA-NCAM^{+} and PSA-NCAM^{-} cells. Hence the reduced transdifferentiation efficiency that we observed was the indirect consequence of the affected cell viability in cells deprived by MLL1 (Figure 49b). Indeed MLL1 was demonstrated to be associated to cell cycle regulator genes[83], but although the straight knock-out embryos showed a widespread cell death at E.10.5[55], there are no any other evidences that directly connect MLL1 deletion to apoptosis. Recently it was demonstrated that endoplasmic reticulum stress-induced apoptosis is enhanced in *Mll1*$^{-/-}$ MEFs[199], however cell death was not the consequence of MLL1 deletion. Because the proportion of dead cells was the same among iNs and untransdifferentiated cells, we could not state that it is the transdifferentiation process *per se* that causes cell death in an already compromised starting cell. Most likely MLL1 activates anti-apoptotic genes independently from the considered cell type and state. In ScanR experiments the efficiency of transdifferentiation of *Mll1*$^{-/-}$ cells varied depending on the cell mortality rate, but MEFs were plated for BAM factors transduction immediately after 4-OHT. Despite the fact that there is no effective anti-MLL1 antibody (reactive with mouse), we can be quite confident about its degradation, also in this experimental setting. Indeed, it has been shown that MLL1 is degraded each late S and M phase[78]. MEFs cultured in low oxygen and at low passages have a doubling time of ~ 24h[200]. Therefore only two additional days before plating the cells for transduction, should be enough to ensure MLL1 degradation and the alteration in the transdifferentiation rate of

Mll1$^{-/-}$ cells, in this experimental setting, should be attributed to their altered cycling in the absence of MLL1. Indeed, depending on the level of cell cycle affection, we observed a lower or higher transdifferentiation efficiency. Finally *Mll1*$^{-/-}$ neurons had the same morphology and matured as well as control iNs, further underling that MLL1 is dispensable during transdifferentiation.

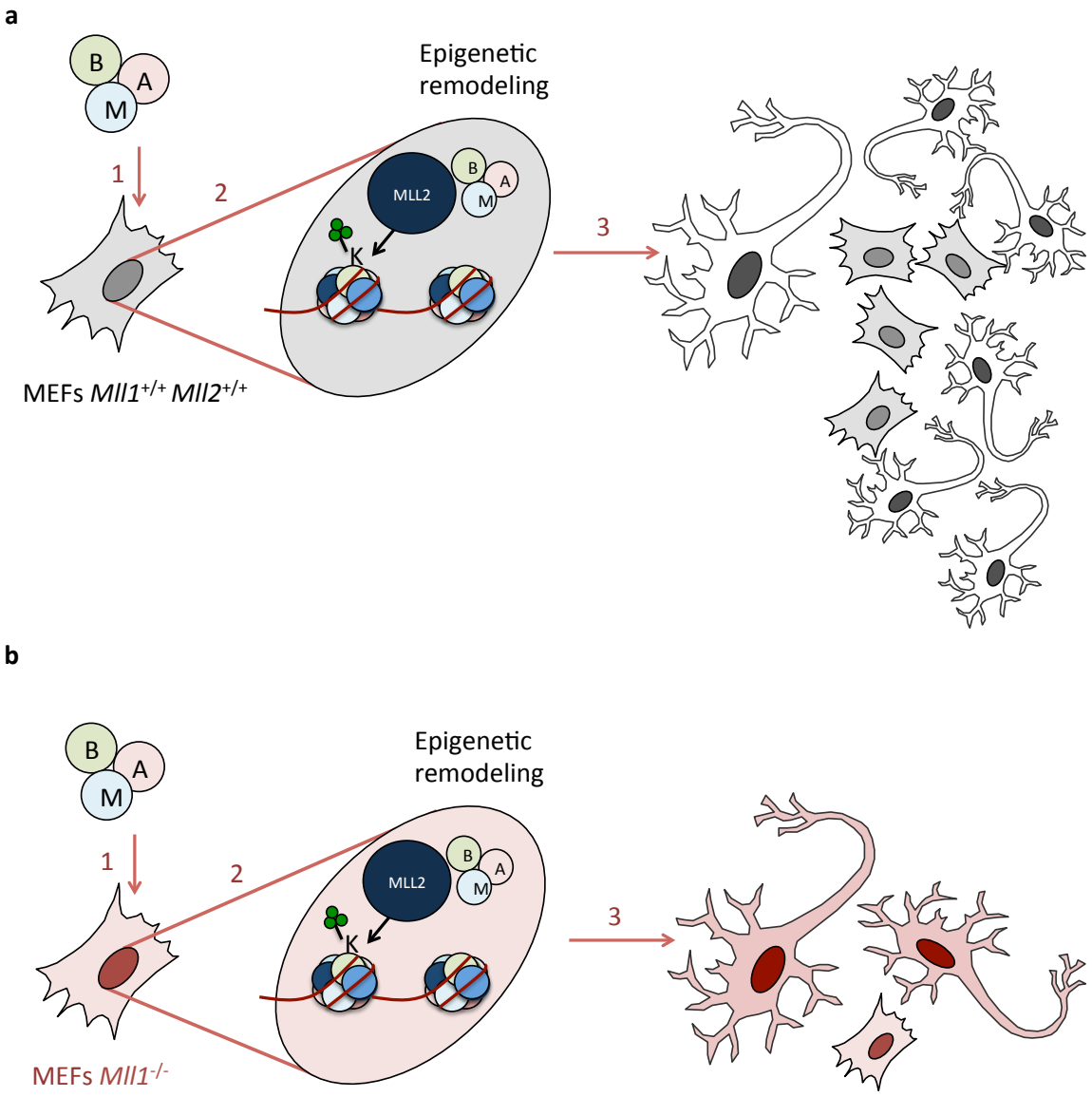
When we performed H3K4me3 ChIP-seq on 5 days transdifferentiating MEFs, *Mll1*$^{-/-}$ cells clustered with *Mll1*$^{+/+}$*Mll2*$^{+/+}$ cells. Therefore, even if only in one sample, the H3K4me3 redistribution, necessary for transdifferentiation, was not affected in the absence of MLL1. This on one side could be reconciled by the lack of a role for MLL1 during neuronal direct cell conversion. On the other side MLL2 could compensate its absence, trimethylating MLL1 specific targets.

Therefore, although the *Mll1* ablation is detrimental for SVZ NSC neuronal differentiation[123], transdifferentiation can proceed even if this methylase is deleted, accordingly to its role during zebrafish development, where the usage of a specific morpholino against *Mll1* does not impede, but it prematurely induces neuronal differentiation[124]. This is not counterintuitive because it is not presume, that transdifferentiation would undertake the same epigenetic changes of "physiologic" differentiation and is even more evident since the starting cells have completely different epigenomes.

The defective neuronal differentiation of *Mll1*$^{-/-}$ SVZ NSC was specifically due to the lack of the erasure of H3K27 trimethylation at critical target genes. To exclude this option and to analyse if also for what concerns this histone mark, *Mll1*$^{-/-}$ transdifferentiating MEFs cluster with control, I have already performed the H3K27me3 ChIP-seq on 5 days *Mll1*$^{-/-}$ transdifferentiating MEFs, which will be analysed in the next future.

In the case of *Mll2*$^{-/-}$ MEFs, plated 2 days after 4-OHT treatment, I observed an initial deficit in neuronal transdifferentiation, overcome at 13 days (both in FACS and in ScanR

experiments). If, instead, MEFs were plated for transdifferentiation one week after 4-OHT treatment the defect overcoming no more occurred (Figure 49c).



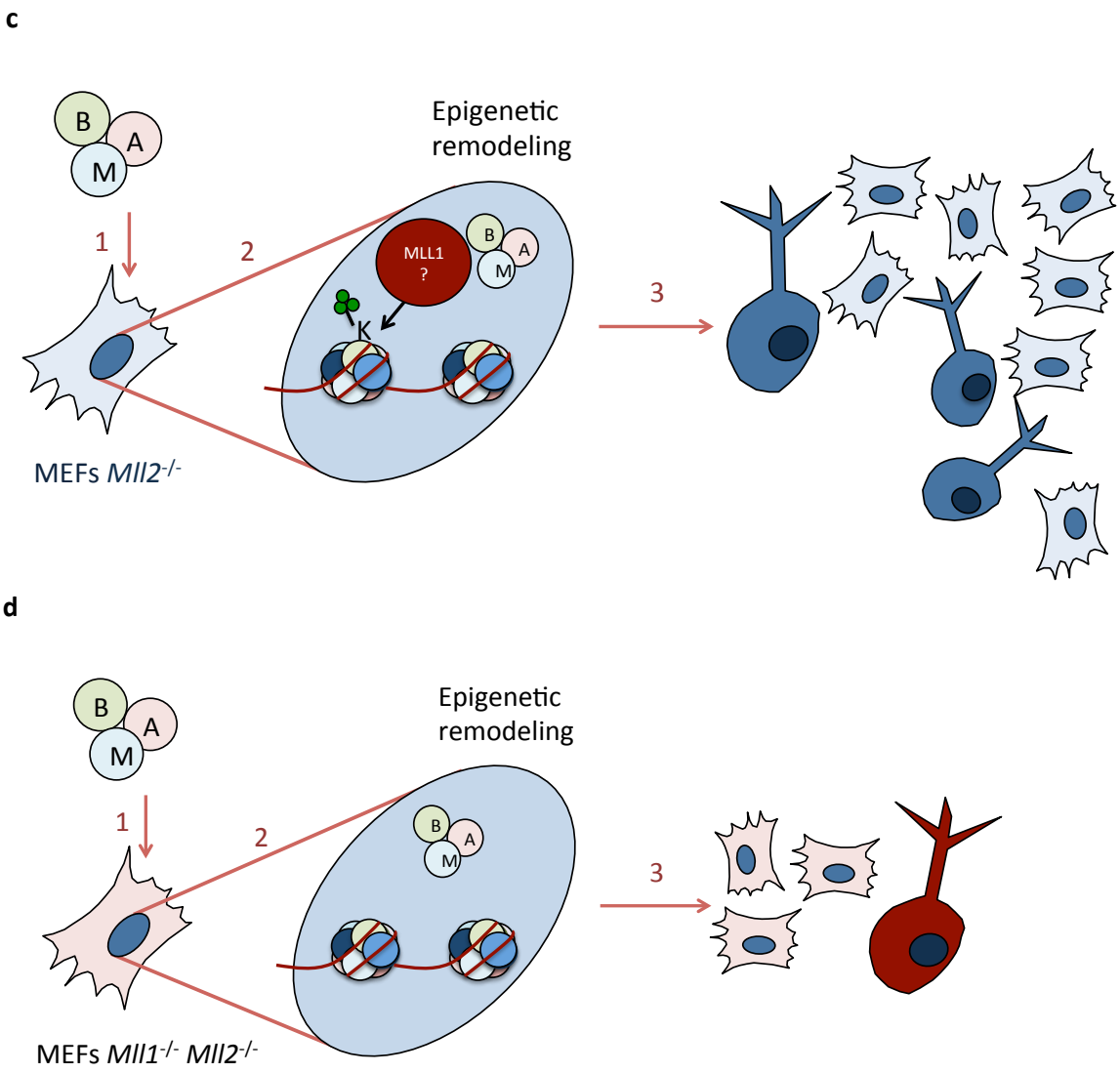


Figure 49: Schematic model of transdifferentiation in the different knock-out

Transdifferentiation efficiency and mortality rate assuming to start from the same number of MEFs, left in culture 7 days after 4-OHT, before plating them for BAM factors transduction. 1- BAM factors transduction; 2- epigenetic remodelling; 3- cell populations observed at 13 days**. a.** Transdifferentiation of *Mll1*$^{+/+}$*Mll2*$^{+/+}$ MEFs. **b.** Transdifferentiation of *Mll1*$^{-/-}$ MEFs. Mortality rate is high with respect to *Mll1*$^{+/+}$*Mll2*$^{+/+}$ transdifferentiating MEFs, but iNs are mature. **c.** Transdifferentiation of *Mll2*$^{-/-}$ MEFs. Mortality rate is comparable to the one of *Mll1*$^{+/+}$*Mll2*$^{+/+}$ transdifferentiating MEFs, but iNs are immature and the efficiency of cell conversion is lower. **d.** Transdifferentiation of *Mll1*$^{-/-}$*Mll2*$^{-/-}$ MEFs. Mortality rate is comparable to the one of *Mll1*$^{-/-}$ transdifferentiating MEFs; very few iNs are generated and they remain immature.

We can confidently exclude that the absence of MLL2 would make MEFs less prone to transdifferentiation. Ascl1 primarily binds to trivalent chromatin states composed by H3K4me1, H3K27Ac and H3K9me[106]. Cells with few trivalent chromatin *loci* convert into

iNs at low efficiency. It has been demonstrated that MLL3 and MLL4 monomethylate H3K4 at enhancer[67,71] and that in active enhancer H3K4me1 is associated with H3K27ac[28]. It is therefore conceivable that MLL3 and MLL4 are also responsible of H3K4 monomethylation at trivalent chromatin states. MLL2 instead is mostly a H3K4 trimethylase, which excludes its possible involvement in structuring the trivalent chromatin state. Moreover MLL1, the homolog of MLL2, is responsible for the H3K4 trimethylation of only 5% of the MEF promoter genes commonly marked by H3K4me3[51]. Because MLL1 and MLL2 possess some specific, but also some shared targets and since SET1A and SET1B are the main responsible for the deposition of H3K4me3 mark[46], we could expect that MLL2, on average, would quantitatively trimethylate the same proportion of targets of MLL1. Therefore the deletion of MLL2 should not lead to a massive reorganization of H3K4me3 and should not affect the specific "MEFs genes" that make this cell type prone to transdifferentiation. However to formally prove it an *ad hoc* ChIP-seq should be performed on control and knock-out MEFs. On the other side, in *Mll1*$^{-/-}$ MEFs, 3% of the genes presented also a reduced expression, concomitantly to the reduced H3K4me3[51]. In particular, these genes were involved in cell signalling and transcriptional regulation and among them also specific *Hox* genes were dysregulated. A change in cell identity/morphology was not reported in this study. Hence, most likely, the down-regulation of these genes does not perturb the MEF lineage specification *per se*. However, since MLL2 has both specific and shared with MLL1 *Hox* targets, the alteration of *Hox* gene expression could affects the *de novo* acquisition of neuronal identity of *Mll2*$^{-/-}$ MEFs. However, in our analysis, *Hox* genes did not result dysregulated, excluding their possible involvement in the defective transdifferentiation observed in the absence of *Mll2*.

Contrary to what has been demonstrated in different cellular types and contexts, such as in ESCs[84], *Mll2*$^{-/-}$ MEFs showed the same cell death rate of control cells during transdifferentiation. Specifically, in *Mll2*$^{-/-}$ ESCs, the higher cell mortality was attributed to the down-regulation of *Bcl-2,* that was indeed lower expressed also in *Mll2*$^{-/-}$ iNs with

respect to control. However *Pdcd5*, which encodes for a protein released during apoptosis and that favours programmed cell death[165,166], was down-regulated in *Mll2*$^{-/-}$ iNs and came out as one of the most differentially expressed genes. Moreover *Bax*, up-regulated in *Mll2*$^{-/-}$ oocytes[74], and *Bad* had similar levels of expression between knock-out and control iNs. It has been recently shown that when *Bcl2* is co-transduced with *Ascl1* during astrocytes-neurons transdifferentiation, it favours the process of cell conversion in an apoptosis unrelated manner[201]. Therefore we could envisage that the lower expression of *Bcl2* would further emphasize the reduced transdifferentiation efficiency of *Mll2*$^{-/-}$ MEFs, more than increasing its mortality rate.

When I immediately plated *Mll2*$^{-/-}$ MEFs after 4-OHT treatment I did not observe any defect in transdifferentiation efficiency at 13 days both in FACS and ScanR experiments. This has two possible explanations. First, 5 days in 4-OHT plus 2 days in normal medium could be not sufficient for the degradation of MLL2 translated prior to the induction of Cre recombinase activity. In ESCs two days of 4-OHT treatment are enough to see the disappearance of the MLL2 protein. However in ESCs this methylase could have a different turnover rate since, as shown in our western blot, MLL2 seems active in ESCs and inactive in MEFs. To check this hypothesis a western blot for MLL2 in MEFs 2 days after the end of 4-OHT treatment has to be performed. However, at FACS, we anyway observed an initial defect in transdifferentiation in *Mll2*$^{-/-}$ cells. In a preliminary experiment, where I transdifferentiated *Mll2*$^{-/-}$ and Mll2$^{+/+}$ MEFs, left in culture after 4-OHT only 2 days, I observed that, after doxycycline administration, knock-out MEFs still proliferate. As I shown in the rest of the thesis, this was not the case for any of the MEFs plated one week after 4-OHT treatment, independently from their genotype, as previously published[91]. Despite the fact that a more comprehensive study of the *Mll2*-devoid MEFs cell cycle, before and along transdifferentiation, is needed, the proliferative advantage of *Mll2*$^{-/-}$ MEFs could explain the higher number of *Mll2*$^{-/-}$ iNs generated with respect to the number of plated cells. Proliferation is not required for neuronal transdifferentiation[107], but

it would lead to start from a higher number of cells from which, consequently, more iNs would be obtained. Moreover, if the slower transdifferentiating MEFs, still PSA-NCAM$^-$, at the beginning would continue to cycle, this would result in a lower percentage of PSA-NCAM$^+$ cells, that indeed we observed. Hence this model would better reconcile most of the observations done on cells immediately plated after 4-OHT. Despite the fact that MLL2, as MLL1, interacts with E2Fs proteins[61], which in turn regulate cell cycle, in *Mll2*$^{-/-}$ ESCs there are no evidences that connect *Mll2* deletion to an alteration either of the cell cycle or of the proliferation rate[84]. However this study was conducted in different cell type with respect to the one I used[84], where the MLL2 function could be different.

The role of MLL2 during transdifferentiation can thus be distinguished from the impact of its absence on cell cycle/cell mortality, plating MEFs one week after 4-OHT treatment. Therefore RNA-seq and ChIP-seq studies were performed only according to this experimental setting. What can be deduced from both the transdifferentiation kinetics and the transcriptomic/epigenomic studies is that the deletion of *Mll2* impinges on transdifferentiation efficiency, but, more importantly, that *Mll2*$^{-/-}$ iNs fail to mature. This is partially in accordance with the results obtained with retinoic acid-mediated neuronal differentiation of ESCs, in which the deletion of *Mll2* led to lower expression of "maturation genes" with respect to the controls, but mature neurons were anyway present in culture[84]. Moreover I showed that the absence of MLL2 severely affected neurite extension, independently from how many days after 4-OHT MEFs were plated, contrary to the conditional deletion of *Mll2* in excitatory neurons *in vivo,* where a defect in neuronal/brain morphology was not observed[125].

Recently the network of genes involved in the MEF-to-iNs transdifferentiation has been described and further subdivided in three subnetworks: the MEF subnetwork, the initiation subnetwork and the maturation subnetwork, on the basis of the phase of the process that the specific group of genes regulates[108]. A conspicuous part of the maturation subnetwork

genes was indeed down-regulated in *Mll2*$^{-/-}$ iNs, such as also the genes previously shown by the same group to be up-regulated to allow transdifferentiation[106]. The only exception came from *Zfp238*, one of the main effector of *Ascl1*[106]. *Zfp238,* during the first transdifferentiation phases, was up-regulated also in *Mll2*$^{-/-}$ transdifferentiating MEFs, even if at a much lower extent, but it failed to be down-regulated at 13 days (assessed both through RT-qPCR and through RNA-seq). However it did not come out among the genes that lost the H3K4me3 at the end of the process. One possible hypothesis that will be tested in the next future through ChIP-seq, is that *Zfp238* fails to be H3K27 trimethylated in the absence of MLL2.

Among the differentially expressed genes, but also the differentially H3K4 trimethylated ones, a relevant fraction was connected to neurite elongation and synapse formation further corroborating the observed phenotype at the molecular level.

When we analysed MENIN binding and the H3K4me3 mark distribution we observed that in *Mll2*$^{-/-}$ MEFs, 839 genes lost MENIN binding, pointing to the absence of a complete compensation of MLL1. The lack of a compensation by MLL1 is further suggested by the absence of an up-regulation of its mRNA in *Mll2*$^{-/-}$ iNs.

On the other side 213 genes acquired *de novo* MENIN binding, suggesting either a mislocalization of the MLL1 and MLL2 COMPASS-like complex or a possible adaptive response of the cells to the absence of MLL2. From gene ontology analysis the genes, which lost MENIN binding, were related to basic cell functions and metabolism, while the genes that acquired MENIN binding were related to non neuronal cell identities, suggesting that the second hypothesis is the most plausible one.

For what concerns the H3K4 trimethylation, 545 genes lost the mark, while 82 genes acquired it, further demonstrating that MLL2 targets specific *loci* during transdifferentiation, whose vulnerability to the loss of MLL2 is not compensated by

MLL1. Also in this case, the differentially expressed genes among them were related to neuronal maturation and synaptic transmission.

Finally, among the differentially H3K4 trimethylated and MENIN bound genes, *Zfp277* was identified[188]. This zinc finger protein has been associated with BMI1, a component of PRC1 complex and, in our datasets, both *Bmi1* and *Cbx7* were differentially expressed between *Mll2*$^{-/-}$ and control iNs. Moreover the differentially expressed genes were enriched for PRC2 targets in mESCs. Therefore also the Polycomb axis could be the cause of the altered gene expression profile we observed in *Mll2*$^{-/-}$ iNs and, hence, of their maturation defects. Trithorax and Polycomb are two families of proteins with opposite and balanced functions. Indeed MLL2 prevents H3K27me3 deposition at specific targets[63,64]. Therefore it is plausible that, in the absence of MLL2 and without the compensation of MLL1, Polycomb would prevail, at least at some targets.

I thus propose a model where the role of MLL2 during transdifferentiation is the H3K4 trimethylation and induction of "neuronal maturation genes". Hence, in its absence, transdifferentiation can start, even if less efficiently, but the generated iNs remain immature (Figure 50). In particular, to confirm this hypothesis, I will perform electrophysiological studies on *Mll2*$^{-/-}$ and control iNs.

On the other side MLL1 cannot compensate for MLL2 absence, as it is the case, instead, during retinoic acid ESCs neuronal differentiation[76]. Hence, again, at least for what concerns MLL1-MLL2-H3K4me3 deposition, the differentiation following the Waddington descent and the crossing of the epigenetic hill are not comparable, probably both because the starting cells possess very different epigenomes and the two systems entail different "routes" to reach the neuronal identity.

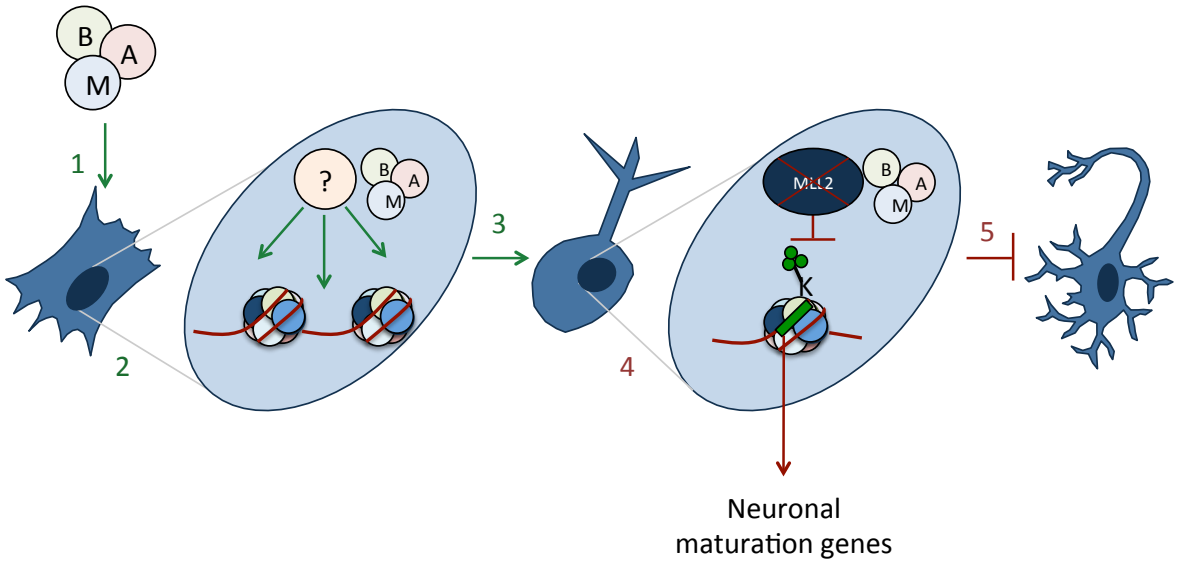


Figure 50: Schematic model of transdifferentiation in the absence of MLL2

Upon BAM factors transduction (1) MEFs start to transdifferentiate (2), probably through the recruitment of a still undefined epigenetic remodeller. However, the absence of MLL2, already affects the efficiency of this step. In iNs generated (3) MLL2 cannot trimethylates fundamental neuronal maturation genes (4). Therefore only iNs with short neurites are detected at 13 days (5).

H3K4me3 has been linked to alternative splicing[142-145], therefore we studied also if in absence of MLL2 this process is affected. Indeed we described few genes, but all subjected to a specific differential exon usage never described in literature. Moreover all genes were strictly linked to neuronal function and maturation. Strikingly the loss of both *Mll1* and *Mll2* severely affected iNs generation and morphology and led to an extremely high cell death (Figure 49d). In particular, despite the fact that the cell death rate was the same between *Mll1*$^{-/-}$ and *Mll1*$^{-/-}$ *Mll2*$^{-/-}$ transdifferentiating MEFs, the efficiency was lower in the double knock-out. This further demonstrates that it is indeed MLL2 the main Trithorax like COMPASS complex component responsible of H3K4me3 remodelling during transdifferentiation, while the absence of *Mll1* only affects cell viability.

9. REFERENCES

1 Whitcomb, S. J., Basu, A., Allis, C. D. & Bernstein, E. Polycomb Group proteins: an evolutionary perspective. *Trends in genetics : TIG* **23**, 494-502, doi:10.1016/j.tig.2007.08.006 (2007).

2 Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1-4, doi:10.1016/S0092-8674(00)80599-1 (1999).

3 Bird, A. *et al.* Studies of DNA methylation in animals. *Journal of cell science. Supplement* **19**, 37-39 (1995).

4 C.H:, W. The epigenotype. *Endeavour* **1**, 18-20 (1942).

5 Waddington, C. H. The epigenotype. 1942. *International journal of epidemiology* **41**, 10-13, doi:10.1093/ije/dyr184 (2012).

6 Meloni, M. & Testa, G. Scrutinizing the epigenetics revolution. *BioSocieties* **9**, 431-456, doi:10.1057/biosoc.2014.22 (2014).

7 Briggs, R. & King, T. J. Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. *Proceedings of the National Academy of Sciences of the United States of America* **38**, 455-463 (1952).

8 Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. *Nature* **182**, 64-65 (1958).

9 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).

10 Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920, doi:10.1126/science.1151526 (2007).

11 Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872, doi:10.1016/j.cell.2007.11.019 (2007).

12 Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250-264, doi:10.1016/j.cell.2008.03.028 (2008).

13 Kim, J. B. *et al.* Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* **454**, 646-650, doi:10.1038/nature07061 (2008).

14 Utikal, J., Maherali, N., Kulalert, W. & Hochedlinger, K. Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. *Journal of cell science* **122**, 3502-3510, doi:10.1242/jcs.054783 (2009).

15 Eminli, S., Utikal, J., Arnold, K., Jaenisch, R. & Hochedlinger, K. Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. *Stem cells* **26**, 2467-2474, doi:10.1634/stemcells.2008-0317 (2008).

16 Papp, B. & Plath, K. Epigenetics of reprogramming to induced pluripotency. *Cell* **152**, 1324-1343, doi:10.1016/j.cell.2013.02.043 (2013).

17 Blau, H. M., Chiu, C. P. & Webster, C. Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell* **32**, 1171-1180 (1983).

18 Taylor, S. M. & Jones, P. A. Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell* **17**, 771-779 (1979).

19 Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000 (1987).

20 Ladewig, J., Koch, P. & Brustle, O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nature reviews. Molecular cell biology* **14**, 225-236 (2013).

21 Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117**, 663-676 (2004).

22 Uhlenhaut, N. H. *et al.* Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* **139**, 1130-1142, doi:10.1016/j.cell.2009.11.021 (2009).

23 Kadkhodaei, B. *et al.* Nurr1 is required for maintenance of maturing and adult midbrain dopamine neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **29**, 15923-15932, doi:10.1523/JNEUROSCI.3910-09.2009 (2009).

24 Lewis, E. B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570 (1978).

25 Ingham, P. W. Differential expression of bithorax complex genes in the absence of the extra sex combs and trithorax genes. *Nature* **306**, 591-593 (1983).

26 Ingham, P. W. trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective. *The International journal of developmental biology* **42**, 423-429 (1998).

27 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **39**, 311-318, doi:10.1038/ng1966 (2007).

28 Creyghton, M. P. *et al.* 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**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

29 Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326, doi:10.1016/j.cell.2006.02.041 (2006).

30 Gao, Z. *et al.* PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Molecular cell* **45**, 344-356, doi:10.1016/j.molcel.2012.01.002 (2012).

31 Cao, R., Tsukada, Y. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Molecular cell* **20**, 845-854, doi:10.1016/j.molcel.2005.12.002 (2005).

32 Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678, doi:10.1016/j.cell.2011.12.029 (2012).

33 Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular cell* **15**, 57-67, doi:10.1016/j.molcel.2004.06.020 (2004).

34 Nekrasov, M., Wild, B. & Muller, J. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO reports* **6**, 348-353, doi:10.1038/sj.embor.7400376 (2005).

35 Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306-310, doi:10.1038/nature08788 (2010).

36 Sanulli, S. *et al.* Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Molecular cell* **57**, 769-783, doi:10.1016/j.molcel.2014.12.020 (2015).

37 Jermann, P., Hoerner, L., Burger, L. & Schubeler, D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E3415-3421, doi:10.1073/pnas.1400672111 (2014).

38 Roguev, A. *et al.* The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO journal* **20**, 7137-7148, doi:10.1093/emboj/20.24.7137 (2001).

39 Milne, T. A. *et al.* MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14765-14770, doi:10.1073/pnas.0503630102 (2005).

40 Yokoyama, A. *et al.* Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Molecular and cellular biology* **24**, 5639-5649, doi:10.1128/MCB.24.13.5639-5649.2004 (2004).

41 Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N. & Herr, W. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes & development* **17**, 896-911, doi:10.1101/gad.252103 (2003).

42 Goo, Y. H. *et al.* Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. *Molecular and cellular biology* **23**, 140-149 (2003).

43 Hughes, C. M. *et al.* Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. *Molecular cell* **13**, 587-597 (2004).

44 Dou, Y. *et al.* Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature structural & molecular biology* **13**, 713-719, doi:10.1038/nsmb1128 (2006).

45 Patel, A., Dharmarajan, V., Vought, V. E. & Cosgrove, M. S. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *The Journal of biological chemistry* **284**, 24242-24256, doi:10.1074/jbc.M109.014498 (2009).

46 Wu, M. *et al.* Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Molecular and cellular biology* **28**, 7337-7344, doi:10.1128/MCB.00976-08 (2008).

47 Rao, R. C. & Dou, Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nature reviews. Cancer* **15**, 334-346, doi:10.1038/nrc3929 (2015).

48 Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Molecular cell* **11**, 709-719 (2003).

49 Kim, J. *et al.* RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**, 459-471, doi:10.1016/j.cell.2009.02.027 (2009).

50 Miller, T. *et al.* COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12902-12907, doi:10.1073/pnas.231473398 (2001).

51 Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Molecular and cellular biology* **29**, 6074-6085, doi:10.1128/MCB.00924-09 (2009).

52 Mo, R., Rao, S. M. & Zhu, Y. J. Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. *The Journal of biological chemistry* **281**, 15714-15720, doi:10.1074/jbc.M513245200 (2006).

53 Kaikkonen, M. U. *et al.* Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular cell* **51**, 310-325, doi:10.1016/j.molcel.2013.07.010 (2013).

54 Guenther, M. G. *et al.* Global and Hox-specific roles for the MLL1 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8603-8608, doi:10.1073/pnas.0503072102 (2005).

55 Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. & Korsmeyer, S. J. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature* **378**, 505-508, doi:10.1038/378505a0 (1995).

56 Glaser, S. *et al.* Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* **133**, 1423-1432, doi:10.1242/dev.02302 (2006).

57 Lee, S. *et al.* Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 15392-15397, doi:10.1073/pnas.0607313103 (2006).

58 Terranova, R., Agherbi, H., Boned, A., Meresse, S. & Djabali, M. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 6629-6634, doi:10.1073/pnas.0507425103 (2006).

59 Hsieh, J. J., Cheng, E. H. & Korsmeyer, S. J. Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell* **115**, 293-303 (2003).

60 Yokoyama, A., Kitabayashi, I., Ayton, P. M., Cleary, M. L. & Ohki, M. Leukemia proto-oncoprotein MLL is proteolytically processed into 2 fragments with opposite transcriptional properties. *Blood* **100**, 3710-3718, doi:10.1182/blood-2002-04-1015 (2002).

61 Takeda, S. *et al.* Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. *Genes & development* **20**, 2397-2409, doi:10.1101/gad.1449406 (2006).

62 Jiang, H. *et al.* Regulation of transcription by the MLL2 complex and MLL complex-associated AKAP95. *Nature structural & molecular biology* **20**, 1156-1163, doi:10.1038/nsmb.2656 (2013).

63 Austenaa, L. *et al.* The histone methyltransferase Wbp7 controls macrophage function through GPI glycolipid anchor synthesis. *Immunity* **36**, 572-585, doi:10.1016/j.immuni.2012.02.016 (2012).

64 Glaser, S. *et al.* The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics & chromatin* **2**, 5, doi:10.1186/1756-8935-2-5 (2009).

65 Ladopoulos, V. *et al.* The histone methyltransferase KMT2B is required for RNA polymerase II association and protection from DNA methylation at the MagohB CpG island promoter. *Molecular and cellular biology* **33**, 1383-1393, doi:10.1128/MCB.01721-12 (2013).

66 Lee, J. E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife* **2**, e01503, doi:10.7554/eLife.01503 (2013).

67 Hu, D. *et al.* The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Molecular and cellular biology* **33**, 4745-4754, doi:10.1128/MCB.01181-13 (2013).

68 Cho, Y. W. *et al.* PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *The Journal of biological chemistry* **282**, 20395-20406, doi:10.1074/jbc.M701574200 (2007).

69 Patel, S. R., Kim, D., Levitan, I. & Dressler, G. R. The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Developmental cell* **13**, 580-592, doi:10.1016/j.devcel.2007.09.004 (2007).

70 Lee, J. H. & Skalnik, D. G. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex,

the analogue of the yeast Set1/COMPASS complex. *The Journal of biological chemistry* **280**, 41725-41731, doi:10.1074/jbc.M508312200 (2005).

71 Herz, H. M. *et al.* Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & development* **26**, 2604-2620, doi:10.1101/gad.201327.112 (2012).

72 Lee, J. H., Tate, C. M., You, J. S. & Skalnik, D. G. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *The Journal of biological chemistry* **282**, 13419-13428, doi:10.1074/jbc.M609809200 (2007).

73 Bach, C., Mueller, D., Buhl, S., Garcia-Cuellar, M. P. & Slany, R. K. Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene* **28**, 815-823, doi:10.1038/onc.2008.443 (2009).

74 Andreu-Vieyra, C. V. *et al.* MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS biology* **8**, doi:10.1371/journal.pbio.1000453 (2010).

75 Ernst, P., Mabon, M., Davidson, A. J., Zon, L. I. & Korsmeyer, S. J. An Mll-dependent Hox program drives hematopoietic progenitor expansion. *Current biology : CB* **14**, 2063-2069, doi:10.1016/j.cub.2004.11.012 (2004).

76 Denissov, S. *et al.* Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**, 526-537, doi:10.1242/dev.102681 (2014).

77 Hu, D. *et al.* The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature structural & molecular biology* **20**, 1093-1097, doi:10.1038/nsmb.2653 (2013).

78 Liu, H., Cheng, E. H. & Hsieh, J. J. Bimodal degradation of MLL by SCFSkp2 and APCCdc20 assures cell cycle execution: a critical regulatory circuit lost in

leukemogenic MLL fusions. *Genes & development* **21**, 2385-2398, doi:10.1101/gad.1574507 (2007).

79 Blobel, G. A. *et al.* A reconfigured pattern of MLL occupancy within mitotic chromatin promotes rapid transcriptional reactivation following mitotic exit. *Molecular cell* **36**, 970-983, doi:10.1016/j.molcel.2009.12.001 (2009).

80 Petruk, S. *et al.* TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell* **150**, 922-933, doi:10.1016/j.cell.2012.06.046 (2012).

81 Liu, H. *et al.* Phosphorylation of MLL by ATR is required for execution of mammalian S-phase checkpoint. *Nature* **467**, 343-346, doi:10.1038/nature09350 (2010).

82 Kotake, Y., Zeng, Y. & Xiong, Y. DDB1-CUL4 and MLL1 mediate oncogene-induced p16INK4a activation. *Cancer research* **69**, 1809-1814, doi:10.1158/0008-5472.CAN-08-2739 (2009).

83 Milne, T. A. *et al.* Menin and MLL cooperatively regulate expression of cyclin-dependent kinase inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 749-754, doi:10.1073/pnas.0408836102 (2005).

84 Lubitz, S., Glaser, S., Schaft, J., Stewart, A. F. & Anastassiadis, K. Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Molecular biology of the cell* **18**, 2356-2366, doi:10.1091/mbc.E06-11-1060 (2007).

85 Tyagi, S., Chabes, A. L., Wysocka, J. & Herr, W. E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Molecular cell* **27**, 107-119, doi:10.1016/j.molcel.2007.05.030 (2007).

86 Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes & development* **21**, 525-530, doi:10.1101/gad.415507 (2007).

87 Vierbuchen, T. & Wernig, M. Molecular roadblocks for cellular reprogramming. *Molecular cell* **47**, 827-838, doi:10.1016/j.molcel.2012.09.008 (2012).

88 Taberlay, P. C. *et al.* Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* **147**, 1283-1294, doi:10.1016/j.cell.2011.10.040 (2011).

89 Heins, N. *et al.* Glial cells generate neurons: the role of the transcription factor Pax6. *Nature neuroscience* **5**, 308-315, doi:10.1038/nn828 (2002).

90 Berninger, B. *et al.* Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 8654-8664, doi:10.1523/JNEUROSCI.1615-07.2007 (2007).

91 Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035-1041, doi:10.1038/nature08797 (2010).

92 Marro, S. *et al.* Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell stem cell* **9**, 374-382, doi:10.1016/j.stem.2011.09.002 (2011).

93 Pang, Z. P. *et al.* Induction of human neuronal cells by defined transcription factors. *Nature* **476**, 220-223, doi:10.1038/nature10202 (2011).

94 Qiang, L. *et al.* Directed conversion of Alzheimer's disease patient skin fibroblasts into functional neurons. *Cell* **146**, 359-371, doi:10.1016/j.cell.2011.07.007 (2011).

95 Yoo, A. S. *et al.* MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* **476**, 228-231, doi:10.1038/nature10323 (2011).

96 Ambasudhan, R. *et al.* Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell stem cell* **9**, 113-118, doi:10.1016/j.stem.2011.07.002 (2011).

97 Chanda, S., Marro, S., Wernig, M. & Sudhof, T. C. Neurons generated by direct conversion of fibroblasts reproduce synaptic phenotype caused by autism-associated neuroligin-3 mutation. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 16622-16627, doi:10.1073/pnas.1316240110 (2013).

98 Pfisterer, U. *et al.* Direct conversion of human fibroblasts to dopaminergic neurons. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10343-10348, doi:10.1073/pnas.1105135108 (2011).

99 Caiazzo, M. *et al.* Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* **476**, 224-227, doi:10.1038/nature10284 (2011).

100 Kim, J. *et al.* Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. *Cell stem cell* **9**, 413-419, doi:10.1016/j.stem.2011.09.011 (2011).

101 Son, E. Y. *et al.* Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell stem cell* **9**, 205-218, doi:10.1016/j.stem.2011.07.014 (2011).

102 Adler, A. F. *et al.* Nonviral direct conversion of primary mouse embryonic fibroblasts to neuronal cells. *Molecular therapy. Nucleic acids* **1**, e32, doi:10.1038/mtna.2012.25 (2012).

103 Meng, F. *et al.* Induction of fibroblasts to neurons through adenoviral gene delivery. *Cell research* **22**, 436-440, doi:10.1038/cr.2011.185 (2012).

104 Ladewig, J. *et al.* Small molecules enable highly efficient neuronal conversion of human fibroblasts. *Nature methods* **9**, 575-578, doi:10.1038/nmeth.1972 (2012).

105 Kim, J. *et al.* Direct reprogramming of mouse fibroblasts to neural progenitors. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7838-7843, doi:10.1073/pnas.1103113108 (2011).

106 Wapinski, O. L. *et al.* Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621-635, doi:10.1016/j.cell.2013.09.028 (2013).

107 Fishman, V. S. *et al.* Cell divisions are not essential for the direct conversion of fibroblasts into neuronal cells. *Cell cycle* **14**, 1188-1196, doi:10.1080/15384101.2015.1012875 (2015).

108 Treutlein, B. *et al.* Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature*, doi:10.1038/nature18323 (2016).

109 Lee, N., Maurange, C., Ringrose, L. & Paro, R. Suppression of Polycomb group proteins by JNK signalling induces transdetermination in Drosophila imaginal discs. *Nature* **438**, 234-237, doi:10.1038/nature04120 (2005).

110 Klebes, A. *et al.* Regulation of cellular plasticity in Drosophila imaginal disc cells by the Polycomb group, trithorax group and lama genes. *Development* **132**, 3753-3765, doi:10.1242/dev.01927 (2005).

111 Zuryn, S. *et al.* Transdifferentiation. Sequential histone-modifying activities determine the robustness of transdifferentiation. *Science* **345**, 826-829, doi:10.1126/science.1255885 (2014).

112 Patel, T., Tursun, B., Rahe, D. P. & Hobert, O. Removal of Polycomb repressive complex 2 makes C. elegans germ cells susceptible to direct conversion into specific somatic cell types. *Cell reports* **2**, 1178-1186, doi:10.1016/j.celrep.2012.09.020 (2012).

113 Davis, F. P. & Eddy, S. R. Transcription factors that convert adult cell identity are differentially polycomb repressed. *PloS one* **8**, e63407, doi:10.1371/journal.pone.0063407 (2013).

114 Testa, G. The time of timing: how Polycomb proteins regulate neurogenesis. *BioEssays : news and reviews in molecular, cellular and developmental biology* **33**, 519-528, doi:10.1002/bies.201100021 (2011).

115 Qian, X. *et al.* Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. *Neuron* **28**, 69-80 (2000).

116 Shen, Q. *et al.* The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. *Nature neuroscience* **9**, 743-751, doi:10.1038/nn1694 (2006).

117 Desai, A. R. & McConnell, S. K. Progressive restriction in fate potential by neural progenitors during cerebral cortical development. *Development* **127**, 2863-2872 (2000).

118 Pereira, J. D. *et al.* Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15957-15962, doi:10.1073/pnas.1002530107 (2010).

119 Hirabayashi, Y. *et al.* Polycomb limits the neurogenic competence of neural precursor cells to promote astrogenic fate transition. *Neuron* **63**, 600-613, doi:10.1016/j.neuron.2009.08.021 (2009).

120 Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell* **30**, 755-766, doi:10.1016/j.molcel.2008.05.007 (2008).

121 Burgold, T. *et al.* The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment. *PloS one* **3**, e3034, doi:10.1371/journal.pone.0003034 (2008).

122 Burgold, T. *et al.* The H3K27 demethylase JMJD3 is required for maintenance of the embryonic respiratory neuronal network, neonatal breathing, and survival. *Cell reports* **2**, 1244-1258, doi:10.1016/j.celrep.2012.09.013 (2012).

123 Lim, D. A. *et al.* Chromatin remodelling factor Mll1 is essential for neurogenesis from postnatal neural stem cells. *Nature* **458**, 529-533, doi:10.1038/nature07726 (2009).

124 Huang, Y. C. *et al.* The epigenetic factor Kmt2a/Mll1 regulates neural progenitor proliferation and neuronal and glial differentiation. *Developmental neurobiology* **75**, 452-462, doi:10.1002/dneu.22235 (2015).

125 Kerimoglu, C. *et al.* Histone-methyltransferase MLL2 (KMT2B) is required for memory formation in mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**, 3452-3464, doi:10.1523/JNEUROSCI.3356-12.2013 (2013).

126 FitzGerald, K. T. & Diaz, M. O. MLL2: A new mammalian member of the trx/MLL family of genes. *Genomics* **59**, 187-192, doi:10.1006/geno.1999.5860 (1999).

127 Huntsman, D. G. *et al.* MLL2, the second human homolog of the Drosophila trithorax gene, maps to 19q13.1 and is amplified in solid tumor cell lines. *Oncogene* **18**, 7975-7984, doi:10.1038/sj.onc.1203291 (1999).

128 Kranz, A. *et al.* An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. *Genesis* **48**, 512-520, doi:10.1002/dvg.20641 (2010).

129 Pool, M., Thiemann, J., Bar-Or, A. & Fournier, A. E. NeuriteTracer: a novel ImageJ plugin for automated quantification of neurite outgrowth. *Journal of neuroscience methods* **168**, 134-139, doi:10.1016/j.jneumeth.2007.08.029 (2008).

130 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).

131 Germain, P. L. *et al.* RNAontheBENCH: computational and empirical resources for benchmarking RNAseq quantification and differential expression methods. *Nucleic acids research* **44**, 5054-5067, doi:10.1093/nar/gkw448 (2016).

132 Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome research* **22**, 2008-2017, doi:10.1101/gr.133744.111 (2012).

133 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).

134 Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology* **11**, R14, doi:10.1186/gb-2010-11-2-r14 (2010).

135 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

136 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

137 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

138 Shen, L. *et al.* diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PloS one* **8**, e65598, doi:10.1371/journal.pone.0065598 (2013).

139 Xu, C. *et al.* Immortalized fibroblast-like cells derived from human embryonic stem cells support undifferentiated cell growth. *Stem cells* **22**, 972-980, doi:10.1634/stemcells.22-6-972 (2004).

140 Huang, Y. C. *et al.* The epigenetic factor Kmt2a/Mll1 regulates neural progenitor proliferation and neuronal and glial differentiation. *Developmental neurobiology*, doi:10.1002/dneu.22235 (2014).

141 Seidenfaden, R., Desoeuvre, A., Bosio, A., Virard, I. & Cremer, H. Glial conversion of SVZ-derived committed neuronal precursors after ectopic grafting into the adult brain. *Molecular and cellular neurosciences* **32**, 187-198, doi:10.1016/j.mcn.2006.04.003 (2006).

142 Davie, J. R., Xu, W. & Delcuve, G. P. Histone H3K4 trimethylation: dynamic interplay with pre-mRNA splicing. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **94**, 1-11, doi:10.1139/bcb-2015-0065 (2016).

143 Podlaha, O., De, S., Gonen, M. & Michor, F. Histone modifications are associated with transcript isoform diversity in normal and cancer cells. *PLoS computational biology* **10**, e1003611, doi:10.1371/journal.pcbi.1003611 (2014).

144 Sims, R. J., 3rd *et al.* Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Molecular cell* **28**, 665-676, doi:10.1016/j.molcel.2007.11.010 (2007).

145 Teoh, P. L. & Sharrocks, A. D. WDR5, ASH2L, and RBBP5 control the efficiency of FOS transcript processing. *Cellular & molecular biology letters* **19**, 215-232, doi:10.2478/s11658-014-0190-8 (2014).

146 Nakanishi, H. *et al.* Neurabin: a novel neural tissue-specific actin filament-binding protein involved in neurite formation. *The Journal of cell biology* **139**, 951-961 (1997).

147 Causeret, F. *et al.* Neurabin-I is phosphorylated by Cdk5: implications for neuronal morphogenesis and cortical migration. *Molecular biology of the cell* **18**, 4327-4342, doi:10.1091/mbc.E07-04-0372 (2007).

148 Orioli, D. *et al.* Rac3-induced neuritogenesis requires binding to Neurabin I. *Molecular biology of the cell* **17**, 2391-2400, doi:10.1091/mbc.E05-08-0753 (2006).

149 Yokoyama, K. *et al.* NYAP: a phosphoprotein family that links PI3K to WAVE1 signalling in neurons. *The EMBO journal* **30**, 4739-4754, doi:10.1038/emboj.2011.348 (2011).

150 Patel, K. G., Liu, C., Cameron, P. L. & Cameron, R. S. Myr 8, a novel unconventional myosin expressed during brain development associates with the protein phosphatase catalytic subunits 1alpha and 1gamma1. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **21**, 7954-7968 (2001).

151 Pettem, K. L. *et al.* The specific alpha-neurexin interactor calsyntenin-3 promotes excitatory and inhibitory synapse development. *Neuron* **80**, 113-128, doi:10.1016/j.neuron.2013.07.016 (2013).

152 Nakayama, M., Kikuno, R. & Ohara, O. Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. *Genome research* **12**, 1773-1784, doi:10.1101/gr.406902 (2002).

153 Morimura, N., Inoue, T., Katayama, K. & Aruga, J. Comparative analysis of structure, expression and PSD95-binding capacity of Lrfn, a novel family of neuronal transmembrane proteins. *Gene* **380**, 72-83, doi:10.1016/j.gene.2006.05.014 (2006).

154 Zheng, Y. L., Li, B. S., Veeranna & Pant, H. C. Phosphorylation of the head domain of neurofilament protein (NF-M): a factor regulating topographic phosphorylation of NF-M tail domain KSP sites in neurons. *The Journal of biological chemistry* **278**, 24026-24032, doi:10.1074/jbc.M303079200 (2003).

155 Wong, T. H. *et al.* PRKAR1B mutation associated with a new neurodegenerative disorder with unique pathology. *Brain : a journal of neurology* **137**, 1361-1373, doi:10.1093/brain/awu067 (2014).

156 Kakizawa, S., Moriguchi, S., Ikeda, A., Iino, M. & Takeshima, H. Functional crosstalk between cell-surface and intracellular channels mediated by junctophilins essential for neuronal functions. *Cerebellum* **7**, 385-391, doi:10.1007/s12311-008-0040-1 (2008).

157 Kwon, S. E. & Chapman, E. R. Synaptophysin regulates the kinetics of synaptic vesicle endocytosis in central neurons. *Neuron* **70**, 847-854, doi:10.1016/j.neuron.2011.04.001 (2011).

158 Gordon, S. L., Leube, R. E. & Cousin, M. A. Synaptophysin is required for synaptobrevin retrieval during synaptic vesicle endocytosis. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**, 14032-14036, doi:10.1523/JNEUROSCI.3162-11.2011 (2011).

159 McMahon, H. T. *et al.* Synaptophysin, a major synaptic vesicle protein, is not essential for neurotransmitter release. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 4760-4764 (1996).

160 Schmitt, U., Tanimoto, N., Seeliger, M., Schaeffel, F. & Leube, R. E. Detection of behavioral alterations and learning deficits in mice lacking synaptophysin. *Neuroscience* **162**, 234-243, doi:10.1016/j.neuroscience.2009.04.046 (2009).

161 Minchin, R. F., Lewis, A., Mitchell, D., Kadlubar, F. F. & McManus, M. E. Sulfotransferase 4A1. *The international journal of biochemistry & cell biology* **40**, 2686-2691, doi:10.1016/j.biocel.2007.11.010 (2008).

162 Nishimura, T., Kubosaki, A., Ito, Y. & Notkins, A. L. Disturbances in the secretion of neurotransmitters in IA-2/IA-2beta null mice: changes in behavior, learning and lifespan. *Neuroscience* **159**, 427-437, doi:10.1016/j.neuroscience.2009.01.022 (2009).

163 Takeyama, N. *et al.* Localization of insulinoma associated protein 2, IA-2 in mouse neuroendocrine tissues using two novel monoclonal antibodies. *Life sciences* **84**, 678-687, doi:10.1016/j.lfs.2009.02.012 (2009).

164 Kakunaga, S. *et al.* Nectin-like molecule-1/TSLL1/SynCAM3: a neural tissue-specific immunoglobulin-like cell-cell adhesion molecule localizing at non-junctional contact sites of presynaptic nerve terminals, axons and glia cell processes. *Journal of cell science* **118**, 1267-1277, doi:10.1242/jcs.01656 (2005).

165 Chen, Y. *et al.* Nuclear translocation of PDCD5 (TFAR19): an early signal for apoptosis? *FEBS letters* **509**, 191-196 (2001).

166 Chen, L. N., Wang, Y., Ma, D. L. & Chen, Y. Y. Short interfering RNA against the PDCD5 attenuates cell apoptosis and caspase-3 activity induced by Bax overexpression. *Apoptosis : an international journal on programmed cell death* **11**, 101-111, doi:10.1007/s10495-005-3134-y (2006).

167 Nathan, B. P., Barsukova, A. G., Shen, F., McAsey, M. & Struble, R. G. Estrogen facilitates neurite extension via apolipoprotein E in cultured adult mouse cortical neurons. *Endocrinology* **145**, 3065-3073, doi:10.1210/en.2003-1707 (2004).

168 Hussain, A., Luong, M., Pooley, A. & Nathan, B. P. Isoform-specific effects of apoE on neurite outgrowth in olfactory epithelium culture. *Journal of biomedical science* **20**, 49, doi:10.1186/1423-0127-20-49 (2013).

169 Karle, K. N., Mockel, D., Reid, E. & Schols, L. Axonal transport deficit in a KIF5A(-/-) mouse model. *Neurogenetics* **13**, 169-179, doi:10.1007/s10048-012-0324-y (2012).

170 Cowper, A. E., Caceres, J. F., Mayeda, A. & Screaton, G. R. Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing. *The Journal of biological chemistry* **276**, 48908-48914, doi:10.1074/jbc.M103967200 (2001).

171 Dwane, S., Durack, E., O'Connor, R. & Kiely, P. A. RACK1 promotes neurite outgrowth by scaffolding AGAP2 to FAK. *Cellular signalling* **26**, 9-18, doi:10.1016/j.cellsig.2013.08.036 (2014).

172 Lai, M. *et al.* Myosin X regulates neuronal radial migration through interacting with N-cadherin. *Frontiers in cellular neuroscience* **9**, 326, doi:10.3389/fncel.2015.00326 (2015).

173 Santiago-Medina, M., Gregus, K. A., Nichol, R. H., O'Toole, S. M. & Gomez, T. M. Regulation of ECM degradation and axon guidance by growth cone invadosomes. *Development* **142**, 486-496, doi:10.1242/dev.108266 (2015).

174 Li, C. M. *et al.* Differential Tks5 isoform expression contributes to metastatic invasion of lung adenocarcinoma. *Genes & development* **27**, 1557-1567, doi:10.1101/gad.222745.113 (2013).

175 Cejudo-Martin, P. *et al.* Genetic disruption of the sh3pxd2a gene reveals an essential role in mouse development and the existence of a novel isoform of tks5. *PloS one* **9**, e107674, doi:10.1371/journal.pone.0107674 (2014).

176 Wang, X., Xu, Q., Bey, A. L., Lee, Y. & Jiang, Y. H. Transcriptional and functional complexity of Shank3 provides a molecular framework to understand the phenotypic heterogeneity of SHANK3 causing autism and Shank3 mutant mice. *Molecular autism* **5**, 30, doi:10.1186/2040-2392-5-30 (2014).

177 Perrelet, D. *et al.* Motoneuron resistance to apoptotic cell death in vivo correlates with the ratio between X-linked inhibitor of apoptosis proteins (XIAPs) and its inhibitor, XIAP-associated factor 1. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **24**, 3777-3785, doi:10.1523/JNEUROSCI.0413-04.2004 (2004).

178 Leaman, D. W. *et al.* Identification of X-linked inhibitor of apoptosis-associated factor-1 as an interferon-stimulated gene that augments TRAIL Apo2L-induced

apoptosis. *The Journal of biological chemistry* **277**, 28504-28511, doi:10.1074/jbc.M204851200 (2002).

179 Linqing, Z. *et al.* Runx1t1 regulates the neuronal differentiation of radial glial cells from the rat hippocampus. *Stem cells translational medicine* **4**, 110-116, doi:10.5966/sctm.2014-0158 (2015).

180 Kehrl, J. M. *et al.* Gain-of-function mutation in Gnao1: a murine model of epileptiform encephalopathy (EIEE17)? *Mammalian genome : official journal of the International Mammalian Genome Society* **25**, 202-210, doi:10.1007/s00335-014-9509-z (2014).

181 Hao, N., Bhakti, V. L., Peet, D. J. & Whitelaw, M. L. Reciprocal regulation of the basic helix-loop-helix/Per-Arnt-Sim partner proteins, Arnt and Arnt2, during neuronal differentiation. *Nucleic acids research* **41**, 5626-5638, doi:10.1093/nar/gkt206 (2013).

182 Chou, D. K., Schachner, M. & Jungalwala, F. B. HNK-1 sulfotransferase null mice express glucuronyl glycoconjugates and show normal cerebellar granule neuron migration in vivo and in vitro. *Journal of neurochemistry* **82**, 1239-1251 (2002).

183 Kim, J. Y. *et al.* Different expression patterns of Phactr family members in normal and injured mouse brain. *Neuroscience* **221**, 37-46, doi:10.1016/j.neuroscience.2012.06.059 (2012).

184 Kaya, L. *et al.* Direct association of the reticulon protein RTN1A with the ryanodine receptor 2 in neurons. *Biochimica et biophysica acta* **1833**, 1421-1433, doi:10.1016/j.bbamcr.2013.02.012 (2013).

185 Melino, S., Nepravishta, R., Bellomaria, A., Di Marco, S. & Paci, M. Nucleic acid binding of the RTN1-C C-terminal region: toward the functional role of a reticulon protein. *Biochemistry* **48**, 242-253, doi:10.1021/bi801407w (2009).

186 Arranz, V. *et al.* Human and mouse MOK2 proteins are associated with nuclear ribonucleoprotein components and bind specifically to RNA and DNA through their zinc finger domains. *Molecular and cellular biology* **17**, 2116-2126 (1997).

187 Smith, B. N. *et al.* Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron* **84**, 324-331, doi:10.1016/j.neuron.2014.09.027 (2014).

188 Negishi, M. *et al.* A novel zinc finger protein Zfp277 mediates transcriptional repression of the Ink4a/arf locus through polycomb repressive complex 1. *PloS one* **5**, e12373, doi:10.1371/journal.pone.0012373 (2010).

189 Parra, L. M. & Zou, Y. Sonic hedgehog induces response of commissural axons to Semaphorin repulsion during midline crossing. *Nature neuroscience* **13**, 29-35, doi:10.1038/nn.2457 (2010).

190 Ueda, S., Negishi, M. & Katoh, H. Rac GEF Dock4 interacts with cortactin to regulate dendritic spine formation. *Molecular biology of the cell* **24**, 1602-1613, doi:10.1091/mbc.E12-11-0782 (2013).

191 Ueda, S., Fujimoto, S., Hiramoto, K., Negishi, M. & Katoh, H. Dock4 regulates dendritic development in hippocampal neurons. *Journal of neuroscience research* **86**, 3052-3061, doi:10.1002/jnr.21763 (2008).

192 Xiao, Y. *et al.* The atypical guanine nucleotide exchange factor Dock4 regulates neurite differentiation through modulation of Rac1 GTPase and actin dynamics. *The Journal of biological chemistry* **288**, 20034-20045, doi:10.1074/jbc.M113.458612 (2013).

193 Pavlopoulos, E. *et al.* Neuralized1 activates CPEB3: a function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. *Cell* **147**, 1369-1383, doi:10.1016/j.cell.2011.09.056 (2011).

194 Lai, E. C. & Rubin, G. M. neuralized functions cell-autonomously to regulate a subset of notch-dependent processes during adult Drosophila development. *Developmental biology* **231**, 217-233, doi:10.1006/dbio.2000.0124 (2001).

195 Lai, E. C., Deblandre, G. A., Kintner, C. & Rubin, G. M. Drosophila neuralized is a ubiquitin ligase that promotes the internalization and degradation of delta. *Developmental cell* **1**, 783-794 (2001).

196 Yang, Y. J. *et al.* Menin mediates epigenetic regulation via histone H3 lysine 9 methylation. *Cell death & disease* **4**, e583, doi:10.1038/cddis.2013.98 (2013).

197 Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11667-11672, doi:10.1073/pnas.0904715106 (2009).

198 Senturk, A., Pfennig, S., Weiss, A., Burk, K. & Acker-Palmer, A. Ephrin Bs are essential components of the Reelin pathway to regulate neuronal migration. *Nature* **472**, 356-360, doi:10.1038/nature09874 (2011).

199 Wang, X. *et al.* Histone H3K4 methyltransferase Mll1 regulates protein glycosylation and tunicamycin-induced apoptosis through transcriptional regulation. *Biochimica et biophysica acta* **1843**, 2592-2602, doi:10.1016/j.bbamcr.2014.06.013 (2014).

200 Piunti, A. *et al.* Polycomb proteins control proliferation and transformation independently of cell cycle checkpoints by regulating DNA replication. *Nature communications* **5**, 3649, doi:10.1038/ncomms4649 (2014).

201 Gascon, S. *et al.* Identification and Successful Negotiation of a Metabolic Checkpoint in Direct Neuronal Reprogramming. *Cell stem cell* **18**, 396-409, doi:10.1016/j.stem.2015.12.003 (2016).