Unravelling the role of NANOG in the regulation of priming gene expression

Raquel Maria Ramos Calçada

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors

Professor Doctor Domingos Manuel Pinto Henrique

Professor Doctor Maria Margarida Fonseca Rodrigues Diogo

Examination Committee

Chairperson: Professor Doctor João Pedro Estrela Rodrigues Conde

Supervisor: Professor Doctor Domingos Manuel Pinto Henrique

Member of the Committee: Doctor Simão José Teixeira da Rocha

November 2016

Acknowledgements

First, I would like to thank my supervisor Prof. Dr. Domingos Henrique for all his challenges, for encouraging me to never stop looking for order in the chaos of pluripotency and for teaching me how to think critically. To my other supervisor, Prof. Dr. Margarida Diogo, my gratitude for being such an inspiring teacher and model scientist.

I am also thankful to Ana Guedes for everything she taught me about *Nanog*, stem cell culture and, of course, our beloved smFISH. A special acknowledgment to all DHenrique lab members: Sara Ferreira, for our awakening Nespresso coffees, and João Carreira, for the magnificent Azeitão cheese and peanut butter Magnums! I am also thankful to Andreia Pereira, for her huge availability and kindness for helping me with the smFISH challenges, and to Jorge Martins, my biomedical and supportive partner on the jungle of development.

I am deeply thankful to the Bioimaging team (António Temudo, Ana Nascimento and José Rino) for their precious assistance given on the transition between microscopes. Gonçalo Fernandes and Carolina Leitão, without your precious help in smFISH data analysis I would have never finish my thesis on time. May the "*Nanog* Fishers" never end!

Nuno Mendes, Beatriz Almeida, Joana Faria, Ricardo Trindade, Rui Lourenço, Nuno Matias, Mariana Ferreira, Elisa Pacheco, Raquel Aguiar, and many other biomedical fellows, it has been a long journey we have taken together and now we have finally reached our academic goal. It was a pleasure to share these last 5 years with you and I hope our friendship never ends. Leonardo Filipe, thank you for your endless patience and for being my very own motivational source to never give up and to always try to be a better version of myself.

To Teresa Cardoso and Mariana Machado, for our long friendship and happy relaxed moments in Sesimbra. Last but not the least, to my family: mum, dad, sister, grandmothers, grandfathers, Zézinha and uncle. Thank you for the unconditional support throughout my academic path and for all the passion and dedication you taught me to have in the things I do. If I arrived so far, it is due to you.

Abstract

The transcription factor NANOG exhibits a heterogeneous expression in pluripotent mouse embryonic stem (mES) cells, both at mRNA and protein levels. This variety appears to arise from stochastic fluctuations in NANOG expression in individual mES cells, creating windows of opportunity to explore pluripotency. Low-NANOG cells, in a “lineage-primed” state, are more susceptible to commit to differentiation and express higher levels of lineage-affiliated genes than high-NANOG cells.

However, it is not clear how NANOG controls the exit of mES cells from pluripotency. Recent studies revealed that NANOG might interact with polycomb repressive complex 2 (PRC2) and ten-eleven translocation 1 (TET1), responsible for H3K27 trimethylation and DNA demethylation, respectively. Moreover, genes upregulated in low-NANOG cells, denoted as priming genes, which include lineage-affiliated genes, are enriched for binding signatures in PRC2. Thus, it was hypothesized that NANOG represses priming gene expression by regulating PRC2 and TET1, but how?

To unravel these mechanisms, the cell lines E14tg2a and Nd (with a Nanog:VNP reporter) were cultured in pluripotent “Serum/LIF” conditions. mES cells were incubated with GSK343, an inhibitor of PRC2 activity, and Ascorbic Acid (AA), a promoter of TET1. Afterwards, the effects on the expression of priming genes were evaluated using single-molecule RNA FISH. In accordance with preliminary data, now confirmed, GSK343 increases lineage-affiliated gene expression in high-*Nanog* cells, whilst AA decreases. Thus, it is proposed that NANOG forms a complex with TET1, which by maintaining a hypomethylated state, contributes to PRC2 recruitment and, consequently, to the silencing of lineage-affiliated gene expression.

Keywords

NANOG; embryonic stem cells; pluripotency; heterogeneity; PRC2; TET1.

Resumo

O fator de transcrição NANOG exibe uma expressão heterogénea nas células estaminais embrionárias de murganho (CEE). Esta variabilidade surge de flutuações estocásticas em cada célula, o que cria “janelas de oportunidade” para estas explorarem a pluripotência. As células baixo-NANOG (“lineage-primed”) são mais suscetíveis a diferenciarem e exibem uma maior expressão de “genes linhagem”, do que as células elevado-NANOG.

Contudo, desconhece-se como o NANOG controla a saída da pluripotência das CEE. Estudos revelaram que o NANOG interage com o Polycomb complexo repressivo 2 (PRC2) e a translocação ten-eleven 1 (TET1), responsáveis pela trimetilação do H3H27 e pela demetilação do DNA, respetivamente. Além disso, os genes “upregulated” nas células baixo-NANOG (“genes priming”, que incluem os de linhagem) ligam-se ao PRC2. Deste modo, formulou-se a hipótese de que o NANOG reprime a expressão de “genes priming” através da regulação de PRC2 e TET1, mas como?

De forma a desvendar estes mecanismos, as linhas celulares E14tg2a e Nd (com repórter Nanog:VNP) foram cultivadas em “Soro/LIF”. As CEE foram incubadas com GSK343, um inibidor do PRC2, e ácido ascórbico (AA), um promotor das TET1. De seguida, a expressão dos “genes priming” foi avaliada através de single-molecule RNA FISH. De acordo com resultados preliminares, agora confirmados, GSK343 aumenta a expressão de “genes linhagem” nas células elevado-*Nanog*, enquanto AA diminui. Desta forma, propõem-se que o NANOG forme um complexo com a TET1, que ao manter um estado hipometilado, contribui para o recrutamento de PRC2 e para a repressão da expressão dos “genes linhagem”.

Palavras-chave

NANOG; células estaminais embrionárias; pluripotência; heterogeneidade; PRC2; TET1.

Table of Contents

Acknowledgements ... **iii**

Abstract ... **v**

Resumo ... **vii**

List of Tables ... **xi**

List of Figures ... **xii**

List of Abbreviations ... **xiv**

1. Introduction ... **1**

1.1 Early Murine Embryonic Development ... *1*

1.1.1 Cell-Fate Choices in the Preimplantation Embryo ... 1

1.1.2 Heterogeneity Creates Possibilities ... 2

1.2 Mouse Embryonic Stem Cells ... *5*

1.2.1 Capturing Pluripotency ... 5

1.2.2 “Naïve” and “Primed” Pluripotency ... 5

1.2.3 Signalling Pathways ... 7

1.3 Gene Regulatory Network Governing Pluripotency ... *8*

1.3.1 *Nanog-Oct4-Sox2* Complex: from Embryo to mES Cells ... 8

1.3.2 Heterogeneity and Lineage Priming in mES Cells ... 10

1.3.3 Stochasticity in mES Cells ... 12

1.4 Chromatin Dynamics in Pluripotency ... *14*

1.4.1 DNA Methylation and Histone Modifications ... 14

1.4.2 NANOG Partners: PRC2 and TET1 ... 15

1.4.3 The Model for NANOG Regulation of Priming Gene Expression ... 19

2. Motivation and Research Aims ... **21**

3. Materials and Methods ... **22**

3.1 Materials ... *22*

3.1.1 Mouse Embryonic Stem Cells ... 22

3.1.2 Reagents ... 22

3.2 Methods ... *26*

3.2.1 Cell Culture ... 26

3.2.1.1 Expansion of Undifferentiated Mouse Embryonic Stem Cells ... 26

3.2.1.2 *Mycoplasma* Test ... 27

3.2.1.3 Chemical Modulators Assay 28
3.2.2 Protein Expression Analysis 29
3.2.2.1 Flow Cytometry Analysis 29
3.2.2.2 Intracellular Stainning - Flow Cytometry (IC-FC) 29
3.2.2.3 Fluorescence Activated Cell Sorting (FACS) 30
3.2.3 Single Molecule RNA Fluorescence *In Situ* Hybridization (smFISH) 31

4. Results and Discussion 34

4.1 Nanog Characterization and Lineage Priming 34
4.1.1 *Nanog* in Pluripotency 34
4.1.2 Analysis of Priming Gene Expression 38

4.2 Nanog Regulation of Priming Gene Expression through PRC2 and TET1 44
4.2.1 Analysis of cell morphology and dynamics of Nanog:VNP 44
4.2.2 Verification of GSK343 Effect 46
4.2.3 Analysis of Priming Gene Expression with GSK343 and/or AA 47

4.3 Sorting into Low and High-Nanog:VNP Cells 52
4.3.1 Analysis of cell morphology and dynamics of Nanog:VNP 52
4.3.2 Expected smFISH Results from VNP_L and VNP_H Cells 56

5. Conclusions 57

6. Future Work 61

7. References 62

8. Annex A A-1

List of Tables

Table 3.1 – List of reagents used in the experiments performed in this project. 22
Table 3.2 – List of solutions/media and respective components which were used in the experiments performed in this project. 23
Table 3.3 – List of chemical modulators used during cell culture in this project. 24
Table 3.4 – List of antibodies used for intracellular staining flow cytometry experiments performed in this project. 24
Table 3.5 – List of smRNA-FISH probes used in the experiments performed in this project. 24
Table 3.6 – Optical filters for RNA detection in smFISH, used in the widefield microscope Zeiss Axio Observer. 25
Table 3.7 – Characteristics of the primers used for *Mycoplasma* PCR. 27
Table 3.8 – Concentration of chemical modulators for WS and final concentration in culture medium. 28
Table 3.9 – Dilutions of the antibodies used for IC-FC. 30
Table 4.1 – Statistical analysis of *Car2*, *Nanog* and *Sox3* mRNA expression. 50
Table 4.2 – Percentages of low- and high-*Nanog* cells that express high levels of *Car2* (on the left) and *Sox3* (on the right) transcripts for the different conditions. 51
Table 4.3 - Expected results after smFISH analysis of FACS-sorted VNP_L cells, after 48 hours of exposure to GSK343 and/or AA. 56

Table A1 – Threshold for each gene analysed by smFISH. A-2
Table A2 – Summary of data analysis obtained by smFISH for *Car2-Nanog-Sox3* and *Nanog-Otx2-Sox3* probe combinations. A-3
Table A3 – Summary of data analysis obtained by smFISH upon exposure to chemical modulators. A-5

List of Figures

Figure 1.1 – Overview of early mouse embryonic development.. 1
Figure 1.2 – Model for specification of TE and ICM, at the 16- to 32-cell transition, in the compact morula.. 3
Figure 1.3 – Model for specification of EPI and PrE, within ICM cells. 4
Figure 1.4 – Differences *in vivo* and *in vitro* between naïve and “primed” pluripotency states.. 6
Figure 1.5 – NANOG heterogeneity. 10
Figure 1.6 – NANOG as a “differentiation rheostat”. 12
Figure 1.7 – Nonbursty versus bursty transcription models. 13
Figure 1.8 – PRC2 role as a transcriptional repressor. 16
Figure 1.9 – Dual roles of TET1 in ES cells. 19
Figure 1.10 – Proposed model for NANOG regulation of lineage-affiliated gene expression. 20
Figure 3.1 – Flow cytometry analysis workflow, in order to obtain the percentages of Nanog:VNP positive cells. 29
Figure 3.2 – smFISH workflow. 32
Figure 4.1 – Scheme of the experiment procedure for ES cell culture in “Serum/LIF” conditions. 34
Figure 4.2 – Nanog:VNP expression in Nd cells.. 35
Figure 4.3 – Analysis of *Sox2* and *Nanog* mRNA expression in E14 mES cells cultured in “Serum/LIF”. 36
Figure 4.4 – Analysis of *Nanog* and *Sox2* expression relatively to their low and high abundance of transcripts in 1349 E14 mES cells. 37
Figure 4.5 – Analysis of gene expression in E14 mES cells cultured in “Serum/LIF”. 39
Figure 4.6 – Correlation analysis of priming gene expression in E14 mES cells cultured in “Serum/LIF” conditions.. 40
Figure 4.7 – Priming gene expression in E14 mES cells cultured in “Serum/LIF”. 41
Figure 4.8 – Correlation analysis of gene expression in E14 mES cells cultured in “Serum/LIF” conditions.. 42
Figure 4.9 – High-*Otx2* expression is equally distributed between high-*Nanog* and low-*Nanog* cells.. 43
Figure 4.10 – Scheme of the experiment procedure for ES cell culture in "Serum/LIF" conditions supplemented with GSK343 and/or AA.. 44
Figure 4.11 – Analysis of the effect of GSK343 and/or AA in mES cells cultured in "Serum/LIF". 45
Figure 4.12 – H3K27me3 positive cells in mES cells grown in "Serum/LIF" supplemented with chemical modulators. 46
Figure 4.13 – Histograms of the distribution of mRNA transcripts for *Car2*, *Nanog* and *Sox3* for different conditions. 50
Figure 4.14 – Scheme of the experiment procedure for the sorting of Nd mES cells into low-Nanog:VNP (VNP_L) and high-Nanog:VNP (VNP_H) subpopulations. 52
Figure 4.15 – Morphology of the sorted cells with chemical modulators. 52
Figure 4.16 – Nanog expression in FACS-sorted Nd mES cells. 53

Figure A1 – Cell morphology of E14 (on the left) and Nd (on the right) mES cells. A-1
Figure A2 – Fold increase and viability calculated for Nd and E14 mES cells. A-1
Figure A3 – Analysis of the Nanog:VNP reporter expression in Nd cells by flow cytometry.. .....A-1
Figure A4 – Correlation scatterplots for the three possible combinations between *Nanog*, *Sox3* and *Car2* within the different conditions. A-4

List of Abbreviations

Abbreviation	Definition
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmc	5-hydroxymethylcytosine
5mC	5-methylcytosine
AA	Ascorbic acid
aPKC	atypical protein kinase C
BAC	Bacterial artificial chromosome
BER	Base excision repair
CGI	CpG Island
ChEA	Chromatin enrichment analysis
CpG	-C-phosphate-G-
CV	Coefficient of variation
DAPI	4', 6-diamino-2-phenylindole dilactate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E	Embryonic day
EC	Embryonal carcinoma
EG	Embryonic germ
EPI	Embryonic Epiblast
EpiS	Epiblast stem
ES	Embryonic stem
ExE	Extraembryonic ectoderm
FACS	Fluorescence activated cell sorting
FBS	Fetal bovine serum
FF	Fano factor
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FI	Fold increase
FISH	Fluorescence *in situ* hybridization
FSS	Forward scatter signal
GAPDH	Glyceraldheyde-3-phosphate dehydrogenase
GMEM	Glasgow minimum essential medium
H3K27me3	Trimethylation at lysine 27 of histone H3
H3K4me1	Monomethylation at lysine 4 of histone H3
H3K4me3	Trimethylation at lysine 4 of histone H3
hES	Human embryonic stem

ICM	Inner Cell Mass
iPS	Induced pluripotent stem
JAK	Janus kinase
LIF	Leukemia inhibitor factor
Max	Maximum
MEF	Mouse embryonic fibroblast
mES	Mouse embryonic stem
mg	Milligram
Min	Minimum
mL	Millilitre
MLL	Mixed lineage leukemia protein
mRNA	Messenger RNA
N or Ncell	Cell number
Nd	Nanog dynamics – Nanog:VNP reporter cell line
nM	Nano molar
NMPs	Neuromesodermal progenitors
NOS	Nanog, Oct4 and Sox2
p-value	Probability value
ParE	Parietal endoderm
PBS	Phosphate buffer saline
PcG	Polycomb group proteins
PCR	Polymerase chain reaction
PrE	Primitive endoderm
PGC	Primordial germ cell
PI	Propidium iodide
PRC	Polycomb repressive complex
R	Spearman correlation coefficient
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
rpm	Rotation per minute
RT	Room temperature
shRNA	Short hairpin RNA
siRNA	Small interfering RNA
smFISH	Single molecule RNA fluorescence *in situ* hybridisation
SS	Side scatter
STAT	Signal transducer and activator of transcription
TDG	Thymine DNA glycosylase
TE	Trophectoderm
TET	Ten-eleven translocation
VE	Visceral endoderm

VNP	Venus-NLS-PEST
VNP_H	High-Nanog:VNP
VNP_L	Low-Nanog:VNP
α-KG	Alpha-ketoglutarate
µM	Micro molar
-/-	Homozygous knock-out

1. Introduction

1.1 Early Murine Embryonic Development

"The concept of an embryo is a staggering one, and forming an embryo is the hardest thing you will ever do. To become an embryo, you had to build yourself from a single cell (…) One of the critical differences between you and a machine is that a machine is never required to function until after it is built. Every animal has to function as it builds itself." (Gilbert 2003)

1.1.1 Cell-Fate Choices in the Preimplantation Embryo

Mammalian embryogenesis starts with the formation of the zygote, as a result of the fertilization between male and female germ cells. The zygote is the original totipotent stem cell, with the ability to generate all embryonic and extraembryonic tissues of an organism. Prior to embryo implantation in the mother's uterus, there is a sequence of consecutive divisions, also known as cleavage, of the fertilized egg into smaller cells, the blastomeres. When a 16-cell stage is reached, at E2.5 (E depicts embryonic day), the embryo is called morula. Afterwards, the two preimplantation cell-fate decisions take place and give rise to the late blastocyst, which will implant on the maternal uterus (Figure 1.1). Thus, this preimplantation period accommodates the transition from zygote's totipotency to cells with progressive restricted developmental potential, as differentiation proceeds.

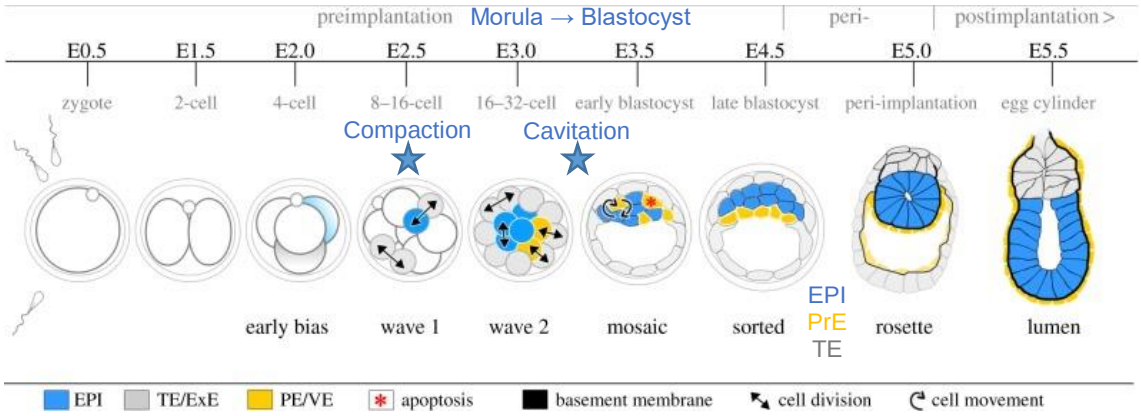


Figure 1.1 – Overview of early mouse embryonic development. *Pre, peri and postimplantation stages are represented, as well as the first cell-fate decision by the distinction between ICM and TE, and the second cell-fate decision by the separation between EPI and PE within ICM cells. EPI – epiblast; ExE – extraembryonic ectoderm; PE or PrE – primitive endoderm; TE – trophectoderm; VE – visceral endoderm. Adapted from Bedzhov* et al. *2014.*

The first cell-fate decision involves two waves of asymmetric cell divisions at E2.5 and E3.0, associated with two morphogenetic events: compaction and cavitation. Firstly, the compaction of the 8-cell embryo, through the development of intercellular junctions, produces the morula. The asymmetric divisions, immediately after compaction, position cells inside and outside of the two-layered morula, and their distinct spatial positions correlate with their fates. The outer layer of the morula will differentiate into extraembryonic trophectoderm (TE), also known as trophoblast, which will generate the placenta. Meanwhile, the inside morula cells will constitute the pluripotent inner cell mass (ICM). At the 32-cell

stage, the morula becomes the blastocyst and the TE will form the blastocyst cavity, which determines the embryonic-abembryonic axis of the embryo, through a process called cavitation. Consequently, the early preimplantation blastocyst, at E3.5, is composed of ICM and TE (reviewed in Saiz & Plusa 2013; Bedzhov *et al.* 2014).

The second cell-fate decision occurs in ICM cells, separating them into pluripotent embryonic epiblast (EPI) and into differentiating extraembryonic primitive endoderm (PrE), also known as hypoblast, which develops into the visceral endoderm (VE) and parietal endoderm (ParE) after implantation, precursors of the yolk sac. Both TE and PrE contribute to extraembryonic tissues, which support the fetus development through the uptake of nutrients, exchange of gases and disposal of waste products from the embryo. Contrarily, the preimplantation EPI consists of pluripotent cells with the potential to differentiate into all three germ layers (endoderm, mesoderm and ectoderm) and the germline of the embryo. These cell-fate choices depend on the cell's spatial location, polarity and signalling. Therefore, at the time of implantation (E5.0), the late blastocyst presents three different cell lineages: TE, PrE and EPI. In the next 24 hours, the blastocyst rapidly invades the maternal tissues, proliferates and transforms into an egg cylinder (reviewed in Saiz & Plusa 2013; Bedzhov *et al.* 2014).

The development in the preimplantation mouse embryo is regulative, meaning that, until the 32-cell stage, it is capable to adapt and compensate for anomalies in position and cell number. This plasticity or fate-modulation is shown by the formation of blastocyst in chimaeras and even when blastomeres are removed, added or rearranged. In fact, heterogeneous gene expression in blastomeres enables a flexible lineage-affiliated gene regulatory network, under the control of mechanical and chemical cues, which ensures a correct lineage-specification (Martinez Arias *et al.* 2013).

1.1.2 Heterogeneity Creates Possibilities

How can a cell know it is different from its neighbour? According to the simplified embryo position model, outside cells would differentiate into TE and inside cells into ICM, besides, outer ICM cells would transform into PrE and deep ICM cells into EPI. Nevertheless, position is not the only factor influencing cell-fate decisions. For instance, cells can be already genetically biased towards their specification, before changing their position, which opens the possibility of differential gene expression being the cause of spatial movement. This hypothesis is supported by the finding that precursors of EPI and PE are initially mixed within the ICM before being sorted into their positions (reviewed in Bedzhov *et al.* 2014). Furthermore, embryo development is also driven by the action of transcription factors (TFs), regulatory proteins that bind to DNA *cis*-regulatory sequences, associated with specific genes, leading to the activation or repression of transcription. It will be further described the role of TFs, position, polarity and signalling pathways on the two cell-fate decisions of the preimplantation embryo.

First Cell-Fate Decision: ICM vs. TE

At the 8-cell stage, all blastomeres occupy equivalent positions. Afterwards, through compaction, blastomeres adhere tightly to each other and become polarized along their apical-basal axis. Inside blastomeres lose polarization and form uniform cell-cell contacts, whilst outside blastomeres

remain polarized and present asymmetric cell-cell contacts. The activation of the TE programme in outside cells is regulated by the Hippo pathway and the transcription factor TEAD4. TEAD4 activity requires the two transcriptional co-activators YAP and TAZ, which are negatively regulated by the Hippo signalling pathway kinase LATS1/2.

Outside cells, which are more prone to differentiate into TE, possess polarity proteins, such as PAR3, PAR6 and atypical protein kinase C (aPKC), at the apical domain. These proteins inhibit LATS from the Hippo pathway, causing the de-repression of YAP and TAZ. As a result, they enter the nucleus and co-activate TEAD4, which will in turn activate the transcription of *Cdx2* and "switch on" the TE programme. Moreover, the TE fate is not only initiated but also maintained by *Cdx2*, which positively regulates its expression and inhibits the expression of pluripotency genes (*Nanog* and *Oct4*).

On the contrary, inside cells, precursors of the pluripotent ICM, display uniform cell-cell contact and a basal domain. The Hippo pathway is active, YAP and TAZ cannot reach the nucleus because they are inhibited by LATS. Consequently, TEAD4 activity is "switched off" and its TE-target genes are not transcribed. Hence, OCT4 expression is promoted and the default pluripotent programme prevails (Figure 1.2).

Position and polarity differences therefore lead to distinct gene regulatory networks, which ultimately end in mutually exclusive cell fates (reviewed in Saiz & Plusa 2013; Schrode *et al.* 2013).

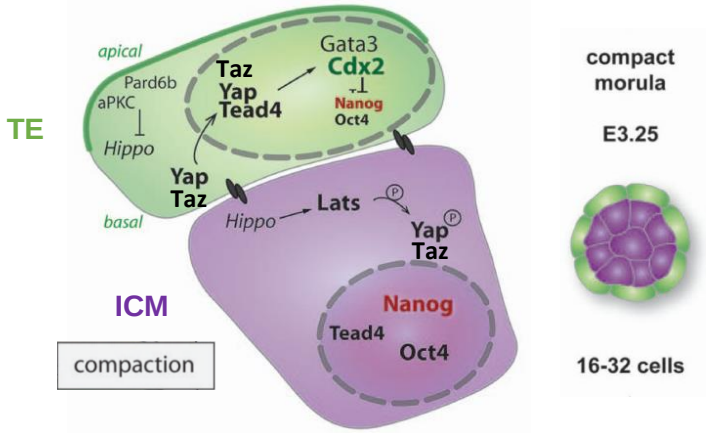


Figure 1.2 – Model for specification of TE and ICM, at the 16- to 32-cell transition, in the compact morula*. Outer cells are polarized, causing the inactivation of the Hippo pathway. Therefore, YAP and TAZ are de-repressed, go to the nucleus and co-activate TEAD4, "switching on" the TE programme. On the other hand, the Hippo pathway is active in unpolarised inner cells, making YAP and TAZ only cytoplasmic, preventing the co-activation of TEAD4 and thereby promoting the default pluripotency ICM programme. Adapted from Schrode* et al. *2013.*

Second Cell-Fate Decision: PrE vs. EPI

At E3.5, ICM cells already express PrE genes (*Gata6*) or EPI genes (*Nanog*) in a mosaic "salt and pepper" manner. Subsequently, the PrE and EPI precursors are sorted into the proper position for each fate by cell migration, positional induction and apoptosis of incorrectly positioned cells.

Alternatively, a recent study with live-cell imaging tracing (Morris *et al.* 2010) proposed that, at cavitation, cells from the first wave of asymmetric cell divisions are internalized earlier and occupy the deeper ICM compartment, being more susceptible to generate EPI than the "later" cells, from the second wave, which give rise to PrE, that can be distinguished as an epithelium at ICM surface. Thus, at E4.5,

the inside ICM cells constitute the EPI while the outside ICM cells compose the PrE. Interestingly, *Nanog*$^{-/-}$ embryos fail to develop EPI but also lack PrE, suggesting that EPI precursors regulate PrE fate.

Indeed, ICM cells internalized first, the EPI precursor cells, upregulate the expression of fibroblast growth factor 4 (FGF4). Meanwhile, cells internalized later, the PrE precursor cells, inherit higher levels of FGF receptor 2 (FGFR2). EPI progenitors produce FGF4 that binds to FGFR2 in PrE progenitors, inhibiting NANOG and de-repressing GATA6 expression (also because NANOG and GATA6 are mutual inhibitors). Moreover, NANOG and OCT4, in EPI precursors, are required for FGF4 expression. Through this non-cell autonomous mechanism, EPI precursors maintain the PrE fate in PrE precursors (Figure 1.3). The final number of ICM cells, from the two waves of divisions, is always the same: if one wave produces less cells, the other wave compensates and vice-versa (reviewed in Saiz & Plusa 2013; Schrode *et al.* 2013).

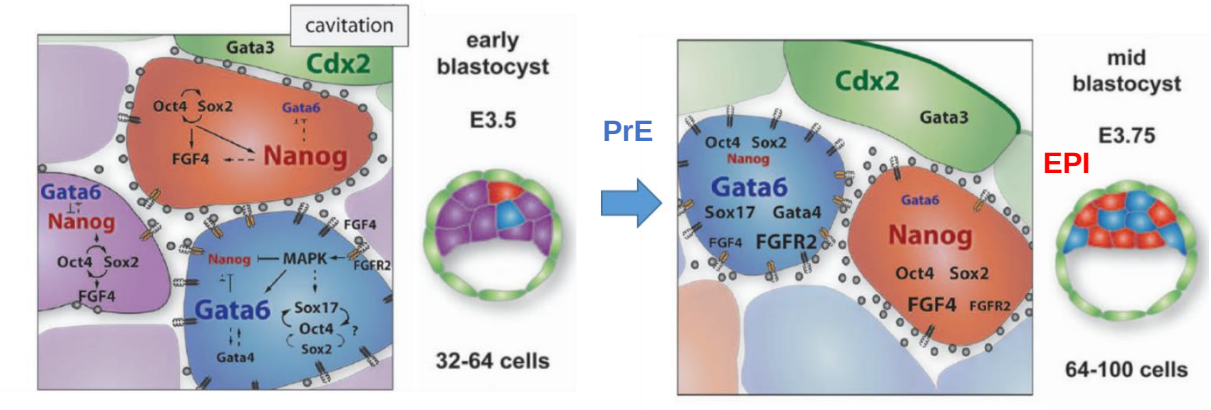


Figure 1.3 – Model for specification of EPI and PrE, within ICM cells. *Cells internalized earlier, the EPI precursors in red, express higher levels of NANOG, which in turn upregulates FGF4 signalling. On the contrary, cells internalised later, the PrE precursors in blue, show higher levels of FGFR2. FGF signalling will inhibit NANOG in PrE progenitors, de-repressing GATA6 expression and therefore biasing cells towards a PE fate. Adapted from Schrode* et al. *2013.*

Therefore, heterogeneity at polarity, position and gene expression, together with inter-cellular communication, creates cues that guide, but do not determine, cell-fate decisions, allowing flexibility during preimplantation embryo development. When all the three preimplantation cell lineages are formed (TE, EPI and PrE), the blastocyst enters the uterus and hatches out of the zona pellucida, an extracellular matrix surrounding the developing oocyte. In few hours, the blastocyst invades the maternal tissue and implants (reviewed in Bedzhov *et al.* 2014).

Following uterine implantation at E5.0, the mouse epiblast converts from a loosely adherent ball of cells into a single-cell layer of columnar epithelium with a cup-shaped structure, the egg cylinder. In female XX embryos one of the X chromosomes undergoes random X inactivation. Afterwards, the egg cylinder epiblast cells receive inductive signals from the adjacent yolk sac and trophoblast, becoming instructively specified according to their location (reviewed in Nichols & Smith 2009).

Molecular landmarks of this pre- to postimplantation embryo are the suppression of naïve pluripotency genes such as *Rex1*, *Klf2/4* and *Tbx3*, as well as, the upregulation of *Pou3f1*, *Otx2* and *Fgf5* (reviewed in Boroviak *et al.* 2014). At E6.5, the egg cylinder undergoes gastrulation, which results in the spatially organized formation of the three germ layers (mesoderm, endoderm and ectoderm).

1.2 Mouse Embryonic Stem Cells

1.2.1 Capturing Pluripotency

Pluripotent stem cells are characterized by the two main hallmarks of unlimited self-renewal and pluripotency with capacity to differentiate into the three cell lineages of an embryo, also known as multi-lineage differentiation potential. Pluripotent stem cells are present in the early embryo, in teratocarcinomas (malignant testicular germ cell tumours), which origin embryonal carcinoma cells (EC cells), and finally in primordial germ cells (PGCs), which are the founders of germline lineage and can give rise to embryonic germ cells (EG cells).

Mouse embryonic stem (mES) cells constitute the most well studied pluripotent stem cells. These immortalized cell lines are derived directly from the ICM epiblast of the mouse preimplantation blastocyst, between E3.5 and E4.5 (Evans & Kaufman 1981; Martin 1981). mES cells retain their pluripotent characteristics if maintained in specific culture conditions, namely under the effect of leukemia inhibitory factor (LIF) (described in section 1.2.3).

Moreover, mES cells have the possibility to form chimaeras, which means that when they are injected into mice blastocyst, they have the capacity to incorporate into the epiblast and to re-enter in the embryonic development, to produce functional soma and germ cells (Bradley *et al.* 1984). mES cells can also generate tumours containing differentiated cells from each of the three germ layers *in vivo*, named teratomas, when transplanted into immune-deficient mice.

Contrarily to mES cells, human ES (hES) cells raise many ethical questions, since human embryos must be destroyed for their derivation. Moreover, tissues obtained from allogeneic hES cell differentiation might induce immunogenic rejection upon transplantation in patients. One solution to overcome these difficulties is the formation of pluripotent cells directly from the patient's own cells. Pluripotent cells can be created in an epiblast-independent context by reprogramming somatic cells, either by fusion with hES cells, transfer of nuclear contents into oocytes, or more efficiently, by transfection with regulatory transcription factors (reviewed in Silva & Smith 2008). Mouse adult and embryonic fibroblasts can be reprogrammed into induced pluripotent stem (iPS) cells by applying four transcription factors: OCT3/4, SOX2, C-MYC and KLF4, under ES cell culture conditions. Interestingly, NANOG is unnecessary for reprogramming. iPS cells exhibit morphology, properties and gene expression similar to ES cells in a pluripotent state (Takahashi & Yamanaka 2006).

Pluripotent stem cells have many applications, namely disease modelling, drug screening and regenerative medicine. It is also essential to characterize pluripotency and to acknowledge the existence of distinct pluripotent stages.

1.2.2 "Naïve" and "Primed" Pluripotency

Two different stages of pluripotency can be distinguished, the "naïve" and "primed" pluripotency (Figure 1.4). The "naïve" state is a functional property attributed to cells with the unbiased capacity to give rise to chimaeras following blastocyst injection, independently of correspondence with the developmental state established *in vivo*. Both ES cells and preimplantation epiblast cells are said to be

in a “naïve” state of pluripotency. ES cells also share an epigenetic feature with the preimplantation epiblast: the presence of two active X chromosomes in female cells (reviewed in Nichols & Smith 2009).

Approximately at E5.0, the blastocyst implants into the uterus, leading to significant modifications in the epiblast. Nevertheless, the postimplantation epiblast remains pluripotent and can give rise to cell lines *in vitro*, named epiblast stem (EpiS) cells. EpiS cells can also arise from ES cell differentiation, by changing the culture conditions to a medium supplemented with activin A and fibroblast growth factor 2 (FGF2) (Tosolini & Jouneau 2015). These cells are developmentally and functionally distinct from ES cells in many aspects. EpiS cells exhibit a flattened morphology and inefficient clonal propagation, contrasting with the round-shape format of ES cells, tightly packed in clusters, and their high clonogenicity (reviewed in Osorno & Chambers 2011).

EpiS cells still retain some pluripotency since they are able to generate teratomas containing differentiated cells of each of the three germ layers, and they can also give rise to PGCs, demonstrating their germline potential. However, in contrast to “naïve” pluripotency, in XX female postimplantation epiblast and EpiS cells, one of the X chromosomes has undergone random inactivation (Heard 2004).

Furthermore, and of utmost importance, EpiS cells fail to contribute efficiently to chimaeras following blastocyst integration, similar to the postimplantation epiblast cells. This may reveal that EpiS cells have a reduced potency compared to ES cells, being more susceptible to differentiation cues (reviewed in Osorno & Chambers 2011). EpiS cells as well as the postimplantation epiblast cells are said to be in a “primed” state, which is an irreversible stage between pluripotency and differentiation, characterized by restricted cell potential and increased susceptibility to lineage commitment. This “primed” state should not be misunderstood with “lineage priming”, explained in section 1.3.2.

On the other hand, “naïve” pluripotency comprises both the “ground state” pluripotency, maintained by “2i” conditions, and other “less pristine” pluripotent states, such as the maintained by ES cells cultured in “Serum/LIF”, as long as they generate chimaeras.

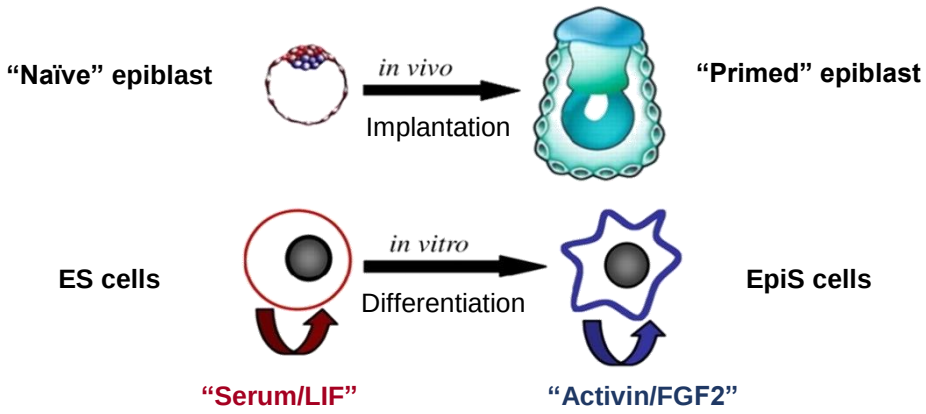


***Figure 1.4 – Differences* in vivo *and* in vitro *between naïve and “primed” pluripotency**. The transition from pre- to postimplantation embryo can be captured* in vitro *by changing environmental cues. “Serum/LIF”” favours ES cells pluripotency while “activin/FGF2” leads to EpiS cell self-renewal. Adapted from Osorno & Chambers 2011.*

1.2.3 Signalling Pathways

"Serum/LIF" Conditions

Originally, ES cells were co-cultured with fetal calf serum together with mitotically inactivated mouse embryonic fibroblasts (MEFs) or "feeders", now known to produce LIF (Williams *et al.* 1988). The cytokine LIF maintains self-renewal and inhibits the differentiation of mES cells through the activation of the signal transducer and activator of transcription 3 (STAT3). For that, LIF induces heterodimerization of the LIF receptor/GP130 complex, which leads to transphosphorylation and activation of receptor-associated Janus kinases (JAKs). JAKs phosphorylate and activate the transcription factor STAT3, which promotes target gene transcription when translocated to the nucleus (Niwa *et al.* 1998).

Both LIF withdrawal and STAT3 inhibition induce mES cell differentiation (Niwa *et al.* 1998). Moreover, mutant mouse embryos deficient in the LIF receptor, GP130 and STAT3 are lethal (Takeda *et al.* 1997). Currently, the standard culture medium for pluripotency in mES cells is supplemented with fetal bovine serum and LIF, named "Serum/LIF" or "conventional" conditions. In these conditions, mES cells are considered "metastable", cycling in and out of a pristine pluripotency state, which reflects into the varied cell morphology and heterogeneous expression of the key pluripotency gene *Nanog*, along with stochastic and reversible expression of lineage-affiliated genes. These mES cells are functionally "naïve" state but not at the "ground state", which means they not exactly mirror the preimplantation epiblast (reviewed in Hackett & Azim Surani 2014).

"2i" Conditions

"Serum/LIF" culture sustains self-renewal by counteracting "downstream" differentiation stimuli. In order to stabilize ES cells in a "upstream" naïve state it is necessary to shield them from these differentiation signals. Hence, the "2i" medium was developed, which is composed of two small molecule kinase inhibitors, the inhibitor of FGF/ERK pathway (PD03) and of GSK3 (CHIRON).

In the absence of serum or LIF, ES cells tend to lose pluripotency and the mechanism behind it involves an auto-inductive FGF/ERK signalling. FGF4, secreted by ES cells, binds to the FGFR on the membrane of neighbouring cells, inducing the mitogen activated protein kinase (MAPK) pathway, that "poises" ES cells for lineage entry (reviewed in Silva & Smith 2008). PD03 prevents phosphorylation and consequent activation of ERK1/2 (part of the MAPK cascade and downstream effector of FGF4), thereby enhancing long-term self-renewal of mES cells. CHIRON not only promotes self-renewal, through β-catenin stabilization, but also maintains ES cell propagation (Ying *et al.* 2008; reviewed in Hackett & Azim Surani 2014).

ES cells cultured in "2i" display a more homogeneous gene expression of *Nanog* and round-shaped morphology than ES cells in "Serum/LIF", corresponding to the "ground state" of pluripotency. ES cells grown in "2i" establish a strong parallel with the preimplantation epiblast, mostly due to a globally hypomethylated DNA, contrarily to cells cultured in "Serum/LIF", which possibly reflect later stages. This erasure of DNA methylation is associated with a downregulation of DNMT3A and DNMT3B (Leitch et al. 2013). The "ground state", which is functionally "naïve", is characterized by the absence of developmental constraints and, consequently, by an unbiased developmental potential, like a *tabula rasa* (reviewed in Silva & Smith 2008; Hackett & Azim Surani 2014).

1.3 Gene Regulatory Network Governing Pluripotency

1.3.1 Nanog-Oct4-Sox2 Complex: from Embryo to mES Cells

Pluripotency in the preimplantation epiblast and in stem cells is commanded by a gene regulatory network of the triumvirate core transcription factors: *Nanog*, *Oct4* and *Sox2* or NOS complex, which act in synchrony. The NOS complex is present in blastomeres, pluripotent early embryo cells (becoming confined to ICM and later to epiblast), PGCs and also in cultured pluripotent stem cells (Nichols *et al.* 1998). Furthermore, the absence of any of these factors leads to ES cell differentiation (Osorno & Chambers 2011). In this subchapter it is described the NOS network within the context of embryo development and in parallel with mES cells.

Oct4

OCT4, encoded by the gene *Pou5f1*, is a mammalian homeodomain transcription factor of the POU (Pit-Oct-Unc) family. Initially, *Oct4* expression is abundant and uniform in all morula's cells, however, as the outer cells differentiate into TE, *Oct4* expression becomes restricted to ICM cells and later to the epiblast (Palmieri *et al.*, 1994). The absence of *Oct4* in embryos causes peri-implantation lethality, before egg cylinder formation (Nichols *et al.* 1998).

After implantation, *Oct4* expression is maintained in the epiblast but it continuously decreases. Loss of pluripotency in embryos, observed at early somitogenesis stage, is coincident with the decline of expression and chromatin accessibility of *Oct4* and *Nanog* regulatory regions. *Nanog* expression is the first to become undetectable, in the somatic cells of 3- to 5-somite embryo (E8.25), whilst *Oct4* only disappears at 12- to 15-somite embryo (E8.75). Embryonic expression will then become restricted to PGCs (Osorno *et al.* 2012).

Oct4$^{-/-}$ embryos only differentiate into trophoblast giant cells and do not contain mature ICM nor ParE. In mES cells, specific levels of *Oct4* expression are critical for the maintenance of pluripotency: in line with the *in vivo* phenotype, *Oct4* absence triggers dedifferentiation to TE whilst *Oct4* overexpression leads to PrE and mesoderm fates (Niwa *et al.* 2000; Osorno & Chambers 2011). Hence, OCT4 seems to be crucial for the maintenance of pluripotency, like a gatekeeper that "locks" it, in both mES cells and ICM and, in its absence, the cell-fate is restricted to TE (Nichols *et al.* 1998). However, forced expression of *Oct4* in mES cells cultured without LIF is not sufficient to prevent differentiation (Niwa *et al.* 2000).

Sox2

Sox2 is a member of the *Sox* (SRY-related HMG box) gene family which encodes for transcription factors with a single HMG DNA-binding domain. Unlike *Oct4*, *Sox2* is expressed by the multipotential cells of the extraembryonic ectoderm (ExE) and it persists throughout the epiblast until the mid-late-streak stages (E7.0 – E7.5), when it becomes restricted to the anterior neuroectoderm (Avilion 2003). Moreover, *Sox2* when jointly expressed with *T-brachyury* characterize the posterior dual-fated neuromesodermal progenitors (NMPs) (reviewed in Henrique *et al.* 2015).

Sox2 null embryos contain a relatively normal ICM but fail to maintain an epiblast, causing early post-implantation lethality. The only surviving cells of null mutants are trophoblast giant cells and extra-embryonic endoderm (Avilion 2003).

Overexpression of *Sox2* does not impair the propagation of undifferentiated ES cells, although, upon release from self-renewal, cells differentiate into neuroectoderm (Zhao *et al.* 2004). On the other hand, similar to *Oct4*, genetic deletion of *Sox2* results in differentiation of ES cells solely into TE-like cells (Masui *et al.* 2007).

Nanog

NANOG is a mammalian homeodomain transcription factor, whose name derives from the mythological Celtic land of the ever young, Tir nan Og. *Nanog* transcripts appear as a temporal wave, upregulated between the late morula and the mid blastocyst, confined to ICM, later to epiblast and downregulated before implantation (Chambers *et al.* 2003). At this point, *Nanog* may act as a "brake" that must be removed so that implantation can proceed. After implantation, *Nanog* is re-expressed in the posterior region of the egg cylinder. Afterwards, it is gradually lost and disappears entirely at the onset of somitogenesis, becoming restricted to PGCs (Osorno *et al.* 2012).

In *Nanog*-null embryos, ICM is trapped in a pre-pluripotent indeterminate state that does not develop into pluripotent epiblast, which results in post-implantation lethality. *Nanog*-null ICM cells only have two options: differentiate into trophoblast or death. Thus, *Nanog* is essential for the generation of epiblast pluripotency (Silva *et al.* 2009).

Nanog expression declines significantly during differentiation, which suggests it has a role in self-renewal. Indeed, forced overexpression of *Nanog* is sufficient to drive autonomous mES cell self-renewal in the absence of LIF, thereby blocking differentiation. Furthermore, continued expression of *Oct4* is necessary for *Nanog*-mediated self-renewal (Chambers *et al.* 2003; Mitsui *et al.* 2003).

NANOG stochastically fluctuates in individual mES cells, exhibiting a heterogeneous expression at population level. Moreover, a transient downregulation of *Nanog* predisposes cells towards differentiation but does not mark commitment (Chambers *et al.* 2007; Abranches *et al.* 2014), further explained in section 1.3.2. Similar to the *in vivo* phenotype, mES cells in the permanent absence of *Nanog* lose pluripotency and differentiate into ParE and VE, accompanied by *Gata6* and *Gata4* expression. Thus, *Nanog* was thought to not only actively maintain pluripotency but also to prevent a PrE fate (Mitsui *et al.* 2003).

More recently, a study with *Nanog*-null mES cells showed that they can still self-renew indefinitely, despite the susceptibility to differentitation into PrE. *Nanog*$^{-/-}$ mES cells maintained an undifferentiated morphology with alkaline-phosphatase positive and LIF-responsive colonies, cells also expressed pluripotency genes, were clonogenic and formed chimaeras, until they differentiate (5-7 days after *Nanog* deletion). Therefore, *Nanog* acts primarily in the generation of ICM (Chambers *et al.* 2007) rather than in the "housekeeping" machinery of pluripotency, contradicting the previous study (Mitsui *et al.* 2003).

Surprisingly, NANOG is not one of the canonical quartet of transcription factors employed to produce iPS cells (Takahashi & Yamanaka 2006). In addition, *Nanog*-deficient iPS cells are

transcriptionally similar to wild-type iPS cells, able to generate chimeric mice and teratomas with the three germ layers represented (Schwarz et al. 2014).

ES Cells as a Reliable Model for the Embryo

It is possible to establish a strong parallelism at the gene regulatory network in pluripotency, between ES cells and the preimplantation embryo. This gives ES cells confidence to be considered as a model for the preimplantation embryo development.

As described in section 1.1.2, in the first cell-fate decision, at compact morula, cells can remain pluripotent as ICM or differentiate into TE. *Oct4* is the major responsible for this choice, since its depletion leads to a TE fate. Afterwards, in the second cell-fate decision, at early blastocyst, ICM cells can continue pluripotent as EPI or differentiate into PrE. At this point, *Nanog* has a key role: cells expressing *Nanog* become EPI and in its absence turn into PrE (Mitsui *et al.* 2003).

In conclusion, cell fate in development seems to be highly dependent on transcription factors. In order to comprehend the exit from pluripotency it is necessary to understand how these "molecular switches" of gene transcription are orchestrated in time and space. Since NANOG exhibits a heterogeneous expression it is a good target for further analysis, detailed in section 1.3.2.

1.3.2 Heterogeneity and Lineage Priming in mES Cells

"Cell-to-cell variation may be integral to the ES cell condition, safe-guarding self-renewal while continually presenting opportunities for lineage specification" (Silva & Smith 2008).

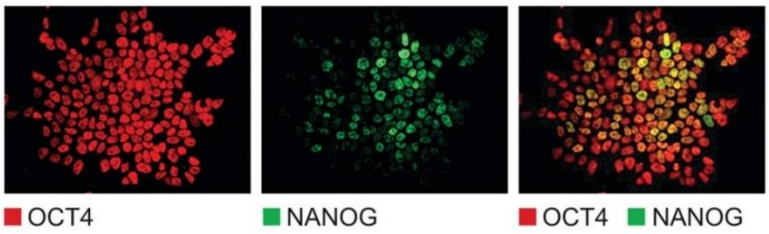


Figure 1.5 – NANOG heterogeneity*. ES cells show variable levels of NANOG protein, opposing to the homogeneous expression of OCT4 protein. Adapted from Torres-Padilla & Chambers 2014.*

Contrarily to OCT4 and SOX2, which exhibit a homogeneous protein expression in ES cells under self-renewing conditions, the expression levels of other pluripotency markers such as NANOG, PECAM1, STELLA and REX1 demonstrate a significant heterogeneity. As it can be observed in Figure 1.5, not all OCT4 positive cells express NANOG. In line with what was described in section 1.1.2, in the second cell-fate decision between PrE and EPI within the blastocyst ICM cells, there is a mosaic "salt and pepper" expression of GATA6 (PrE) and NANOG (EPI).

The existence of NANOG heterogeneity was confirmed by immunodetection (Singh *et al.* 2007) and by the use of fluorescent reporters to monitor the temporal dynamics of NANOG expression, such as a Nanog:VNP reporter, expressed by the Nd cell line (Abranches *et al.* 2013). This heterogeneity

seems to arise from stochastic fluctuations of NANOG expression, widespread amongst individual mES cells. NANOG fluctuations are a cell-autonomous property of pluripotent mES cells, with essentially all expressing cells showing variations in NANOG levels, even when cultured in “2i/LIF” conditions.

Only (56.2 ± 8.0)% of mES cells cultured in “Serum/LIF” express NANOG whilst there is an increase of NANOG expression up to (91.1 ± 3.1)% in “2i/LIF”-cultured cells (Abranches *et al.* 2013). Despite the minor subpopulation of low-NANOG cells in “2i/LIF”, there is a higher percentage of NANOG fluctuating cells (Abranches *et al.* 2014). Nevertheless, whether these fluctuations have a functional impact on the pluripotent state is still an open question.

The finding that *Nanog*$^{-/-}$ mES cells have an increased tendency to spontaneously differentiate (Chambers *et al.* 2007) led to the hypothesis that the low-NANOG state is more permissive to commitment whilst the high-NANOG state is a pure state of pluripotency, unresponsive to differentiation signals, like the “ground state” pluripotency. Indeed, low-NANOG cells are more prone to differentiate, have less self-renewal capacity and transiently express lineage-affiliated genes, when compared to high-NANOG cells. Lineage-affiliated genes are the genes responsible for lineage specification, such as *Fgf5* for ecto-, *Gata6* for endo- and *T-brachyury* for mesoderm. On the other hand, high-NANOG cells exhibit a more stable expression profile in different culture media, with minimal or null expression of lineage-affiliated genes. Therefore, loss of NANOG might be an early sign of pluripotency exit (Abranches *et al.* 2013; Abranches *et al.* 2014) (Figure 1.6).

This increased expression of lineage-affiliated genes in low-NANOG cells has been associated to “lineage priming” or simply “priming”, a process within the pluripotency window during which mES cells display reversible and upregulated expression of lineage-affiliated genes, reflecting an increased predisposition to commit into a cell-fate. It is “the seeding of a particular fate on the way to commitment”, which might involve chromatin modulation. “Lineage priming” is essential for the pluripotency capacity of ES cells, conferring the plasticity that allows differentiation (Martinez Arias *et al.* 2013).

The stochastic NANOG fluctuations were proposed to provide windows of opportunity for mES cells to explore their pluripotency, by responding to different signalling cues and testing multiple lineage differentiation programmes before definitive commitment. NANOG heterogeneity confers a stochastic advantage towards the ES cell population: not only preserves a pristine pluripotent identity, in a high-NANOG state, but also enables a fast response to signalling stimuli, required for a rapid transition to differentiation, which occurs in a low-NANOG state. However, the low-NANOG state does not mark definitive commitment, but rather defines a “lineage-primed” state: it “poises” or predisposes cells on a specific fate and they might revert their decision by re-expressing NANOG (Figure 1.6).

The “lineage-primed” state is characterized at the molecular level by low expression of *Nanog* accompanied by high expression of *Oct4*, *Sox2*, as well as reversible and increased expression of lineage-affiliated genes. However, this definition does not mean that all low-*Nanog* cells will be “lineage-primed”. When a cell differentiates, it has a decreased expression of these three pluripotency factors balanced by the upregulation of lineage-affiliated genes. Meanwhile, in a pristine state of pluripotency all pluripotency factors are upregulated and lineage-affiliated gene expression is absent (Figure 1.6).

Previous work in DHenrique Lab led to the identification by RNA-sequencing of genes preferentially upregulated in low-Nanog:VNP cells cultured in “Serum/LIF”, named priming genes.

Initially it was thought that priming genes would only correspond to lineage-affiliated genes, hence the attribution of the name “priming”, nevertheless, they comprise other genes as well. In fact, priming genes include two classes: (1) Lineage-affiliated genes (such as *Sox3* for neural fate); (2) Sporadic genes, with an unknown role in development and irregular expression (such as *Car2* and *Cld6*). Thus, it was hypothesized that NANOG regulates priming gene expression, although the mechanism behind this control is still unknown, which is exactly the question this project aims to address.

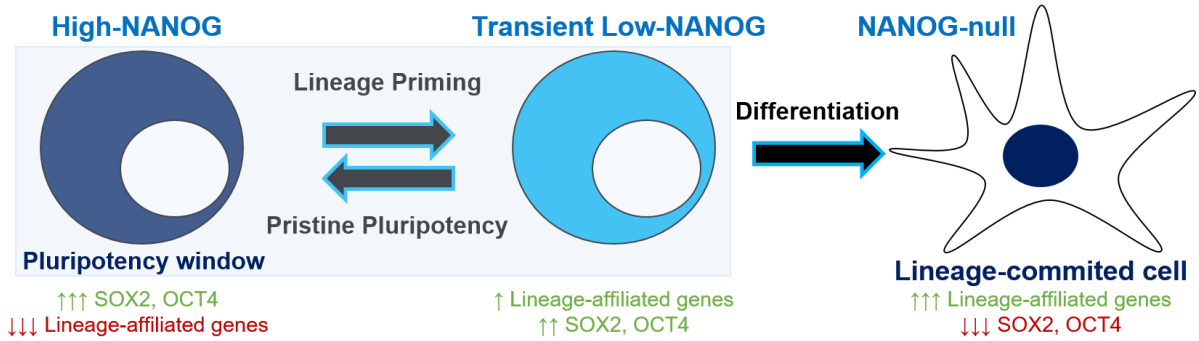


Figure 1.6 – NANOG as a “differentiation rheostat”. *ES cells maintain a balance between self-renewal and differentiation. This equilibrium is influenced by NANOG, which is constantly fluctuating between high and low states. Indeed, overexpression of Nanog shields the cell from differentiation induction and renders self-renewal constitutive. A transient Nanog negative ES cell might undergo “lineage priming”, characterized by reversible and high expression of lineage-affiliated genes, being more prone to differentiate. Nevertheless, cell’s “decision” may be reverted by the re-expression of Nanog, returning to a state of pristine pluripotency. A cell that does not express any of the three pluripotency factors, but shows upregulation of lineage-affiliated genes, has probably already differentiated into a lineage programme. Adapted from Chambers* et al. *2007.*

This type of stochastic models were already applied to the lineage commitment in hematopoietic stem cells (Pina *et al.* 2012). Gene fluctuations might be controlled at the transcriptional level through feedback loops, namely an auto-repression mechanism was proposed to control *Nanog* expression independent of OCT4 and SOX2 (Navarro *et al.* 2012). Moreover, dynamic allele switching (Miyanari & Torres-Padilla 2012) and crosstalk between signalling pathways and pluripotency transcription factors (reviewed in Silva & Smith 2008) have been proposed to explain NANOG fluctuations. The next section 1.3.3 will try to elucidate what is the origin of mES cell’s stochasticity.

1.3.3 Stochasticity in mES Cells

Originally, it was thought that mRNAs were produced and degraded randomly, hence the probability to find a transcript produced within any given time would be constant, following a Poissonian distribution. Nevertheless, mRNA production does not occur with a constant probability in time but rather in transcriptional bursts or pulses. Most of the time, genes are in a transcriptionally inactive state, mRNA synthesis does not occur and the already formed mRNAs are degraded at variable rates (Figure 1.7.b/d) (Raj & van Oudenaarden 2009).

Therefore, a nonbursty or continuous transcriptional activity, which reflects into a perfect Poisson mRNA distribution, might be an idealized model, difficult to find in mES cells (Figure 1.7.a/c). On the other hand, a stochastic switch behaviour between "ON" and "OFF" states, observed in a bursty or discontinuous transcription, generates a huge variability in gene expression, which in turn might allow room for developmental decisions in mES cells (Figure 1.7.b/d). So far, genes present transcriptional bursts that display different probabilities of transcriptional firing, pulse duration, frequency and intensity (Torres-Padilla & Chambers 2014).

In order to define whether the transcription of a particular gene is bursty it is necessary to obtain the number of mRNA molecules per cell, within a population. Nowadays, this single-cell analysis is possible through single molecule RNA-FISH, single-cell RT-PCR, molecular beacons, among other techniques (Raj & van Oudenaarden 2009).

In bursts, genes stochastically switch between active and inactive transcriptional states due to noise, which can be intrinsic or extrinsic. The intrinsic noise refers to the random events that dictate which reactions occur and in what order during gene expression. By contrast, the extrinsic noise arises from fluctuations in cellular components, such as location or concentration of transcriptional activators (like RNA polymerases) (Elowitz *et al.* 2002). Both types of noise contribute to gene transitions between states, nevertheless these are most likely to be due to intrinsic noise, namely dependent on chromatin remodelling (Raj *et al.* 2006).

Both the transcription of mRNAs and consequent translation into proteins are highly stochastic processes, which lead to cell-to-cell variability, namely relative to *Nanog*, as described in section 1.3.2. Development can be considered an intrinsically noisy system due to fluctuations in transcriptional regulation, but where is the order in this chaotic noise? Coming back to our fundamental question, how does *Nanog* heterogeneity influence priming gene expression? In chapter 1.4 we shall look into NANOG's partners, involved in transcriptional modulation.

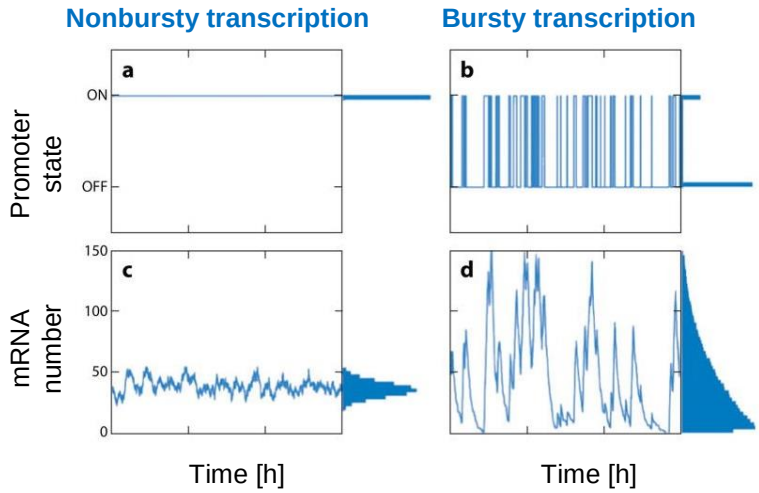


Figure 1.7 – Nonbursty versus bursty transcription models. *Promoter dynamics for a gene that is (a) constantly active (nonbursty, continuous or "one-state" model) versus (b) a gene which oscillates between ON and OFF states (bursty, discontinuous, pulsatile or "two-state" model) due to transcriptional bursts. mRNA dynamics for (c) nonbursting and (d) bursting genes. The nonbursty transcription originates a Poisson distribution of mRNAs per cell across the population, depicted in the marginal histogram, whereas in the bursty transcription is much wider than a Poisson distribution. Adapted from Raj & van Oudenaarden 2009.*

1.4 Chromatin Dynamics in Pluripotency

1.4.1 DNA Methylation and Histone Modifications

Despite the fact that every single cell in an organism contains the same genetic code, not all cells present the same phenotype nor function. Only a few thousands of genes are expressed in a specific cell, within a given time. How does each cell differentially express a certain selection of genes? The answer to this question involves the concept of epigenetics: “an epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence” (Berger *et al.* 2009).

The DNA of eukaryotic cells is wrapped around a set of evolutionarily conserved histone proteins, which together form the chromatin. The basic structural unit of chromatin is the nucleosome, which is composed by a histone octamer with two molecules of each of the four core histones H2A, H2B, H3 and H4 and 147 base pairs of DNA, wrapped around it in a superhelix (reviewed in Barth & Imhof 2010). Chromatin can be remodelled from a tightly compacted, transcriptionally inactive structure where DNA is less accessible (heterochromatin or “closed chromatin”) to a more “loosely packed”, transcriptionally active conformation, which allows DNA access (euchromatin or “open chromatin”) (Elgin & Grewal 2003).

A phenotypic difference is likely triggered by environmental signals, which are received by an “initiator”. “Initiators” can be TFs, that “perceive” the signal and define the location on a chromosome where the “epigenetic chromatin state” is going to be established. Then, a “maintainer”, which can be TET1 or PRC2, sustains this “state” through DNA (de)methylation, histone post-translational modifications or other process, that ultimately influence gene transcription (Berger *et al.* 2009).

DNA methylation occurs almost exclusively at CpG dinucleotides in mammals and it was generally associated with transcriptional silencing. CpGs are clustered into CpG islands (CGIs), often at gene’s promoters, and are usually unmethylated. A methyl group is laid down at the 5’ group of cytosine by *de novo* methyltransferases DNMT3a and DNMT3b, and methylation can be maintained by the action of DNMT1 or diluted through demethylation pathways. DNA demethylation can be either passive, through dilution with cell division, when DNMT1 is not in the nucleus, or active, through ten-eleven translocation (TET) enzymes (reviewed in Schübeler 2015).

Histone post-translational modifications, such as methylation or acetylation, can either activate or repress gene transcription. Histone methylation can involve the transference of one, two or three methyl groups (mono-, di- or trimethylated, respectively). It is of particular interest the trimethylation of lysine 27 on histone 3 (H3K27me3), usually mapped to regions where transcription is repressed and catalysed by the polycomb repressive complex 2 (PRC2) complex. On the other hand, H3K4me3 is a modification frequently located in actively transcribed regions and is catalysed by the trithorax/MLL complex (reviewed in Barth & Imhof 2010).

“Bivalent domains” are regions characterized by the co-localization of both active (H3K4me3) and repressive (H3K27me3) marks and are present at promoters of around 3000 developmental regulators in mES cells cultured in “Serum/LIF”. In addition to the active H3K4me3 mark, bivalent genes also display a RNA polymerase II (RNAPII), which keeps the promoter “poised” for timely activation

despite the fact that the gene is silenced, like a “transcriptional pause” ready to “replay”. Current models propose that the presence of H3K4me3 and RNAPII allows a rapid transition from transcriptional repression to activation of developmental genes, upon differentiation cues. Therefore, bivalent domains are quickly resolved into actively transcribed genes (characterized by the presence of H3K4me3 and loss of H3K27me3) or silent genes (vice-versa), according to the differentiation process. Thus, this fine-tuned regulation of gene expression seems to be essential for cell-fate decisions in mES cells on their “way to commitment” (Azuara *et al.* 2006; Bernstein *et al.* 2006; reviewed in Aloia *et al.* 2013). However, how this chromatin state switching is controlled and correlated to ES cell heterogeneity remains unclear.

DNA methylation and histone modifications are important in maintaining a pluripotency program in ES cells and in guiding correct differentiation during embryonic development. Nevertheless, TFs are the true “master regulators” of gene expression, giving instructions to the chromatin remodellers, which act as “workers” following the “master’s rules”. Interestingly, NANOG was found to bind to PRC2 and to TET1, this interaction is further described in section 1.4.2.

1.4.2 NANOG Partners: PRC2 and TET1

PRC2 Complex

The polycomb group (PcG) proteins are involved in the control of gene repression via chromatin compaction. In mammals, PcG proteins are found in many multiprotein complexes, namely PRC1 and PRC2. While PRC1 mediates the monoubiquitylation of histone 2A on lysine 119, PRC2 catalyses the di- and trimethylation of H3K27, both responsible for the silencing of target genes. Nevertheless, PcG proteins should not be considered “absolute silencers”: active PRC targets switch between PRC-repressed and active states, possibly associated to the regulation of RNAPII in metabolic genes (Brookes *et al.* 2012). Moreover, PRC2 occupancy has been associated with large unmethylated CGIs. PcG proteins play a role in several biological processes, such as cell cycle control, tissue homeostasis, tumorigenesis and, most importantly, in stem cell differentiation (reviewed in Aloia *et al.* 2013).

Indeed, PcG proteins co-localize at genes encoding developmental regulators, which constitute direct targets for PRC2-mediated transcriptional repression. These developmental genes (*Hox*, *Fox*, *Sox*, *Gata* and *Tbx*) are essential for cell-fate decisions in both mouse and human ES cells, being reactivated during ES cell differentiation. Thus, it was suggested that the genes repressed by PRC2, in ES cells cultured in pluripotent conditions, maintain the potential to become later activated upon commitment, under PRC2 temporal control (Boyer et al. 2006; Lee et al. 2006; reviewed in Surface *et al.* 2010). However, in this project, it is hypothesized that PRC2-mediated repression stops before definitive commitment and still within the pluripotency window, in a transient low-NANOG “lineage-primed” state (Figure 1.8).

In mammals, the PRC2 complex is composed by three core components, SUZ12 and EED, for complex assembly and proper enzymatic activity, and EZH2, the catalytic subunit. EZH2 is a histone lysine methyltransferase that catalyses the trimethylation of the H3K27 in the nucleosome substrate, via transfer of a methyl group from the cofactor S-(S’-adenosyl)-L-methionine (SAM) (Verma *et al.* 2012).

EZH2 is expressed during development, including in ES cells, whereas its paralog EZH1 is preferentially expressed in adult differentiated tissues (Villasante *et al.* 2011).

Aberrant histone hypermethylation, caused by *Ezh2* mutations, contributes to tumour initiation and progression, thus its inhibition is an attractive therapeutic approach. This motivation led to the creation of GSK343, a highly potent and specific small molecule inhibitor of EZH2 activity, blocking the *de novo* trimethylation of H3K27 by competing with SAM (Verma *et al.* 2012).

Despite some proposed mechanisms, the recruitment of PRC2 to DNA in mammals is still not clear, nor how PRC2 influences mES cell's "lineage priming". Knockout mice for PRC2 components die during early postimplantation stages (reviewed in Aloia *et al.* 2013). In line with this result, ES cells deficient in *Eed* (Boyer *et al.* 2006) or *Suz12* (Pasini *et al.* 2007) present an increased expression of lineage-affiliated genes, suggesting a role for PRC2-mediated repression during pluripotency. *Ezh2*$^{-/-}$ mES cells fail to abolish H3K27me1/3 at some genes due to redundancy with *Ezh1* (Shen *et al.* 2008).

Ezh2-null iPS cells present lower levels of H3K27me3 at the *Nanog* promoter and higher levels of NANOG, resulting in an expansion of high-NANOG subpopulation, compared to control iPS cells. In addition, levels of EZH2 and H3K27me3 at the *Nanog* promoter were increased in the low-NANOG subpopulation of ES/iPS cells, when compared to the high-NANOG subpopulation. Together, these results indicate that EZH2 is a direct regulator of *Nanog* expression and affects the NANOG equilibrium towards a low-NANOG state, which might be "lineage-primed" (Villasante *et al.* 2011).

Interestingly, priming genes, identified by their higher expression in low-NANOG cells, are enriched for binding signatures in PRC2 components. In contrast, genes with higher expression in high-NANOG cells are enriched for pluripotency TFs (unpublished results from DHenrique Lab). Moreover, in recent NANOG interactome studies many proteins were identified as NANOG's partners, including PRC2 (Gagliardi *et al.* 2013) and TET1 (Costa *et al.* 2013). But what is exactly the role of TET enzymes?

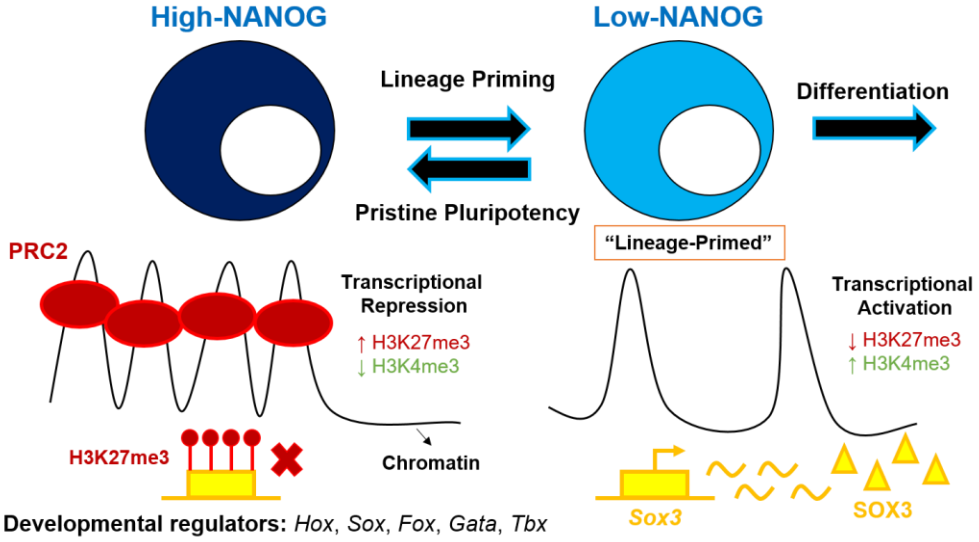


Figure 1.8 – PRC2 role as a transcriptional repressor. *PRC2 silences the expression of developmental regulators in ES cells cultured in "pure" or pristine pluripotent conditions, though chromatin compaction, associated to an increase of H3K27me3 repressive mark (and decrease of H3K4me3). Nevertheless, upon "lineage priming" in a transient low-NANOG state, it is hypothesized that PRC2 stops repressing lineage-affiliated gene expression. In mES cell differentiation, chromatin decompaction and gene activation are observed, associated to a reduction in H3K27me3 repressive mark and an increase in H3K4me3 activation mark. Adapted from Surface* et al. *2010.*

TET Enzymes

DNA methylation was generally described as being responsible for long-term and stable repression of gene expression in eukaryotic cells. Nevertheless, recent data shows increased turnover of methylation during differentiation, challenging the concept of static DNA methylation into a more dynamic event.

TET proteins are a family of DNA hydroxymethylases or dioxygenases that promote DNA hydroxymethylation and its mediated active demethylation through the catalytic conversion of 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). Later, 5fC or 5caC can be excised by thymine DNA glycosylase (TDG) and subsequently replaced with unmodified C through base excision repair (BER) (Figure 1.9.A). The three successive oxidation reactions are dependent of Fe^{2+}, alpha-ketoglutarate (α-KG) and 2-oxoglutarate. Whilst 5hmC accumulates in cells at a hundred to several thousand modified bases per million unmodified C, 5fC and 5caC expression levels are almost undetectable. The formed 5hmC can be also diluted by DNA replication, suggesting a passive DNA demethylation pathway (Yin *et al.* 2013; reviewed in Kohli & Zhang 2013).

The TET family is composed of three members, TET1, 2 and 3, and the enzymatic activity of TET1 is conserved from human to mouse. The mouse *Tet1* and *Tet2* genes, but not *Tet3*, are redundant, both are highly expressed in pluripotent mES cells, downregulated upon differentiation and coincident with the 5hmC enrichment. Through chromatin immunoprecipitation coupled with high-throughput DNA sequencing (ChIP-Seq), it was possible to infer that TET1 is enriched at genomic regions with high-density CpG sites (CGIs), usually hypomethylated (Wu *et al.* 2011).

Two different studies using *Tet1* knockdown in mES cells, by lentiviral short hairpin RNA (shRNA), identified the *Tet1*-activated and *Tet1*-repressed targets. Amongst the *Tet1*-activated genes was *Nanog*, *Tcl1* and *Esrrb*, concluding that, in addition to binding to CpG-rich promoters, TET1 also binds to a subset of actively transcribed CpG-poor promoters, essential for mES cell pluripotency (Wu *et al.* 2011). In line with this result, *Tet1* knockdown leads to the loss of mES cell morphology and self-renewal, probably caused by the increase in DNA methylation and consequent downregulation of *Nanog*. In this context, TET1 promotes a transcriptionally active state of *Nanog* by maintaining the promoter hypomethylated (Ito *et al.* 2010).

Unexpectedly, despite the fact that DNA demethylation has been associated to transcriptional activation, there were more *Tet1*-repressed genes rather than activated, suggesting a novel repressive function for TET1. Therefore, TET1 has two fundamental roles: (1) as a transcriptional activator sustains pluripotency by maintaining the expression of key pluripotency genes, such as *Nanog*; (2) as a transcriptional repressor silences the expression of lineage-affiliated genes, such as *Cdx2* (TE), *Krt8* (ectoderm), *Sox17* and *Gata6* (endoderm) (Ito *et al.* 2010; Wu *et al.* 2011).

This result raised the possibility that TET1 may play an important role on ICM formation, where it is enriched, in comparison to TE. To address this hypothesis, small interfering RNAs (siRNAs) against *Tet1* were injected into single blastomeres at the 2-cell stage and its effect on ICM and TE was assessed. Indeed, *Tet1* knockdown cells favoured an embryonic cell specification towards TE fate in detriment of ICM, as observed by the increase of CDX2 and decrease of OCT4 (Ito *et al.* 2010).

Moreover, TET1 contributes to the genome DNA demethylation during PGCs specification and to the dedifferentiation of adult cells during reprogramming to iPS cells. On the other hand, in adult tissues, *Tet1* downregulation might be involved in tumor initiation and progression. Hence, the need to find a small molecule able to promote TET1 activity: the ascorbic acid (AA), also known as vitamin C (Yin *et al.* 2013).

The addition of AA to mES cells and mice leads to a fast and global increase in all 5mC oxidation products, from 5mC loss. Consequently, there is DNA demethylation of many promoters, including germline and pluripotency genes, as mentioned before. In order to test if the effects of AA were mediated by TET, *Tet1/2* double knockout mES cells were analysed, resulting in no differences of 5mC or its oxidation products at gene promoters, upon AA exposure. Thus, the effects of AA are TET-dependent. AA also reduces DNA methylation at CGIs that usually gain methylation during the blastocyst to epiblast transition, leading to a state reminiscent of the ICM (Blaschke *et al.* 2014).

The mechanism underlying AA enhancement of the catalytic activity of TET for the oxidation of 5mC involves a direct interaction with the C-terminal catalytic domain of TET. This interaction might promote the folding of TET enzymes and/or the recycling of the co-factor Fe^{2+}, by reducing the intermediate Fe^{3+} to Fe^{2+} (Yin *et al.* 2013). AA was also found to improve the speed and efficiency of iPS cell generation from mouse and human somatic cells, by promoting the transition of pre-iPS cell colonies to a fully reprogrammed state (Esteban *et al.* 2010). Furthermore, there is a recent study which affirms that the exposure of AA to mES cells leads to the increase of NANOG (Wu *et al.* 2014).

In an interactome study, TET1 was identified as a novel partner of NANOG. It was found a physical association between NANOG and TET1, which potentiates the 5hmC increase and enhances the efficiency of reprogramming MEFs into iPS cells. TET1 and NANOG co-occupy genomic *loci* of genes associated with maintenance of pluripotency (such as *Esrrb*) and lineage commitment (such as *Pax6*) in ES cells. Moreover, TET1 binding to these common targets is reduced upon NANOG deletion, suggesting that NANOG recruits TET1 to these *loci* (Costa *et al.* 2013). Interestingly, one recent study found that the transcriptional activation of *Tet1* is promoted by the ES cell-specific factors OCT3/4, NANOG and MYC, suggesting a regulatory positive feedback loop (Neri *et al.* 2015).

PRC2 and TET1 have in common genomic *loci*, NANOG binding and repression of lineage-affiliated genes. Do they interact, if yes, how?

Crosstalk between PRC2 and TET1

Indeed, PRC2 occupancy has been also associated to large unmethylated CGIs, linked to the repression of many developmental regulators. Approximately 95% of PRC2-binding sites overlap with TET1-bound *loci* in mES cells. Moreover, *Tet1*-repressed genes are preferentially associated to bivalent chromatin states, whereas *Tet1*-activated targets are H3K4me3-only genes (Wu *et al.* 2011).

Furthermore, *Tet1* knockdown impairs the ability of EZH2 to bind to 72% of PRC2-binding sites, whereas *Ezh2* depletion does not affect TET1 binding, suggesting that TET1 acts “upstream” of PRC2 and is required for the chromatin binding of PRC2 in mES cells (Wu *et al.* 2011).

Interestingly, one study found that *Tet1* downmodulation is mediated by EZH2 in MEFs. Through ChIP analysis, the researchers discovered that, during MEFs passages, there is a gradual increase of

H3K27me3 and reduction of H3K4me3 on the *Tet1* bivalent promoter, suggesting a progressive repression by EZH2 recruitment. Knockdown of *Suz12* and also the application of GSK343 in MEFs resulted in TET1 and 5hmC upregulation, with a correspondent reduced EZH2 binding and consequent decrease of H3K27me3 deposition on the *Tet1* promoter (Neri *et al.* 2015).

However, no stable interactions have yet been observed between TET1 and EZH2. Moreover, *Tet1* knockdown does not reduce PRC2 expression or stability, which raises the question how of TET1 affects PRC2 binding without a direct interaction with its subunits?

All these results, accompanied with some unanswered questions, led to the hypothesis that TET1, by maintaining a hypomethylated state at CpG-rich promoters (PRC2-bound *loci*), contributes indirectly to PRC2 recruitment (Figure 1.9.B) and, consequently, to the silencing of developmental regulators (Wu *et al.* 2011; Sui *et al.* 2012). In fact, this is supported by another study showing that DNA methylation blocks PRC2 binding to chromatin (Wu *et al.* 2010).

Nevertheless, other models were proposed to explain TET1 repressive role, namely one states that PRC2 recruits TET1 to chromatin of bivalent promoters, to maintain their hypomethylated state (Neri *et al.* 2013). In sum, TET1 might mediate a crosstalk between DNA methylation and the surrounding histone modifications, which may result on the observed TET1 repressive and active effects on gene expression depending on its distribution throughout the genome and co-localization with PRC2. This crosstalk might be essential for "lineage priming" in mES cells.

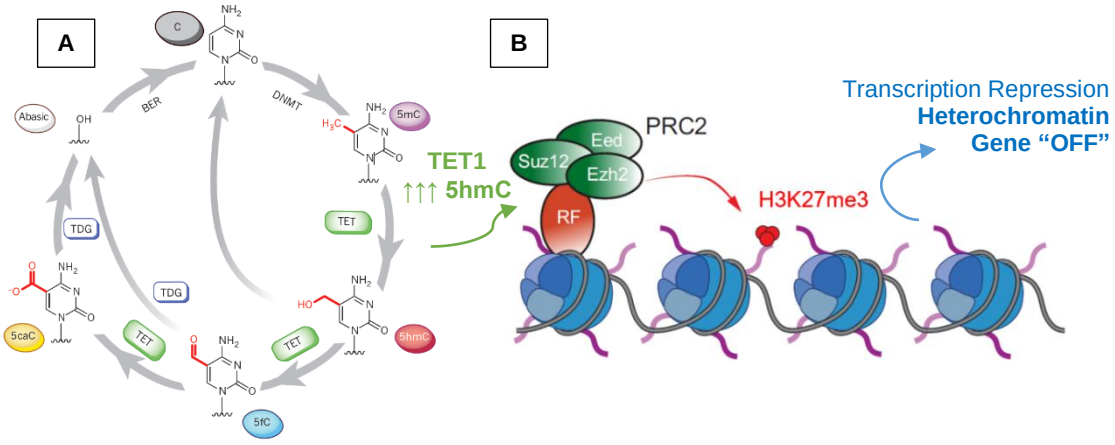


Figure 1.9 – Dual roles of TET1 in ES cells. (A) *TET1 role as a transcriptional activator of pluripotency genes, such as Nanog, through the active DNA demethylation pathway by the oxidation reactions of 5mC into 5hmC, 5fC and 5caC.* ***(B)*** *The new role of TET1 as a transcriptional repressor: TET1 indirectly, through 5hmC increase, "attracts" PRC2 to lineage-affiliated gene promoters, leading to the deposition of H3K27me3 and causing the silencing of lineage-affiliated gene expression. Adapted from Kohli & Zhang 2013 and Aloia* et al. *2013.*

1.4.3 The Model for NANOG Regulation of Priming Gene Expression

As mentioned in section 1.3.2, priming genes are genes upregulated in low-NANOG cells, which comprise lineage-affiliated genes, specific for a certain cell-fate, and sporadic genes, with unknown role in development. To explain NANOG regulation of priming gene expression, the working hypothesis at the basis of this thesis proposes that NANOG recruits TET1 (Costa *et al.* 2013), which would lead to the

conversion from 5mC to 5hmC in NANOG-regulated priming gene promoters. The hypomethylated state of the promoters "recruits" PRC2 to these binding sites and leads to the trimethylation of H3K27 (Wu *et al.* 2011), causing the silencing of priming gene expression in high-NANOG mES cells (Figure 1.10.A).

Furthermore, this model justifies the occurrence of "lineage priming": the increased lineage-affiliated gene expression in low-NANOG cells, where NANOG would be almost absent and no NANOG-TET1 complex would be formed. Consequently, PRC2 would not be "recruited", which would lead to the prevention of repression of lineage-affiliated gene expression (Figure 1.10).

In order to test this hypothesis, small molecules were used to modulate the activity of EZH2 and TET1 in mES cells. To inhibit EZH2 activity, GSK343 was used, which was shown to block *de novo* methylation of H3K27 (Verma *et al.* 2012). AA was used to increase TET1 activity and cause DNA demethylation, translated into higher levels of 5hmC (Yin *et al.* 2013; Blaschke *et al.* 2014).

Preliminary data from DHenrique Lab revealed that GSK343 causes increased expression of *Sox3*, a lineage-affiliated gene, in high-NANOG cells (Figure 1.10.A). On the other hand, exposure to AA causes a decrease of *Sox3* expression in high-NANOG cells (Figure 1.10.B). In addition, simultaneous treatment of mES cells with GSK343 and AA leads to the reversion of AA effect, indicating that PRC2 acts "downstream" of TET1.

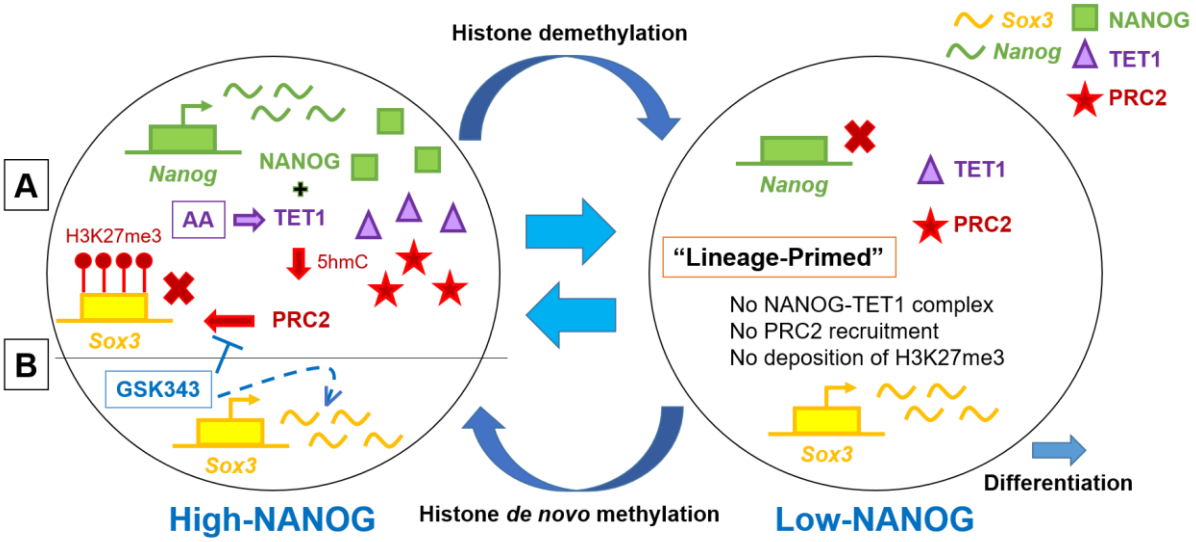


Figure 1.10 – Proposed model for NANOG regulation of priming gene expression*. NANOG stochastically fluctuates between high- and low-NANOG states. It is predicted that histone* de novo *methylation occurs from low- to high-NANOG state, due to PRC2 recruitment, "attracted" by hypomethylated regions promoted by the NANOG-TET1 complex. Consequently, PRC2 mediates the silencing of the expression of Sox3, a lineage-affiliated gene, in the high-NANOG state, through trimethylation of H3K27. When AA is added, it promotes TET1 action, leading to the decrease of Sox3 expression in high-NANOG cells. This represents what is expected to occur in a high-NANOG state (even without AA interference)* ***(A)****. GSK343, through EZH2 inhibition, leads to an increase in Sox3 expression in high-NANOG cells* ***(B)****. By contrast, in a low-NANOG state, NANOG is absent and there is no NANOG-TET1 complex, nor consequent PRC2 recruitment, which leads to an increase in Sox3 expression.*

2. Motivation and Research Aims

Despite the fact that mES cells have been extensively used by the scientific community during the last 25 years, the mechanisms underlying the exit from pluripotency have not been totally elucidated. The main goal of this project is to unravel the molecular machinery behind the NANOG regulation of priming gene expression in mES cells, through its interaction with PRC2 and TET1. Thus, my motivation behind this goal is to provide new insights on the exit from pluripotency of mES cells, which might help to optimize differentiation protocols, and to ultimately contribute to regenerative medicine or drug screening tests, amongst other mES cell's applications.

The working model is that NANOG forms a complex with TET1 (Costa *et al.* 2013), which increases 5hmC levels and favours the recruitment of PRC2 (Wu *et al.* 2011) and trimethylation of H3K27 in priming gene promoters, causing their repression. In order to test this hypothesis, small molecules that interfere with PRC2 and TET1 activity will be used: GSK343 and AA, respectively. A single-cell analysis is required due to mES cell heterogeneity, made possible by single-molecule RNA FISH (smFISH), which will be optimized for the first time in a new microscope system.

More specifically, this master thesis proposes to achieve the following aims:

1. (A) Describe *Nanog* transcriptional dynamics during pluripotency through smFISH quantification of *Nanog* transcripts in single mES cells; (B) Monitor NANOG dynamic expression at protein level though flow cytometry (FC) analysis of Nanog:VNP cells cultured in "Serum/LIF" conditions.
2. Understand if low-*Nanog* cells are in a "lineage-primed" state and characterize it. To accomplish this aim I will verify how priming gene expression correlates to *Nanog* expression in pluripotency at mRNA level by smFISH.
3. (A) Verify how NANOG expression is affected by GSK343 and/or AA exposure, by FC; (B) Verify if GSK343 treatment reduces H3K27me3 levels, by intracellular staining-flow cytometry (IC-FC).
4. Dissect the mechanisms by which NANOG regulates priming gene expression, in an mES cell unsorted population, by analysing how priming gene expression, in correlation to *Nanog*, is affected by the exposure of GSK343 and/or AA at the mRNA level by smFISH.
5. Dissect the mechanisms by which NANOG regulates priming gene expression, in low- or high-NANOG FACS-sorted subpopulations. (A) Study the effect of GSK343 and/or AA on NANOG expression; (B) Verify how priming gene expression correlates to *Nanog* expression at the mRNA level by smFISH.

The accomplishment of all the previous aims shall confirm the already existing model for NANOG regulation of priming gene expression through its interactions with PRC2 and TET1.

First, there is a brief "Introduction", followed by "Motivation and Research Aims", and the next chapter compiles all "Materials and Methods" used. Afterwards, the chapter of "Results and Discussion" is subdivided in three topics, the first is about *Nanog* in pluripotency and priming gene expression (aims 1. and 2.). The second subchapter is relatively to the mechanisms of NANOG regulation of priming gene expression through PRC2 and TET1, by the exposure of GSK343 and/or AA to mES cells (aims 3. and 4) and the third is dedicated to sorting (aim 5.). Finally, there is the chapter of "Conclusions", accompanied by limitations and strengths of the project, followed by "Future Perspectives".

3. Materials and Methods

3.1 Materials

3.1.1 Mouse Embryonic Stem Cells

The mES cell lines used in this work were: E14tg2a (E14), derived from 129 mouse strain, and Nd (for Nanog dynamics). Nd mES cells were derived from E14tg2a and are a BAC-transgenic line for VNP-tagged *Nanog* gene (Abranches *et al.* 2013). E14tg2a was a gift from Austin Smith (Wellcome Trust Centre for Stem Cell Research, University of Cambridge, UK).

3.1.2 Reagents

Below it is presented a set of tables relative to the reagents, solutions/media, chemical modulators, antibodies, smRNA-FISH probes and filter sets for smFISH experiments used throughout the work described in this master thesis (Table 3.1, Table 3.2, Table 3.3, Table 3.4, Table 3.5 and Table 3.6 respectively). For cell passaging, P60 dishes from Nunc were used and for production of GMEM x1, filters from Millipore were used, with 0.22 µm-size pores (in order to avoid bacterial or fungi contamination).

Table 3.1 – List of reagents used in the experiments performed in this project. *It is given information relative to its suppliers, catalogue numbers, stock and working stock concentrations and temperatures.*

Reagent	Supplier	Cat. Number	Stock	Working Stock
2-mercaptoethanol	Sigma	M-7522	RT	0.1 M in H2O, 4ºC
Catalase	Sigma	C-3515	-	4ºC
DAPI	Sigma	D9542	1 mg/mL in PBS, -20ºC	1.5 µg/mL in PBS, 4ºC
Dextran Sulphate	Sigma	D8906	-	4ºC
DMSO	Sigma	D-2650	RT	RT
Dow corning high vacuum silicone grease	Sigma	Z273554	-	RT
Dream Taq	Fermentas	EP0701	-20ºC	-20ºC
FBS ES-qualified	Hyclone	SH30070	-20ºC	Heat-inactivate, -20ºC
Formaldehyde	Sigma	252549	-	RT
Formamide	Ambion	AM-9342	-	4ºC
Gel red	Biotium	41603-01	10000x, RT	500x
Gelatin 2%	Sigma	G-1393	4ºC	0.1% in PBS, 4ºC
Glucose	Sigma	C-6152	-	RT
Glucose oxidase	Sigma	G2133	37 mg/mL in 50 mM Sodium Acetate, -20ºC	3.7 mg/mL in 50 mM Sodium Acetate, -20ºC

Glutamine	GIBCO	25030-123	200 mM	100x, -20ºC
GMEM	GIBCO	21710-025	-	1x, 4ºC
Non-essential aminoacids	GIBCO	11140-035	-	100x, 4ºC
PenStrep	GIBCO	15140-122	-	100x, -20ºC
Propidium Iodide	Invitrogen	P-3566	1 mg/mL, 4ºC	1 ng/mL, 4ºC
rTaq Polymerase	GE Healthcare	27-0798-05	-20ºC	-20ºC
Saponin	Sigma	S7900	-	RT
SeaKem LE Agarose	Lonza	50001	-	RT
Sheep Serum	Invitrogen	S3772	-	-20ºC
Sodium Pyruvate	GIBCO	11360-039	-	100x, -20ºC
Triton	Sigma	T8787	-	RT
Trypan blue	Sigma	T8154	RT	0.4% in PBS, RT
Trypsin	GIBCO	25090-028	2.5% (v/v), -20ºC	0.25% or 0.025% in PBS, -20ºC/4ºC

Table 3.2 – List of solutions/media and respective components which were used in the experiments performed in this project.

Solution	Components
1x TAE buffer	40 mM Tris; 1mM EDTA; 0.35% glacial acetic acid
4% PFA	4% (w/v) paraformaldehyde; PBS
Anti-fade buffer	1% (v/v) of catalase and glucose oxidase (1 µL each) in glox buffer (100 µL)
Gelatin 0.1%	2% gelatin; PBS
Glox buffer	85% (v/v) H_2O; 10% (v/v) 20x SSC; 4% (v/v) glucose; 1% (v/v) Tris 1M pH=8; 1% (v/v) Triton (850 µL total)
GMEM 1x	80% (v/v) GMEM; 1% (v/v) Glutamine; 1% (v/v) Pen-Strep; 1% (v/v) Sodium Pyruvate; 1% (v/v) non-essential Aminoacids; 10% (v/v) inactivated FBS; 0.001% (v/v) of 2-mercaptoethanol
Hybridization buffer	14% (w/v) dextran sulphate in H_2O; 10% (v/v) formamide; 10% (v/v) 20x SSC
Solution A	10 mM Tris-HCl pH 8.3; KCl 100mM; $MgCl_2$ 2.5 mM
Solution B	10 mM Tris-HCl pH 8.3; $MgCl_2$ 2.5 mM; 1% (v/v) Tween20; 1% (v/v) Triton x100; 120 µg/mL proteinase K
Trypsin 0.025%	0.25% Trypsin; PBS
Trypsin 0.25%	2.5%(v/v) Trypsin; 0.01% (v/v) chicken serum; 0.02% (v/v) 0.5M EDTA; PBS
Wash buffer for PCR	10 mM Tris-HCl pH 8.3; KCl 50mM; $MgCl_2$ 1.5 mM
Wash buffer for smFISH	10% (v/v) 20x SSC; 10% (v/v) formamide; H_2O

Table 3.3 – List of chemical modulators used during cell culture in this project.

Chemical Modulator	Supplier	Cat. Number	Stock	Working Stock
GSK343	Sigma	SML0766	1 mM in DMSO, -80ºC	1 mM in DMSO, 4ºC
L-Ascorbic Acid2-Phosphate (AA)	Sigma	A8960	50 mg/mL in sterile H_2O, -80ºC	50 mg/mL in sterile H_2O, 4ºC

Table 3.4 – List of antibodies used for intracellular staining flow cytometry experiments performed in this project. *It is shown the supplier, number, animal in which it was produced and dilution, for each antibody.*

Antibody	Supplier	Number	Host	Dilution
Anti-α-rabbit Alexa 647	Molecular Probes	A21244	Goat	1:400
Anti-H3K27me3	Cell Signalling	9733	Rabbit	1:200

Table 3.5 – List of smRNA-FISH probes used in the experiments performed in this project. *The probes were ordered to Biosearch Technologies and its hybridization temperature is 37ºC (Biosearch Technologies).*

Gene	Probe Set
Car2 (Alexa 594)	gtgacaggcagaggtgacag**;** aaggggaggagaccgtggag**;** tgattggggcagagcagaag**;** ctccattggcaatggggaag**;** tgctgtgtcaatgtccacag**;** gatatgagcagaggctgtag**;** gttgacaatgctcttggacg**;** tcaacgttaaaggagtggcc**;** attgtcctgagagtcatcaa; atctgtaggagtcactgagg; ccccagtgaaagtgaaactg**;** gttccagtgaaccaagtgaa**;** caaaacagccaatccatccg**;** gaagttagcaaaggccgcac**;** caggaagaagggagcaagga**;** tatgtccagtagtccaagtt**;** acgatccaggtcacacattc**;** ctcgctgctgacagtaatgg**;** cattgaagttcagcgtacgg**;** ctattctttagcggctgagc**;** cttaaaggacgctttgatct**;** ttagctacagagaggcggtc**;** caaatcacccagcctaactg**;** acaataccagatgcgagtcg**;** agcacaacggatgagaggta**;** gtctcatgatgtggacttgt**;** tttgcctaagttacttcagt**;** atccattgtgttgtggtatg
Nanog (Alexa 594; Cy5; TMR)	aaatcagcctatctgaaggc; cagaaagagcaagacaccaa; gaagtcagaaggaagtgagc; actcagtgtctagaaggaaa; ggttttaggcaacaaccaaa; cgagggaagggatttctgaa; cacactcatgtcagtgtgat; cagaactaggcaaactgtgg; ttcccagaattcgatgcttc; aaaaactgcaggcattgatg; agcaagaatagttctcggga; cagagcatctcagtagcaga; gaagaggcaggtcttcagag; tgggactggtagaagaatca; tcaggacttgagagcttttg; cttgttctcctcctcctcag; gagaacacagtccgcatctt; ctgtccttgagtgcacacag; tgaggtacttctgcttctga; gagagttcttgcatctgctg; atagctcaggttcagaatgg; gaaaccaggtcttaacctgc; ttgcacttcatcctttggtt; tcaaccactggtttttctgc; ttctgaatcagaccattgct; gatactccactggtgctgag; ggatagctgcaatggatgct; cagatgcgttcaccagatag; aagttgggttggtccaagtc; gtctggttgttccaagttgg; aaagtcctccccgaagttat; ctgcaactgtacgtaaggct; caaatcactggcagagaagt; tagtggcttccaaattcacc; ctaaaatgcgcatggctttc; ataattccaaggcttgtggg; tggagtcacagagtagttca; agatgttgcgtaagtctcat; gctttgccctgactttaagc; tttggaagaaggaaggaacc; caaatcactggcagagaagt; tagtggcttccaaattcacc; ctaaaatgcgcatggctttc; ataattccaaggcttgtggg; tggagtcacagagtagttca; agatgttgcgtaagtctcat; gctttgccctgactttaagc; tttggaagaaggaaggaacc
Otx2 (Cy5)	ggatggagtgagaccagata; ggcacaggcttttaaaggag; cacgttccagcactaactaa; gtgggtagatttggagtgac; ttgtttggaggcgcaaagtc; ggcggttgctttagataaga; catacccgaagtggtcagac; ctagtaaatgtcgtcctctc; aacagagcttccagaacgtc; catgaagatgtctgggtacc; cacttagctcttcgattctt; ctggagagctcttcttcttg; acttgttccactctctgaac; ctagaggggggactgaactg; tggcaatggttgggactgag; cagatagacactggagcact; ctgagtataggtcatgggat; catagccttgactataacct; ccaaagtaggaagttgagcc; ataagatccacagtccatgc; taacagcattggtacccatg; ctggggactgattgagatgg; atatccctgggtggaaagag; agtggttgagttaaaaccca; ttggtccttataatccaagc; aagttaagcttccaagaggc; ataatccaagcagtcagcat; tggaatttccatgaggacgt; ttcagcccagcatatttaaa; tttaaccaatgcctggctaa; agagcatcgttccatctaac; ctcggtaactttgatcagga; gttgatggacccttctaagg; taaaacaccggatcacctct; ccattcctaagattcaacca; gaaacgtgaatgagcctggg; ttttcagtgccaactacctg; aatccacacagccctgaaaa; catctaggacaatcagtcgc; catattgactccgtatgagc; cagttttgaagtctagcaca; ggagttcaaggttgcataca; ttaatcacagaagaacccct; cagttgctctgaattttgct; cccaaggtaatctttctaga; ggtgaattagggtccttttg; aaagtcatcagggtcagagc; tccagttaacatctgcaagc

Sox2 (Alexa 594)	ccgtctccatcatgttatac; tccgggctgttcttctggtt; ataccatgaaggcgttcatg; ttctcctgggccatcttacg; atctccgagttgtgcatctt; tcggacaaaagtttccactc; ttataatccgggtgctcctt; tcatgagcgtcttggttttc; ggaagcgtgtacttatcctt; tagctgtccatgcgctggtt; ttgctccagccgttcatgtg; tcctgcatcatgctgtagct; tgcatcggttgcatctgtgc;tcatggagttgtactgcagg; ttcatgtaggtctgcgagct; agtaggacatgctgtaggtg; ttgaccacagagcccatgga; tgggaggaagaggtaaccac; aggtacatgctgatcatgtc; tgggccatgtgcagtctact; agtgtgccgttaatggccgt; aaaatctctcccccttctcca; cccaattcccttgtatctct; tactctcctctttttgcacc; ctgcggagatttttttcct; tttttccgcagctgtcgttt; aatttggatgggattggtgg; tagtcggcatcacggttttt; gaagtcccaagatctctcat; ctgtacaaaaatagtccccc; tatacatggtccgattcccc; gcgtagtttttttcctccag; cctaacgtaccactagaact; aagacttttgcgaactccct; ccggagtctagctctaaata; ctgtacaaaagttgcttgca; gattgccatgtttatctcga; caagaaccctttcctcgaaa; aagctgcagaatcaaaaccc; ccttgtttgtaacggtccta; ccagtacttgctctcatgtt; aacaagaccacgaaaacggt; acaatctagaacgtttgcct; gatatcaacctgcatggaca; gggtaggattgaacaaaagc; cggaaaataaaaggggggaa; ccaataacagagccgaatct; tatacatggattctcggcag
Sox3 (TMR)	ttctctcgagctggtcgcat; cgggcttctctcacctgatg; acaccatgaacgcgttcatg; ggttctccagggccatcttg; atctcggagttgtgcatctt; catcggtcagcagtttccag; cttggcctcgtcgatgaacg; gtacttgtagtccgggtact; gagcagcgtcttggtcttgc; gcagcgagtacttgtccttc; ttcacgtgcgtgtacgtgtc; ctcctgcacgagcgagtagg; atgtcgtagcggtgcatctg; gagagctgggctccgacttc; ggtacatgctgatcatgtcg; accgttccattgaccgcagt; gagcaaagctaaacagcaag; catcttcggtacaaggcaac; gacagttacggccaaacttt; ggacttctcgcttttgtaca; gctctagcaagtcccatttc; gaacctaggaatccgggaag; gacattttcaactgcaacag; gggcaacctcactcagttct; tggaggcattgcagttcttg; aacattggctttagctgtcg; aactcaacagcctaaacgcg; agcaaatagatcactgcaga; gaacgaaatgcgtacacgaa; actttgaaaaaacctggaac

Table 3.6 – Optical filters for RNA detection in smFISH, used in the widefield microscope Zeiss Axio Observer. *A new TMR filter was bought, with similar settings to the previous TMR filter. Excitation and emission wavelength values are depicted, in nanometres, as well as the supplier in the last column. The first and second numbers refer to the centre and width of the bandpass region, respectively. The beam splitter is the cut-off wavelength. Adapted from Guedes* et al. *2016.*

	Excitation [nm]	**Beam splitter or dichroic [nm]**	**Emission [nm]**	**Supplier**
TMR (new)	539/21	556	576/31	Chroma
TMR (old)	546/10	560	580/30	Chroma
Alexa 594	590/10	610	630/30	Omega
Cy5	640/30	660	700/75	Chroma
DAPI	365/12	395	> 397	Zeiss

3.2 Methods

3.2.1 Cell Culture

3.2.1.1 Expansion of Undifferentiated Mouse Embryonic Stem Cells

Stem cell culture of E14 and Nd cell lines was performed in a sterile laminar flow hood class II, type A/B3. The hood and equipment were decontaminated with ultraviolet (UV) radiation before and after each utilization, during at least 20 minutes.

The medium used in stem cell manipulation during pluripotency was Glasgow Modified Eagle Medium 1x, as described in Table 3.2. In “Serum/LIF” conditions, the leukemia inhibitory factor (LIF) was added to GMEM 1x, at 2 ng/mL.

Frozen mES cells ($3x10^6$ cells per vial), preserved at -80ºC or in liquid nitrogen, were thawed by heating for approximately 1 minute in a 37ºC hot water-bath, until the media colour changed from orange to pink. Afterwards, cells were rapidly ressuspended in pre-heated GMEM+LIF (until defrosting) and centrifuged at 1200 rpm for 4 minutes, in order to remove the remnant dimethyl sulfoxide (DMSO). Then, the supernatant was removed, cells were again ressuspended in pre-heated GMEM+LIF and transferred into a plate previously coated with gelatin (at least 10 minutes). It was necessary to change the medium for new GMEM+LIF 6 hours after and, then, to pass the cells in the next 24h.

The stem cell expansion occurred in a 37ºC incubator with 5% CO_2 (v/v), completely sterile. The medium used for expansion during pluripotency was also GMEM 1x. Stem cells were passed every 48h with a constant plating density of $3x10^4$ cells/cm^2, which corresponds to $6.45x10^5$ cells in a P60.

Firstly, cells were washed twice with phosphate buffer saline (PBS), to remove dead cells, debris and medium. Then, cells were dissociated by adding 0.025% trypsin, incubated for 2 minutes in an incubator at 37ºC. At this step, it was necessary to confirm if cells were correctly dissociated by plate movements. A serum containing medium (GMEM 1x, 4x the volume of trypsin) was added to stop trypsinization and cells were collected into a 15 mL falcon tube, for centrifugation at 1200 rpm for 4 minutes. Later, the supernatant was removed, cells were ressuspended in GMEM 1x and counted in a counting camera, in order to determine the cell density. Also, the cell viability was inferred through the trypan blue exclusion test. Afterwards, the “mixes” containing cells, new GMEM 1x and LIF were prepared and, immediately before plating, the gelatin was removed from the plate.

The morphology of mES cells was observed before each passage and pictures were taken in an inverted bright field microscope. Moreover, the percentages of Nanog:VNP positive cells were checked after each passage, through flow cytometry analysis, in Accuri C6 flow cytometer.

In order to prepare cell stocks, stem cells were dissociated as described and $3x10^6$ cells were collected for each vial, with 1 mL GMEM+LIF supplemented with 10% DMSO (v/v). Then, cells were rapidly moved into a -80º camera for short-term storage, or to liquid nitrogen, for longer periods of storage. Simultaneously, a cell suspension sample was collected from each cell stock for *Mycoplasma* detection (described in section 3.2.1.2).

3.2.1.2 Mycoplasma *Test*

Mycoplasma testing was performed every time cells were frozen, in order to verify if there was a contamination amongst the stem cell culture. From each sample, $1x10^6$ cells were collected and washed with PBS, followed by centrifugation for 5 minutes at 2000 rpm. Afterwards, cells were ressuspended in wash buffer (described in Table 3.2), centrifuged again as described and the supernatant was discarded. At this point, pellets were stored at -20ºC until testing.

Stem cell pellets were ressuspended in a mix of solution A+B (in 1:1 proportion, composition in Table 3.2), incubated for 1 hour at 60ºC and proteinase K-inactivated for 1 hour at 90ºC. Then, samples were tested by Polymerase Chain Reaction (PCR) or stored at -20ºC until tested.

For the amplification step, primers Pr_{27} and Pr_{22} were used to amplify a *Mycoplasma* specific genomic region corresponding to the highly conserved *16s ribosomal RNA (rRNA)* gene, allowing the detection of 30 different *Mycoplasma* species. The reaction mix, performed in a final volume of 25 µL, included: 16 µL of H_2O miliQ, 1 µL of Pr_{27}, 1 µL of P_{r22}, 2.5 µL of PCR buffer 10x, 1 µL of each deoxyribonucleotide phosphate (0.2 mM for each dNTP: dCTP, dGTP, dATP and dTTP, Table 3.7) 10 mM, 0.4 µL of rTaq polymerase and 3 µL of DNA sample.

As a quality control of the DNA extraction, the PCR amplification of each sample was done in parallel with primers (Pr_{DIR} and Pr_{REV}) for *glyceraldehyde 3-phosphate dehydrogenase (GAPDH)*, a house-keeping gene. This control was performed to prevent false-negative results caused by the unsuccessful inactivation of proteinase K, leading to the degradation of the Taq polymerase within the reaction mix. Only the samples that produce no band in *Mycoplasma* specific PCR but a band in *GAPDH* PCR were considered as *Mycoplasma*-free. Furthermore, a plasmid carrying the *Mycoplasma* 16S *rRNA* was used as a positive control for the amplification and an ultrapure water sample as a negative control.

The PCR conditions included an initial cycle of denaturation for 5 minutes at 95ºC, followed by 30 cycles of denaturation for 30 seconds at 95ºC, annealing for 1.5 minutes at 58ºC and extension for 1.5 minutes at 72ºC, and, at the end, a final cycle of extension for 10 minutes at 72ºC. After the PCR reaction, gel electrophoresis was performed to analyse the DNA fragments produced, in 1.5% agarose gel. Agarose was heated in 1x TAE buffer until complete melting, followed by adding GelRed in a 1:20 dilution, to stain the PCR products. The gel was run at 60 volts for 60 to 90 minutes and the PCR products were observed under UV light, using Chemidoc XRS+. Data was analysed using Bio-Rad Image Lab Software and the fragment's size were predicted by comparison to linear DNA strands of known molecular weight (1 kb Plus DNA Ladder-Invitrogen).

Table 3.7 – Characteristics of the primers used for* Mycoplasma *PCR.

Gene	Company	Sense Primer	Antisense primer	Product size [bp]	Annealing temperature [ºC]
16S rRNA	Sigma	TGCACCATCT GTCACTCTGTT AACCTC	ACTCCTACGGG AGGCAGCAGTA	717	58
GAPDH	Sigma	ATTCAACGGC ACAGTCAAGG	TGGATGCAGGG ATGATGTTC	580	60

An additional *Mycoplasma* test was performed through the company GATC Biotech, entitled "MYCOPLASMACHECK", which employs quantitative PCR (qPCR) technology. Samples for test were prepared under manufacturer's conditions.

3.2.1.3 Chemical Modulators Assay

Chemical modulators were used in order to study the influence of the chromatin environment on *Nanog* expression. The tested conditions were GSK343 and Ascorbic Acid (AA). Cells were grown in "Serum/LIF", as it was explained in section 3.2.1.1 and incubated, for 48h, with each of these drugs or in combination (AA+GSK343). After 24h of drug exposure, the medium was replaced for new "Serum/LIF" medium, supplemented with the chemical modulator. After 48h of drug exposure, cells were dissociated, as described in section 3.2.1.1, and fixed for smFISH or for intracellular stainning - flow cytometry (in future sections 3.2.3 and 3.2.2.3, respectively).

Moreover, following dissociation, the percentages of Nanog:VNP positive cells were analysed by flow cytometry (detailed in future section 3.2.2.1). DMSO was always done in parallel with the chemical modulators, to act as a control. The working stock (WS) and final concentration of each chemical modulator is shown in Table 3.8 (values obtained by $c_i v_i = c_f v_f$, in which c stands for concentration, v for volume, i for initial and f for final). The concentration of DMSO used was the highest in a condition with DMSO as a solvent (GSK343).

Table 3.8 – Concentration of chemical modulators for WS and final concentration in culture medium.

Chemical Modulator	Working Stock (WS)	Final Concentration
GSK343	**1 mM:** 1 mg stock + 1.864 mL DMSO	**1 µM (c_f):** 1 µL (v_i) from 1 mM WS (c_i) per 1 mL of culture medium (v_f) (1:1000 dilution)
AA	**50 mg/mL:** 1 g stock + 10 mL H_2O	**100 µL/mL:** 2 µL from 50 mg/mL WS per 1 mL of culture medium

3.2.2 Protein Expression Analysis

3.2.2.1 Flow Cytometry Analysis

In order to quantify the percentages of Nanog:VNP positive cells in live cells, flow cytometry analysis was performed. After dissociation, a sample of approximately $6x10^5$ cells in suspension was taken from each condition (Nd or E14, with or without chemical modulator or DMSO) and ressuspended in PBS, in a dilution of 1:4.

Afterwards, cells were stored at 4ºC and shortly after (maximum 1 hour) analysed in a BD Accuri C6 flow cytometer. Firstly, cells were gated based on the Forward Side Scatter (FSS) and the Side Scatter (SSC), which characterize the size and complexity of cells, respectively. Through this first gate, low-size objects were discarded and only cells were further analysed. Secondly, non-viable cells were excluded by propidium iodide dye exclusion assay. In the end, 10.000 events of viable cells were acquired and Nanog:VNP positive cells were gated based on the exclusion of the non-fluorescent parental cell line E14 (which does not contain Nanog:VNP signal). The flow cytometry analysis workflow is described in Figure 3.1. Data analysis was done using *FlowJo* software.

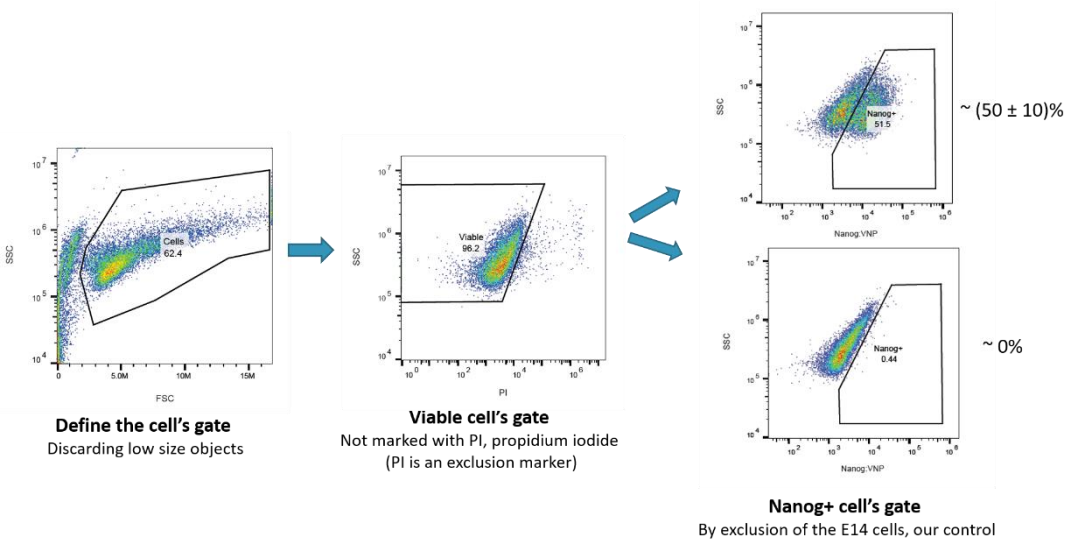


Figure 3.1 – Flow cytometry analysis workflow, in order to obtain the percentages of Nanog:VNP positive cells. *First a gate for cells was defined, discarding the low size objects, afterwards, a gate for viability was draw based on the PI levels. Finally, Nanog:VNP positive cells were gated by exclusion of the E14 cell's gate.*

3.2.2.2 Intracellular Stainning - Flow Cytometry (IC-FC)

After dissociation, as described in section 3.2.1.1, cells were ressuspended in PBS. Then, cells were centrifuged for 5 minutes at 4000 rpm, the supernatant was removed, cells were ressuspended in 500 µL of 4% PFA in PBS and incubated at room temperature for 10 minutes. Afterwards, 1 mL of PBS was added and cells were centrifuged for 5 minutes at 4000 rpm. Cells were washed in PBS to remove PFA residues. The supernatant was removed, cells were ressuspended in 1 mL of PBS and stored at 4ºC for a maximum of one month.

For each experiment, approximately $1x10^6$ cells were collected (enough to form a visible pellet). Cells were centrifuged 5 minutes at 4000 rpm, the supernatant was removed, cells were ressuspended in 100 µL of PBS/Saponin 0.25% in PBS/Sheep serum 5% and incubated for 60 minutes at room temperature. Later, primary antibodies were distributed per each sample (identification in Table 3.4 and dilutions in Table 3.9) and incubated for 60 minutes at 4ºC.

Afterwards, cells were washed twice with 1 mL of Saponin 0.25%, with centrifugation for 5 minutes at 4000 rpm. Cells were ressuspended in 100 µL of the Alexa 647 secondary antibody diluted to the 1:1000 in PBS/Saponin 0.25%/Sheep serum 5% and incubated for 60 minutes at 4ºC. Then, cells were washed twice with 1 mL of Saponin 0.25%. Finally, cells were ressuspended in 500 µL of PBS and transferred to a flow cytometry tube coated with BSA, in order to prevent cell adhesion to the tube surface.

Samples were further analysed in BD Accuri C6 flow cytometer. Firstly, cells were gated according to the FSS and SSC and antibody-positive cells were gated based on the non-fluorescent negative controls. Control samples included cells stained exclusively with secondary antibodies (*Anti-α-rabbit Alexa 647*) and single stainning for anti-H3K27me3 antibody. In each acquisition 10000 events of viable cells were recorded and subsequently analysed using the *FlowJo* software.

Table 3.9 – Dilutions of the antibodies used for IC-FC.

Antibody	Intermediate Dilution (Working Stock - WS)	Final Dilution	Total Dilution
Anti-α-rabbit Alexa 647 (secondary)	**1:10** 10 µL stock + 90 µL PBS + 1% BSA	**1:40** 2.5 µL WS + 97.5 µL PBS/Saponin/SS	**1:400**
Anti-H3K27me3	**1:10** 10 µL stock + 90 µL PBS + 1% BSA	**1:20** 5 µL WS + 95 µL PBS/Saponin/SS	**1:200**

3.2.2.3 Fluorescence Activated Cell Sorting (FACS)

FACS was used to separate the Nd cell line population into low-Nanog:VNP and high-Nanog:VNP. First, approximately $120x10^6$ cells were collected after dissociation and ressuspended in PBS, in FACS tubes coated with 1% BSA in PBS. Cells were sorted in ARIA III cell sorter, defining the cell and Nanog:VNP gates as described in section 3.2.2.1. After each sorting, sorter purity was measured by analysis of Nanog:VNP expression in live cells (both at ARIA III cell sorter and BD Accuri C6 flow cytometer) and cell viability was assessed using the trypan blue dye exclusion method, as described in section 3.2.1.1. Afterwards, cells were replated in “Serum/LIF” conditions with chemical modulators or DMSO for 48h, followed by fixation for smFISH. In each acquisition 10000 events of viable cells were recorded and subsequently analysed using the *FlowJo* software.

3.2.3 *Single Molecule RNA Fluorescence* In Situ *Hybridization (smFISH)*

Measurements of gene expression within single cells have revealed a huge variability otherwise hidden in bulk averages. The smFISH is a method capable of detecting individual mRNA molecules in each cell, thus allowing the accurate quantification and localization of mRNAs. The strategy involves probing target mRNAs using more than 30 short DNA oligonucleotides (20 bases), each of which hybridize to a different portion of the same target mRNA. Each oligonucleotide is labelled with a single fluorophore at its 3' end. Therefore, upon hybridization, many fluorophores are brought close to the mRNA target, emitting a detectable signal (Raj *et al.* 2006; Raj & Tyagi 2010; Batish *et al.* 2011). The Stellaris RNA FISH probes used in these experiments were ordered to LGC Biosearch Technologies. In the next paraphs, it will be described the experimental procedure used (Figure 3.2).

Following dissociation, cells were washed with PBS, centrifuged for 2 minutes at 3000 rpm (same centrifugation parameters along the experiment) and ressuspended in 4.5 mL of PBS. Then, 500 µL of 37% formaldehyde were added and homogeneously mixed, followed by an incubation for 10 minutes at room temperature. Afterwards, cells were washed twice with PBS, to remove formaldehyde residues, and ressuspended in 1-3 mL of ethanol 70% for permeabilization (1 mL per $3x10^6$ cells). Cells fixed for smFISH in ethanol can be stored at 4ºC for years without RNA degradation.

It was collected 200 µL of fixed cells for each experiment, followed by centrifugation. Later, cells were ressuspended in 850 µL of wash buffer with 0.25% triton. Simultaneously, the hybridization mix was prepared by joining 100 µL of hybridization buffer with 1 µL of each probe (1:100 dilution). Cells were centrifuged, ressuspended in the hybridization mix and incubated overnight at 37ºC, protected from light. On the next day, 850 µL of wash buffer were added and cells were centrifuged. Then, cells were washed with 850 µL of wash buffer and incubated for 30 minutes at 37ºC. Afterwards, cells were centrifuged, ressuspended in 850 µL of wash buffer plus 1 µL of 1 mg/mL of DAPI and incubated, again, for 30 minutes at 37ºC. Cells were centrifuged and ressuspended in 850 µL of glox buffer. Finally, cells were centrifuged and ressuspended in 10 µL of anti-fade buffer (to prevent signal degradation). After suspension, cells were mounted between slide and coverslip and carefully smashed to reduce their volume and to increase signal quality. In the end, the sample was carefully sealed with silicone-based vacuum grease Dow Corning.

Cells were imaged within 24 hours on an inverted wide-field fluorescence microscope, the Zeiss Axio Observer, using a high numerical aperture (1.40) 100x oil-immersion objective, a Zeiss Axiocam 506 mono camera and filter sets suitable to the fluorophores used (Alexa 594, Cy5 and TMR). For each image were taken 20 stacks, with 0.3 µm step size each and a total of 80 to 130 positions were acquired, each channel imaged with long exposure times (5-6 seconds). Segmentation of cells in each position and thresholding of mRNA molecules (for each channel in individual cells) were done using code developed by the Raj Lab (Bitbucket, 2016), for Matlab software. Finally, statistical analysis was done using the RStudio software, with a code written according to the user's necessity. Some of the most important R functions used were: *melt*, *ddply*, *ggplot* and *transform*.

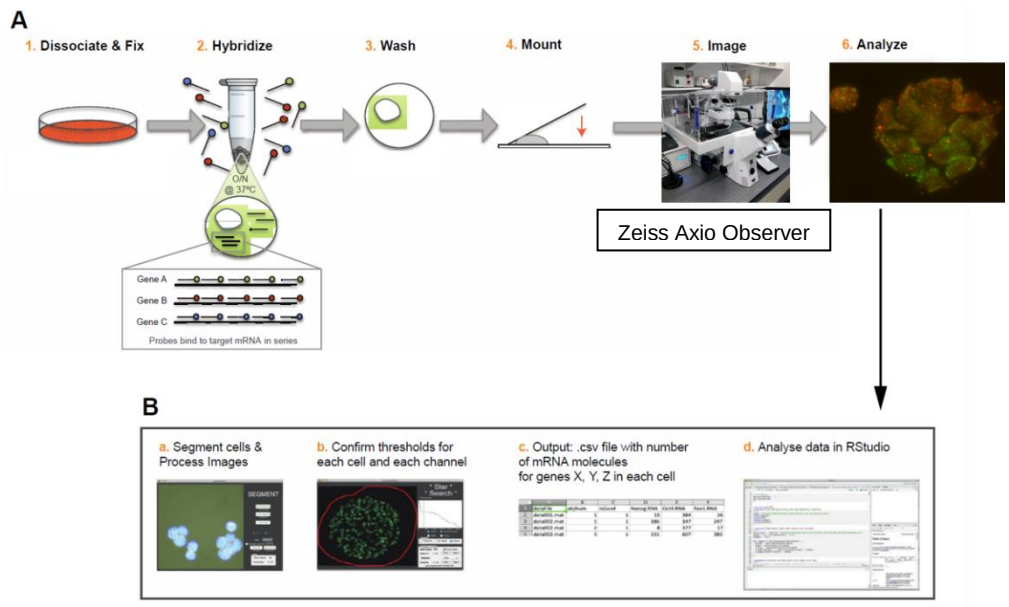


Figure 3.2 – smFISH workflow. *Part A refers to sample preparation and signal acquisition whilst part B refers to data analysis. First, it is necessary to design and synthetize fluorescently labeled oligonucleotides complementary to the target mRNA (order Stellaris RNA FISH probes). Then, mES cells should be fixed in ethanol (1), followed by hybridisation with until three probes at the same time (2). After washing steps (3) and sample mounting (4), the signal is acquired in the widefield microscope Zeiss Axio Observer (5) and finally* in silico *analysis is performed (6). Part B involves data analysis using Matlab (segmentation and thresholding) and RStudio (statistical analysis). Adapted from Guedes* et al. *2016.*

Differences between Microscopes: Zeiss Axiovert 200M and Zeiss Axio Observer

Due to technical limitations, it was necessary to use the Zeiss Axio Observer instead of Zeiss Axiovert 200M microscope system, for the smFISH. One of the most important differences between these systems lies in the fact that the new camera, the Axiocam 506m (which works at 18 ºC), is not a cooled CCD (charge-coupled device) camera, like the Coolsnap HQ of Axiovert 200M (works at -30 ºC), which is a specific requirement for smFISH and might influence the results (Zeiss 2013; Photometrics).

Nevertheless, the read noise (random unwanted signal in the image) and dark current (small electric current that flows in the device even when there are no photons entering it) are very similar, thus it should be suitable for smFISH as well (QSI 2013; McFee). Moreover, the quantum efficiency, the percentage of photons hitting the device's photoreactive surface that produce charge carriers, of the new camera is higher for the filter's wavelengths used (~ 75% compared to 60% of Coolsnap HQ) (Zeiss 2013; Photometrics). Since the systems are distinct, an optimization for smFISH on the new system was carried during this project and the next topics will cover some differences.

Split channels

The new microscope system (Zeiss Axio Observer) provides as an output a czi file comprising all the positions (80 to 130) acquired. This czi does not serve as an input for the Matlab software used, which only accepts separate tiff images named "alexa#", "cy#" and "tmr#", where # is the number of each position. Thus, the Matlab script "SplitChannels" was developed in collaboration with the iMM

Bioimaging unit in order to convert the unique czi file into tiff images, separating individually images per position and also per channel.

Hot pixel problem

Another limitation relatively to the new system is the occurrence of hot pixels for long acquisition times, which is the case of the smFISH experiments (5 to 6 seconds). A hot pixel is an individual bright high-intensity pixel, caused by electrical charges that leak into the sensor wells, which can be related to an increase of temperature in the camera sensor (Premium Beat, 2016). The existence of hot pixels interferes with the smFISH analysis since the software detects the brightest pixels as mRNA molecules, which is not the case, affecting the determination of the correct threshold, for each cell. In order to efficiently remove these hot pixels, it is now used the “Noise Filter” option, available in the Zeiss software.

TMR filter

In the first smFISH analysis from the new system it was systematically detected a higher number of mRNA molecules in the TMR channel than the expected by the previous system Zeiss Axiovert 200M. Therefore, controls were made to ensure that the filters were suitable to be used for smFISH analysis. Samples were marked with probes from only one fluorophore and the signal was verified in the other channels, in order to detect if some signal was being transmitted non-specifically (“bleedthrough”). When the Alexa-only sample was used, there was signal in TMR, which supported the hypothesis that TMR filter was no longer suitable to use. Due to the long exposition times and recurrent use, this filter might have become damaged. Therefore, a new TMR filter, with very similar settings, was bought and tested.

Temperature stabilization

The Zeiss Axio Observer displays an incubator which allows for efficient temperature stabilization of the sample. This was a feature unavailable in the previous system Zeiss Axiovert 200M, and whose absence caused many complications associated to sample unfocus in Z during overnight acquisitions. Consequently, the data produced would not be reliable and much more difficult to analyse, since it was unfocus. In the new system, with temperature stabilization at 23ºC (optimal temperature found) it is possible to keep the sample focus during longer periods of time. The new microscope displays a definitive focus system; however, it did not work properly. The most probable cause is the use of a medium which is not 100% water, changing the ideal refraction index.

4. Results and Discussion

4.1 *Nanog* Characterization and Lineage Priming

E14 and Nd mES cells were cultured in monolayer, in "Serum/LIF" conditions for seven days until their fixation for single molecule RNA FISH (smFISH) experiments (Figure 4.1). The Nd cell line contains a dynamic fluorescent Nanog:VNP reporter to monitor NANOG expression (Abranches *et al.* 2013). Every other day mES cells were passed and the expression of Nanog:VNP reporter in Nd cells was assessed by flow cytometry (FC) analysis, using E14 cells as a control. Cell morphology was also daily monitored on an inverted bright field microscope.

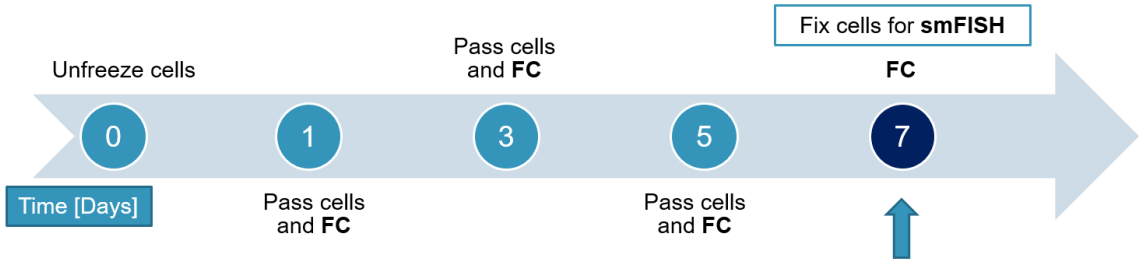


Figure 4.1 – Scheme of the experiment procedure for ES cell culture in "Serum/LIF" conditions. *After 3 cell passages, with Nanog:VNP expression monitored every 48 hours by flow cytometry, cells were fixed for smFISH.*

4.1.1 Nanog in Pluripotency

Both E14 and Nd mES cells were cultured in pluripotency conditions ("Serum/LIF") and characterization was performed by monitoring cell morphology and Nanog:VNP expression in Nd cells. Contrarily to the homogeneity in morphology and expression observed in "2i" conditions, cells cultured in "Serum/LIF" showed a marked morphological heterogeneity, as a result of the differential expression of pluripotency regulators and lineage-affiliated genes that characterizes these mES cells (reviewed in Marks & Stunnenberg 2014). Both cell lines grew in clusters through several passages (at least 7) and cells presented large nucleoli, scant cytoplasm and a wide variety of shapes: some more round-shaped than others, the differentiated-like cells, which presented elongations and were more adherent to the plate disk, previously coated with gelatin. Differentiated-like cells were found at clusters' periphery and isolated between colonies (Figure A1, from Annex A). Fold increase (FI) and viability were calculated to both cell lines and are within normal values for these cultures (Abranches *et al.* 2013). The mean value for viability is above 92% and there are similar proliferation rates between cell lines (FI from 5 to 7 after the first passage) (Figure A2).

The Nanog:VNP reporter expression in Nd cells was monitored every 48 hours and it was within the range of (56.2 ± 8.0)% of positive cells (Figure A3), previously described for Nd cells (Abranches *et al.* 2013). For example, in Figure 4.2.A is depicted one measurement for Nd population showing that approximately 55.6% are Nanog:VNP positive cells. Nanog:VNP values were constant within the same experiment (Figure 4.2.B) and also across experiments (Figure A3).

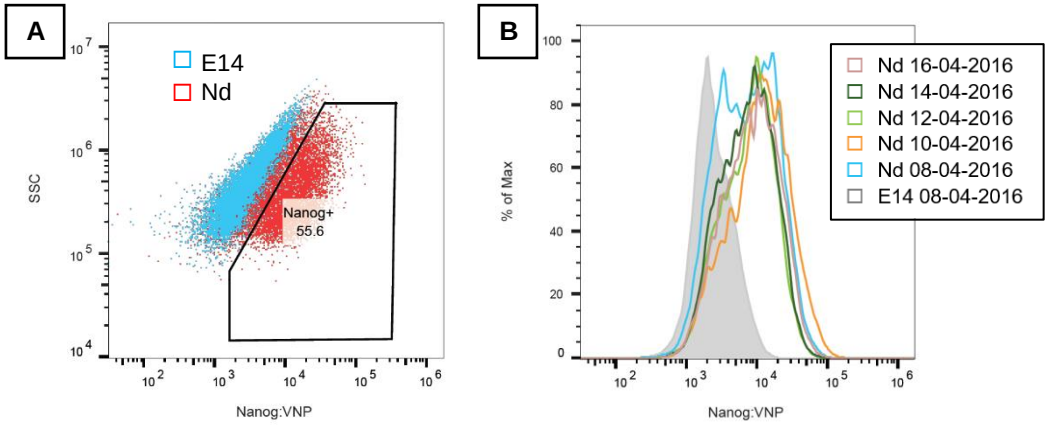


Figure 4.2 – Nanog:VNP expression in Nd cells. (A) Scatterplot *of Nanog:VNP expression in relation to SSC, of Nd cells at 16-04-2016, where E14 cells are depicted in blue and Nd cells in red.* ***(B)*** *Representative flow cytometry profile of Nanog:VNP expression for Nd cells grown in "Serum/LIF" and replated every 48 hours. E14 cells were used as a control (depicted in filled grey).*

While population studies give a perspective of the average behaviour of a group of cells, the heterogeneity that characterizes ES cells during pluripotency requires a single cell analysis. Indeed, the number of mRNAs and proteins can vary significantly from cell-to-cell within the same population, due to the inherent stochastic nature of the biochemical events involved in gene expression, as explained in section 1.3.3.

It was hypothesized that the observed NANOG heterogeneity might arise from stochastic fluctuations within individual cells (as described in section 1.3.2), which would, in turn, be caused by cell-to-cell mRNA variability. Therefore, to study how *Nanog* expression correlates with priming gene expression, at a single-cell level, a modified mRNA FISH method, smFISH, was used. This is a quantitative method that allows to identify the number of mRNA transcripts per each cell and gene, in a large population of cells (Raj *et al.* 2006). It was possible to correlate until three genes at the same time in each smFISH experiment, always considering *Nanog* as one of those genes, not only to obtain correlations with priming gene expression, but also to prove the reproducibility of this technique.

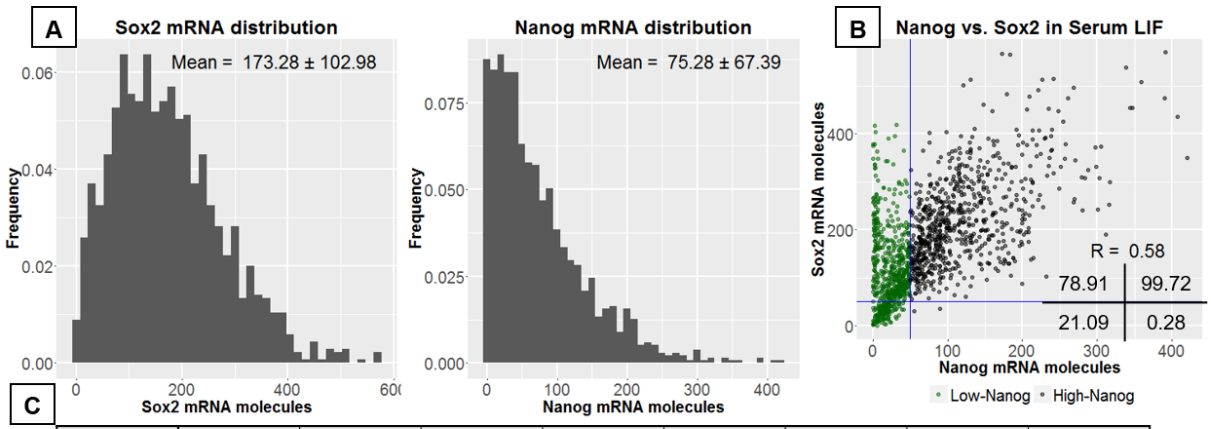
A histogram representing the distribution of the number of mRNA transcripts per cell frequency was chosen to represent the results obtained in this work. The mRNA distribution can have distinct shapes, which might be informative about the transcriptional activity. A gene can present a bell-shaped or "Gaussian-like" distribution (but not a Poisson model), in other words, most of the cells expresses an average level of mRNA molecules and only a very small group of cells shows mRNA values distinct from the mean. In these cases, transcriptional bursts might occur more frequently giving the illusion that transcription is continuously active. The gene promoter is "ON" most of the time, activating transcription, which results in less cell-to-cell variability and more homogeneous mRNA expression. This might be the case of *Sox2* in pluripotency conditions ("Serum/LIF") (Figure 4.3.A).

Other genes, such as *Nanog*, present a long-tailed non-Poissonian distribution (perhaps approximates to an exponential), in which many cells exhibit lower levels of mRNA transcripts and only few cells express higher levels of mRNA transcripts (Figure 4.3.A). This distribution reflects a pulsatile or bursty transcription, which occurs in short and rare periods of time when the transcription is active,

since most of time the promoter is in a transcriptionally inactive state. The end result is the observed high cell-to-cell variability of mRNA levels (Raj & van Oudenaarden 2009). These differences also mean that genes can be transcribed with widely different bursting kinetics.

Nanog has a short mRNA half-life (average time required to degrade 50% of the mRNA) of (4.7 ± 2.5) h for E14 cells (Abranches *et al.* 2013), which means it is a good "read-out" of the transcriptional activity. If *Nanog* had a long mRNA half-life, the mRNA smFISH analysis would probably not reflect the transcriptional activity: mRNA would still be detectable but transcription would be already inactive. Simultaneously, Nanog:VNP protein half-life is 1.8 h for Nd cells, which is very short and similar to NANOG half-life of 2.3 h (Abranches *et al.* 2013).

The correlation between the expression of multiple genes at single-cell level was obtained and represented in the form of scatterplots (where each dot represents a cell), such as the one in Figure 4.3.B, and a statistical analysis was performed for each gene, by calculating parameters such as mean ± standard deviation (SD), median, variation, minimum (Min), maximum (Max), Fano factor (FF), coefficient of variation (CV) and number of cells analysed (N) (Figure 4.3.C).



C

Condition	Mean	SD	Median	Min	Max	FF	CV	N
Sox2	173.28	102.98	161.00	0.00	570.00	61.20	0.59	1349
Nanog	75.28	67.39	57.00	0.00	420.00	60.32	0.90	1349

Figure 4.3 – Analysis of Sox2 and Nanog mRNA expression in E14 mES cells cultured in "Serum/LIF". (A) *Histograms of the distribution of mRNA transcripts/cell for Sox2 and Nanog. Mean ± standard deviation is shown for each gene.* ***(B)*** *Correlation scatterplot between Nanog and Sox2 mRNA molecules, in which R is the Spearman correlation coefficient. The numbers depicted are the percentages of low-Nanog (green dots, on the left) and high-Nanog (black dots, on the right) cells with high (up) or low (down) levels of Sox2 transcripts.* ***(C)*** *Statistical measurements for Sox2 and Nanog. The parameters shown are the mean, standard deviation (SD), median, minimum (Min), maximum (Max), Fano factor (FF), coefficient of variation (CV) and the number of cells (N) analysed by smFISH.*

The spearman correlation (r_s or R) was calculated for each relation and its value varies between -1 and 1. It measures the strength and direction of a monotonic relationship between two ranked variables. The sign indicates the direction of association between genes X and Y. When R is positive, as the mRNA expression of gene X increases, gene Y tends to increase too. By contrast, when R is negative, as the mRNA expression of gene X increases, gene Y tends to decrease. When R = 0, there

is no relationship between the two variables at all. The closer to 1 or -1, the stronger is the monotonic relationship. There is a mild and positive correlation between *Nanog* and *Sox2* mRNA expression, observed by the moderate Spearman correlation value of R = 0.58 (Figure 4.3.B) (McDonald, 2014).

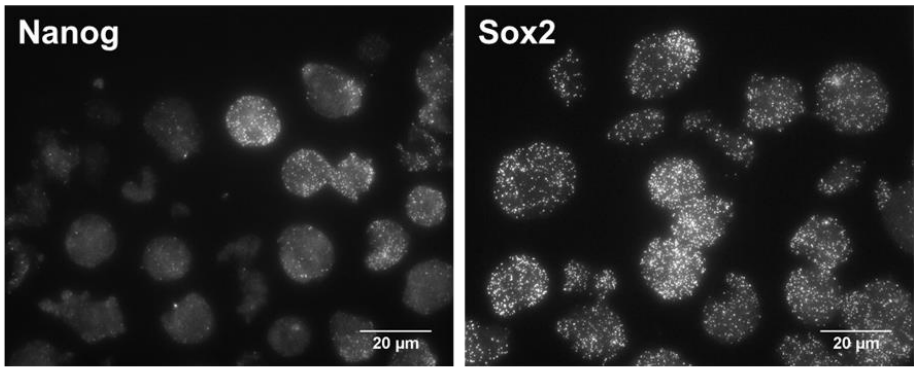
Sox2 and *Nanog* present a wide range of mRNA transcripts, varying from 0 to 570 and 420 transcripts/cell, respectively. The average of mRNA molecules per cell for *Sox2* is 173 mRNAs/cell and for *Nanog* is 75 mRNAs/cell (Figure 4.3.C).

In order to mathematically identify if a distribution is Poissonian, the Fano factor (FF) was defined as the ratio of variance to mean, being 1 for a Poisson distribution. By contrast, transcriptional bursts can generate a striking variability even when the mean is high, thus FF is much larger than 1 (Raj & van Oudenaarden 2009). Both for *Sox2* and *Nanog*, the FF values are much higher than 1 (61.20 and 60.32, respectively), which supports a bursty transcription for both genes (Figure 4.3.C).

Moreover, the coefficient of variation (CV), the ratio of SD to mean, was also calculated for each mRNA distribution, to measure the dispersion of mRNA molecules relatively to the mean. If the CV is higher than 1, the distribution has a high dispersion; if it is lower than 1, the dispersion is low. CV was calculated for *Sox2* (0.59) and *Nanog* (0.90), which supports a more heterogeneous expression of *Nanog* (since the CV is higher) compared to *Sox2* (Figure 4.3.C).

It was necessary to define thresholds for the values of mRNA transcripts/cell for each gene in order to distinguish cells in an active ("high-") from inactive ("low-") state of transcription. A "high-" cell has a higher expression of mRNA transcripts, above the threshold, whilst a "low-" cell has a lower mRNA expression, below the threshold. For *Nanog* and *Sox2* a threshold of 50 mRNA transcripts/cell was used to distinguish "high-" and "low-" expressing cells. Threshold definition was done by visual analysis of the histogram, as previously performed in other smFISH experiments. In Table A1 from Annex A are depicted the threshold values used in smFISH.

A



B

Nanog	Sox2	N	Total [%]	Nanog Total [%]	Sox2 Total [%]	Nanog [%]
+	+	721	53.45	53.60	90.07	99.72
+	-	2	0.15		9.93	0.28
-	+	494	36.62	46.40		78.91
-	-	132	9.78			21.09

Figure 4.4 – Analysis of Nanog and Sox2 expression relatively to their low and high abundance of transcripts in 1349 E14 mES cells. ***(A)*** *Heterogeneous expression of Nanog compared to a more homogeneous expression of Sox2. Scale bar = 20 µm.* ***(B)*** *Summary of the number (N) and percentage of cells expressing high (denoted by +) and low (-) levels of Nanog and Sox2 relatively to the total number of cells and only relatively to the Nanog subpopulation (in the last column).*

The observed heterogeneity in NANOG expression at the protein level is underlined by heterogeneity at mRNA level, with a population composed of approximately 46.40% low-*Nanog* cells and 53.60% high-*Nanog* cells (Figure 4.4.A and Figure 4.4.B). The percentage of low-*Sox2* cells is approximately 9.93%, revealing the existence of cells with low levels of *Sox2* mRNA even in pluripotency conditions such as “Serum/LIF” (Figure 4.4.B).

There is a very low percentage (0.15%) of cells expressing high number of *Nanog* transcripts and low number of *Sox2* transcripts, while 53.45% of cells express high abundance of transcripts of both genes, 36.62% of cells express high-*Sox2* and low-*Nanog*, and only 9.78% of cells express low number of transcripts from both genes (Figure 4.4.B).

Within the low-*Nanog* cells, 78.91% are high-*Sox2*, whilst only 21.09% are low-*Sox2* (Figure 4.4.B), which indicates that most low-*Nanog* cells are still pluripotent. These low-*Nanog* and high-*Sox2* cells are likely to correspond to the “lineage-primed” state, whereas the cells expressing low levels of both transcripts probably are already committed to differentiation.

Like *Nanog, Sox2* is a good “read-out” of the transcriptional state because it has a short mRNA half-life of (1.6 ± 0.2) h for E14 cells (Abranches *et al.* 2013). Adding to the fact that *Sox2* exhibits a bell-shape distribution, these results suggest that cells spend most of the time with the *Sox2* promoter in a transcriptionally active state. Furthermore, since *Sox2* mRNA half-life is much shorter than that of *Oct4*, which is more than 6 h for E14 cells (Abranches *et al.* 2013), *Sox2* allows a better classification of *Nanog* states within the pluripotency window than *Oct4*. A longer mRNA half-life can lead to incorrect assumptions: in the case of *Oct4*, for instance, which has a stable mRNA, the presence of transcripts can occur when *Oct4* transcription has been inactive for already some time, thereby masking an undergoing burst-like transcription.

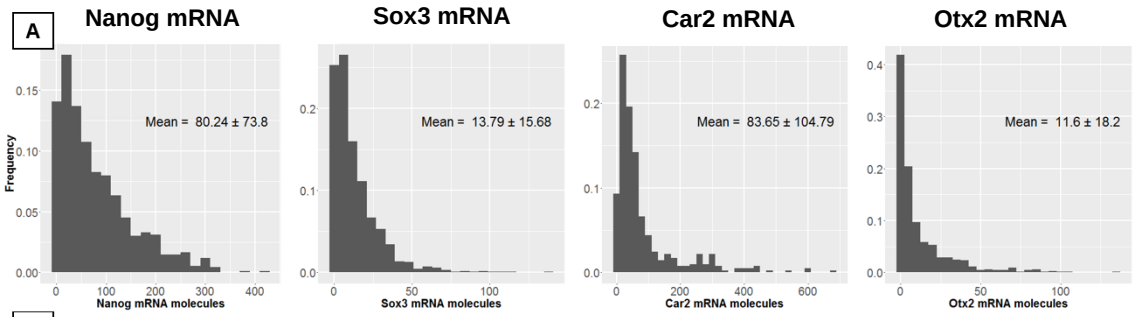
4.1.2 Analysis of Priming Gene Expression

Analysis of mRNA Distributions

Previous work in DHenrique Lab led to the identification, through RNA-sequencing data analysis, of genes preferentially upregulated either in low-Nanog:VNP cells or in high-Nanog:VNP cells cultured in “Serum/LIF”. The genes upregulated in low-NANOG cells were named priming genes, already mentioned in section 1.3.2. Amongst the priming genes, there were *Car2* and *Sox3*. *Car2* is a sporadic gene with infrequent expression, which has no known role in development and encodes for a carbonic anhydrase II, that catalyses the reversible hydration of carbon dioxide (Lindskog 1997). By contrast, *Sox3* is a lineage-affiliated gene that encodes for the SOX3 transcription factor associated to the regulation of embryonic brain development and neural lineage (Sarkar & Hochedlinger 2013).

Otx2 is a marker of the transition between ES cells, in a “naïve” pluripotent state, to the EpiS cells, in a “primed” state, defined in section 1.2.2 (Acampora *et al.* 2013). The mRNA expression of *Sox3*, *Car2* and *Otx2* in E14 cells grown in “Serum/LIF” was analysed in detail and represented in the form of histograms (Figure 4.5.A), accompanied by a statistical analysis (Figure 4.5.B).

A



B

Gene	Mean	SD	Median	Min	Max	FF	CV	N
Nanog	80.24	73.80	57.00	0.00	426.00	67.88	0.92	1088
Sox3	13.79	15.68	9.00	0.00	141.00	17.82	1.14	1088
Car2	83.65	104.79	47.00	0.00	686.00	131.27	1.25	408
Otx2	11.60	18.20	4.00	0.00	137.00	28.54	1.57	680

Figure 4.5 – Analysis of gene expression in E14 mES cells cultured in "Serum/LIF". (A) *Histograms of the distribution of mRNA transcripts/cell for Nanog, Sox3, Car2 and Otx2. Mean ± standard deviation is shown for each gene.* ***(B)*** *Statistical measurements for Nanog, Sox3, Car2 and Otx2. The parameters shown are the mean, standard deviation (SD), median, minimum (Min), maximum (Max), Fano factor (FF), coefficient of variation (CV) and the number of cells (N) analysed by smFISH.*

As it was already described for *Nanog*, these genes also present a bursty transcription, revealed by their long tailed distributions in the histograms (Figure 4.5.A). Thus, it is predicted that their transcription is inactive ("OFF" state) most of the time except when transcription bursts occur ("ON" state), leading to an increase in mRNA expression. The bursty transcription explains, in part, why priming gene expression (*Sox3* and *Car2*) is rare, with most cells presenting low mRNA levels whilst only a few cells express higher and variable levels of mRNA transcripts. In fact, priming gene expression is expected to be rare, because, in theory, only low-*Nanog* cells should present it and cells were cultured in pluripotency conditions, in which lineage-affiliated genes are not expected to be upregulated.

Priming gene and *Otx2* expression are highly variable at the single-cell level, as it can be observed by the wide range of mRNA values/cell for each gene (between minimum and maximum). The average number of mRNA transcripts/cell for *Nanog* is 80 (similar to the value obtained in the previous *Nanog*-*Sox2* analysis), for *Sox3* is 14, for *Car2* is 84, and for *Otx2* is 12 transcripts/cell. The FF is higher than 1 for all genes (67.88, 17.82, 131.27 and 28.54 for *Nanog*, *Sox3*, *Car2* and *Otx2*, respectively), which indicates that these distributions are non-Poissonian, supporting a bursty transcription (Figure 4.5.B). The FF calculated for *Nanog* (67.88) is approximately the same obtained in the previous analysis (60.32).

Furthermore, the CV is also higher than 1 for *Sox3* (1.14), *Car2* (1.25) and *Otx2* (1.57), showing that there is a high variance and dispersion of data (Figure 4.5.B). The CV calculated for *Nanog* (0.92) is approximately the same obtained in the previous analysis (0.90).

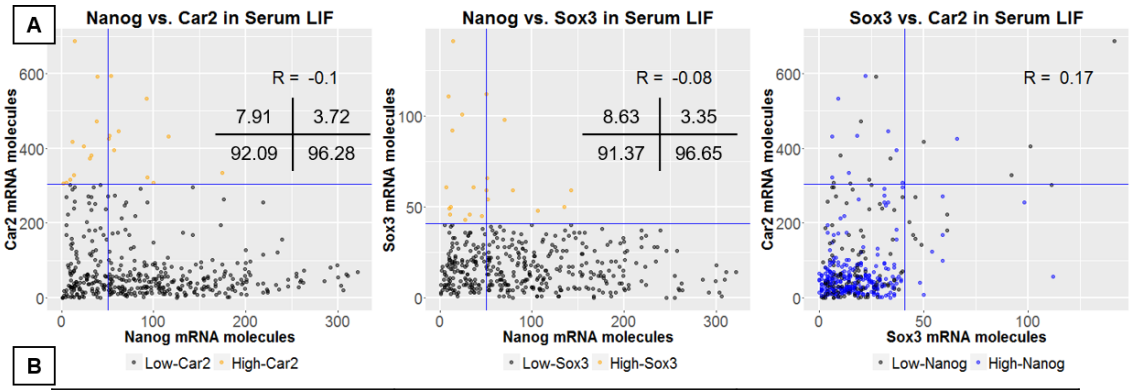
The fact that both *Nanog* mRNA analysis (first with *Sox2* and now with priming genes) provided similar values for different statistical parameters gives confidence to the results, proving the reproducibility of the technique, which was tested for the first time in the new system Zeiss Axio Observer.

In the next pages, it will be described the analysis of *Car2-Nanog-Sox3* and *Nanog-Otx2-Sox3* probe combinations, in E14 mES cells. The probe order is coincident with the fluorophore order Alexa 594-Cy5-TMR. A more detailed analysis of both smFISH experiments is depicted in Table A2 from Annex A.

Car2-Nanog-Sox3 smFISH Analysis

For *Nanog* it was used a threshold of 50 mRNA molecules/cell: low-*Nanog* cells, with less than 50 mRNAs/cells, are depicted in black and high-*Nanog* cells in blue. Relatively to *Car2* and *Sox3*, only the 5% top-expressing cells were selected as high-expressing, corresponding to more or equal than 305 mRNAs/cells and 41 mRNAs/cell, respectively, both depicted in in orange, in the correlation scatterplots for 408 E14 mES cells (Figure 4.6.A).

A



B

High-Car2 [%]			High-Sox3 [%]			High-Car2 High-Sox3 [%]		
Low-Nanog (A)	High-Nanog		Low-Nanog (B)	High-Nanog		Low-Nanog Observed	Low-Nanog Predicted (A x B)	
7.91	>	3.72	8.63	>	3.35	2.88	>	0.68

Figure 4.6 – Correlation analysis of priming gene expression in E14 mES cells cultured in "Serum/LIF" conditions. (A) *Correlation scatterplots for Car2, Nanog and Sox3. R is the Spearman correlation coefficient and the numbers on the table correspond to the percentage of cells expressing higher or lower levels of priming gene within the respective Nanog subpopulation: low-Nanog cells on the left and high-Nanog cells on the right.* ***(B)*** *Percentage of cells expressing higher levels of priming genes (Car2 and Sox3) within the low- and high-Nanog subpopulations, suggesting the occurrence of "lineage priming" for Sox3. The last two columns are for low-Nanog cells expressing high levels of both priming genes simultaneously, comparing the observed to the expected results if they were independent events.*

Most of the cells express lower levels of both priming genes (86.33% of low-*Nanog* cells and 93.31% of high-*Nanog* cells), since they were grown in pluripotency conditions where lineage-affiliated genes like *Sox3* are almost absent (Table A2). *Nanog* is weakly anti-correlated to either *Car2* and *Sox3*, as evident by the negative and low Spearman correlations (-0.10 and -0.08, respectively) (Figure 4.6.A).

These results also show that there is a higher percentage of low-*Nanog* cells expressing increased levels of priming genes, when compared to high-*Nanog* cells. This observation is valid for both *Car2* (7.91% > 3.72%) and *Sox3* (8.63% > 3.35%), further confirming that both priming genes are upregulated in a low-*Nanog* state, as pointed initially by the RNA-sequencing data, and suggesting the occurrence of "lineage priming" in the case of *Sox3* (Figure 4.6.B).

Moreover, there is a poor correlation in mRNA expression between *Car2* and *Sox3*, inferred by the very weak Spearman correlation coefficient (R = 0.17), which implies that the expression of one priming gene is not a good indicator for the other. It is thus unlikely that there is a common upstream regulatory pathway to activate priming gene expression in a low-*Nanog* state.

Most low-*Nanog* cells (86.33%) do not express any of these priming genes. Nevertheless, in the low-*Nanog* cells that do, 7.91% only express high levels of *Car2* transcripts, 8.63% only high-*Sox3* and 2.88% express both genes simultaneously. In order to test if the expression of each gene is an independent event, they must follow the mathematical rule (1), in which P(A) is the probability of expressing higher levels of *Car2*, P(B) of *Sox3* and P(A ∩ B) of both, always in low-*Nanog* cells. Indeed, 7.91% x 8.63% = 0.68%, which is lower than the observed 2.88%, concluding that the events are not independent (Figure 4.6.B).

$$P(A)\times P(B) = P(A \cap B) \quad (1)$$

Thus, the expression of *Sox3* and *Car2* within low-*Nanog* cells shows some dependency. It is hypothesized that in the absence of NANOG, TET1 and PRC2 would not act at priming genes promoters, which would lead to the de-repression of priming gene expression in the low-NANOG state. On the other hand, sporadic genes like *Car2*, unrelated to lineage commitment, might be upregulated in low-*Nanog* cells due to neighbouring chromatin effects that allow simultaneous active transcription, by crosstalk or near location.

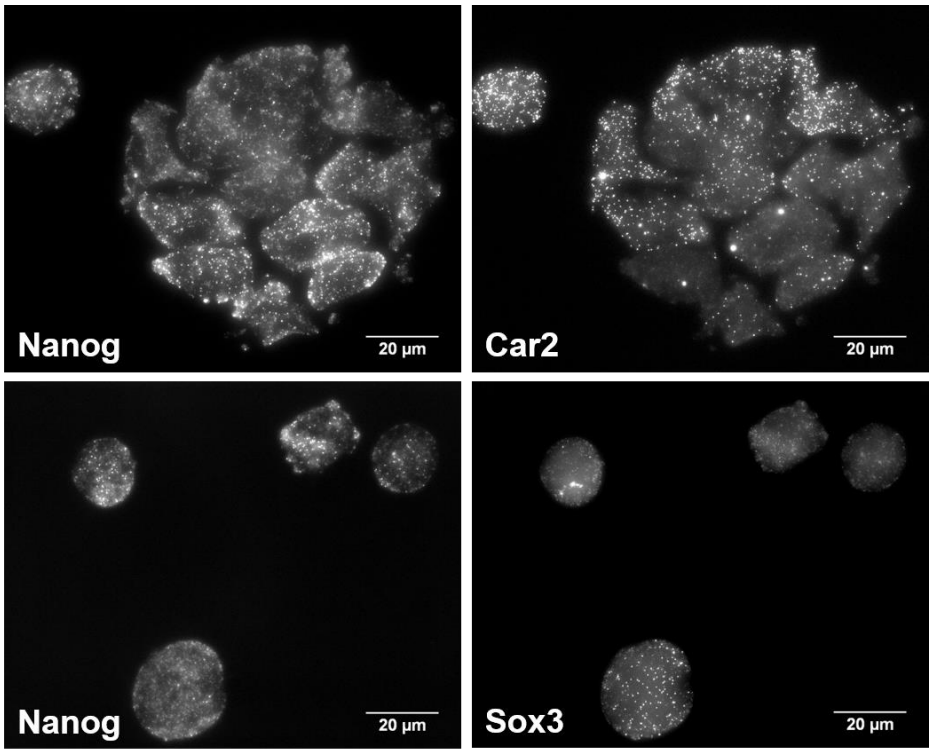
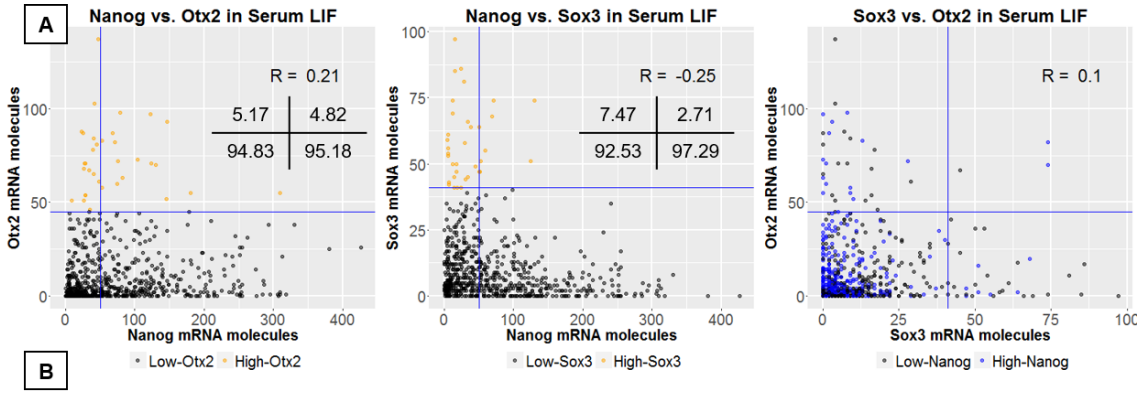


Figure 4.7 – Priming gene expression in E14 mES cells cultured in "Serum/LIF". *Nanog is depicted on the left and priming genes on the right. Car2 is analysed in the first row and Sox3 in the second. There is an upregulation of priming gene expression in low-Nanog cells. Scale bar = 20 µm.*

In summary, this smFISH analysis proved that “lineage priming” occurs in pluripotent mES cells, as it was shown by *Sox3* upregulation in low-*Nanog* cells (Figure 4.7). Nevertheless, “lineage priming” is a very rare event, since it only occurs to 8.63% of low-*Nanog* cells, which correspond to 2.94% of the total population (Table A2). Therefore, some low-*Nanog* cells can explore lineage programmes (such as neural, denoted by *Sox3*) before definitive commitment by expressing, in a reversible manner, higher levels of lineage-affiliated genes.

Nanog-Otx2-Sox3 smFISH Analysis

Additionally, the combination *Nanog-Otx2-Sox3* was analysed in 680 E14 cells cultured in “Serum/LIF”. The aim was to establish if there is an effect of *Nanog* on *Otx2* and if *Otx2* correlates with *Sox3*. The threshold for *Otx2* was 45 mRNAs/cells and for *Sox3* was 41 mRNAs/cell (the same used in the previous analysis). The scatterplots representing the correlations between these three mRNA molecules are depicted in Figure 4.8.A.



B

High-Otx2 [%]		High-Sox3 [%]		High-Otx2 High-Sox3 [%]	
Low-Nanog (A)	High-Nanog	Low-Nanog (B)	High-Nanog	Low-Nanog Observed	Low-Nanog Predicted (A x B)
5.17 ≈	4.82	7.47 >	2.71	0.29 ≈	0.39

Figure 4.8 – Correlation analysis of gene expression in E14 mES cells cultured in “Serum/LIF” conditions. ***(A)*** *Correlation scatterplots for Otx2, Nanog and Sox3. R is the Spearman correlation coefficient and the numbers on the table correspond to the percentage of cells expressing higher or lower levels of Otx2 or Sox3 within the respective Nanog subpopulation: low-Nanog cells on the left and high-Nanog cells on the right.* ***(B)*** *Percentage of cells expressing higher levels of Sox3 and Otx2 within the low- and high-Nanog subpopulations, suggesting the occurrence of “lineage priming” for Sox3. The last two columns are for low-Nanog cells expressing high levels of Sox3 and Otx2 simultaneously, comparing the observed to the expected results if they were independent events.*

As expected, there was a higher percentage of low-*Nanog* cells expressing increased levels of *Sox3* when compared to high-*Nanog* cells (7.47% > 2.71%), which implies the occurrence of “lineage priming” for *Sox3*. Nevertheless, the same does not occur with *Otx2*, since there is a similar number of high-*Otx2* cells in both low-*Nanog* (5.17%) and high-*Nanog* (4.82%) states (Figure 4.8.B). These results elucidate that *Otx2* is not upregulated in low-*Nanog* cells when compared to high-*Nanog* cells (as *Sox3* is) and thus, it cannot be considered a priming gene (which is coincident with the RNA-sequencing data prediction). Therefore, it is hypothesized that *Otx2*, an EpiS cell marker of the “primed” state (definition

in section 1.2.2), expression might not be increased in a low-NANOG “lineage-primed” state (definition in section 1.3.2) in “Serum/LIF” pluripotency conditions, and only become upregulated in later stages of ES cell “way” to commitment or, alternatively, *Otx2* might be regulated by another mechanism.

In low-*Nanog* cells, only 5.17% express high levels of *Otx2*, 7.47% of high-*Sox3* and 0.29% express both genes simultaneously. In case these are independent events, the expected probability of low-*Nanog* expressing high levels of both genes would be 5.17% x 7.47% = 0.39%, which is very similar to the observed 0.29%, supporting the fact that the expression of *Otx2* and *Sox3* in low-*Nanog* cells are completely independent events (Figure 4.8.B), contrarily to what was previously observed for *Car2* and *Sox3*.

As depicted in Figure 4.9, there can be an increased *Otx2* expression almost equally distributed between low-*Nanog* (yellow arrow) and high-*Nanog* cells (blue arrow). This result is different from a protein analysis study, already published, which states that ES cells exhibiting high levels of OTX2 preferentially express low levels of NANOG, whereas those with low levels of OTX2 exhibit a preference for higher NANOG expression (Acampora *et al.* 2013). Therefore, OTX2 was considered anti-correlated with NANOG and a marker of pluripotency exit.

However, protein information might not correspond to what happens at the mRNA stage, *Otx2* has a short mRNA half-life of 2.26 h (Sharova *et al.* 2009), which is shorter than *Nanog*’s mRNA half-life, and OTX2 protein might have a longer half-life. If this is the case, OTX2 protein would still be present while *Otx2* mRNA was already degraded, explaining why there would be OTX2 protein within low-NANOG cells when the *Otx2* mRNA is no longer there. Nevertheless, this type of comparisons should be studied in more detail.

In the next subchapter 4.2, it will be discussed the mechanism of NANOG regulation of priming gene expression, through the use of small molecules, such as GSK343 and/or AA.

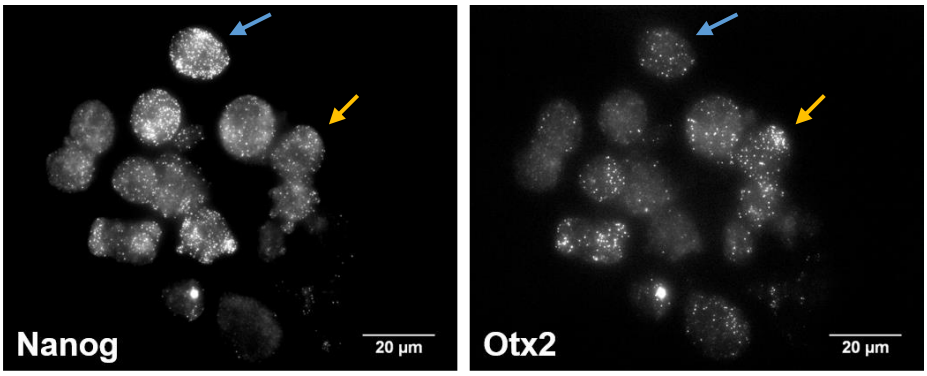


Figure 4.9 – High-Otx2 expression is almost equally distributed between high-Nanog and low-Nanog cells. *Despite the fact that OTX2 is anti-correlated with NANOG in mES cells, at a protein level, this relation is not clear at the mRNA level. This image pretends to illustrate the variety of combinations that can occur. Scale bar = 20 µm.*

4.2 Nanog Regulation of Priming Gene Expression through PRC2 and TET1

Previous work in DHenrique Lab found, by Chromatin Enrichment Analysis (ChEA), that priming genes, genes preferentially upregulated in low-Nanog:VNP cells, are highly enriched for binding signatures in PRC2 components. Thus, it was hypothesized that the repression of priming gene expression in a high-NANOG state might involve NANOG interaction with PRC2. Moreover, it was thought that this regulation might involve TET1 due to its crosstalk with PRC2 (Wu *et al.* 2011) and also because TET1 binds to NANOG (Costa *et al.* 2013).

In order to test this hypothesis mES cells were exposed during 48 hours to GSK343, ascorbic acid (AA), the combination AA+GSK343 and DMSO (the solvent of GSK343, to serve as a control). GSK343 and AA are from now on named chemical modulators due to their modulation of PRC2 and TET1 activities, respectively. The medium was changed in 24 hours with fresh "Serum/LIF" supplemented with new chemical modulators, to ensure its maximum effect. After 48 hours of exposure, cells were fixed, half for smFISH and the other half for IC-FC (Figure 4.10).

GSK343 is a specific inhibitor of EZH2, the PRC2 catalytic subunit, blocking the *de novo* trimethylation of H3K27 whilst AA stimulates TET1 activity. According to preliminary data from DHenrique Lab, GSK343 increased the expression of *Sox3*, a lineage-affiliated gene, in a high-NANOG state. On the other hand, AA decreased *Sox3* expression, also in high-NANOG cells. When GSK343 was added to AA, the effect of AA was reverted. Both GSK343 and AA should only influence those mES cells that have transited from low- to high-NANOG state during 48 hours of exposure to chemical modulators. This chapter will confirm or contradict these results by analysing *Car2* and *Sox3* expression in single-cells by smFISH, to unravel the mechanisms of NANOG regulation of priming gene expression.

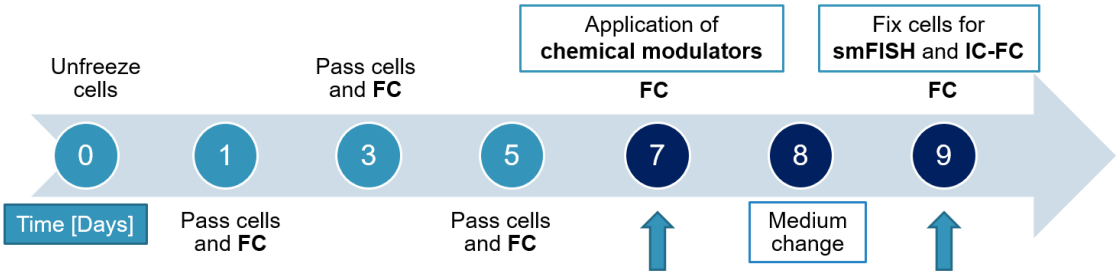


Figure 4.10 – Scheme of the experiment procedure for ES cell culture in "Serum/LIF" conditions supplemented with GSK343 and/or AA. *mES cells were also exposed to DMSO, to serve as a control. After 24h of exposure, the medium was changed for fresh "Serum/LIF" supplemented with new chemical modulators or DMSO, 24h after that mES cells were fixed for smFISH and IC-FC.*

4.2.1 Analysis of cell morphology and dynamics of Nanog:VNP

mES cells were grown in "Serum/LIF" and exposed, for 48 hours, to GSK343 at a concentration of 1 µM (concentration associated with the highest H3K27me3 decrease without causing toxic effects to cells tested in DHenrique lab). Cells incubated with GSK343 showed no major changes on morphology nor evidence of toxicity, when compared to the cells incubated with DMSO (Figure 4.11.A).

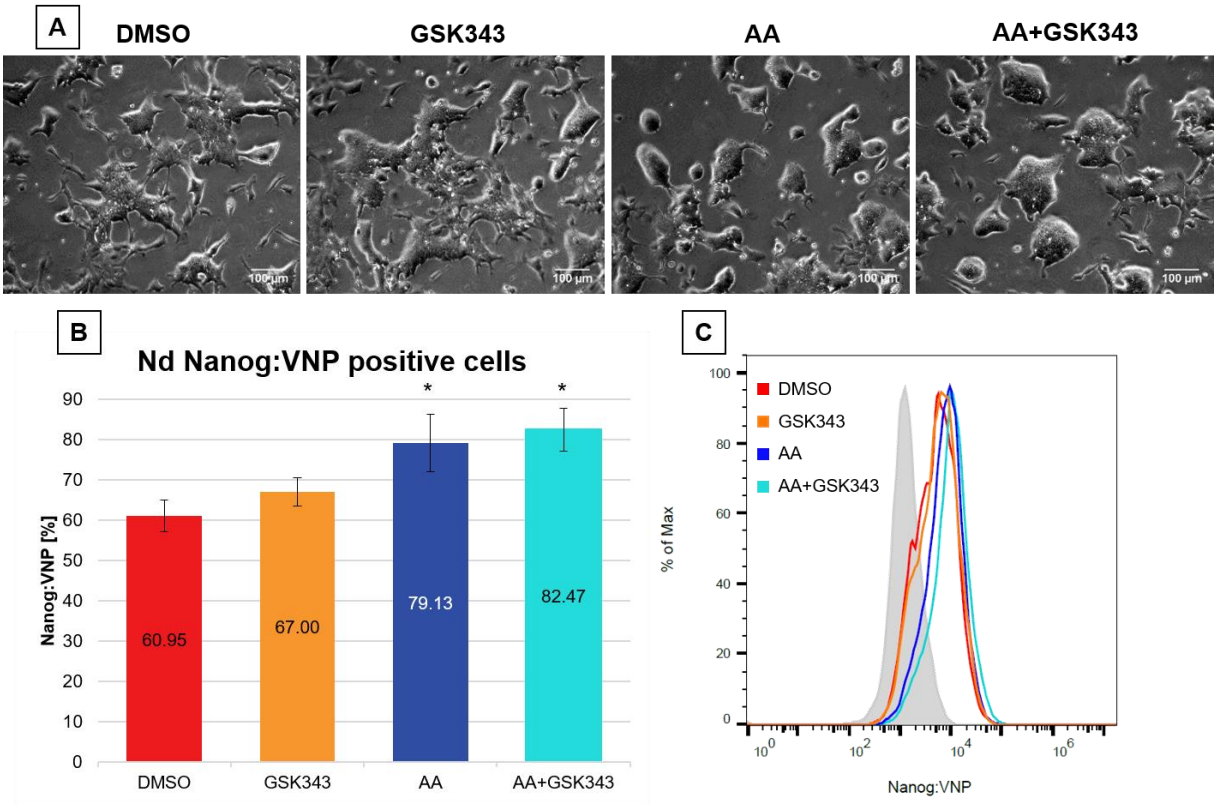


Figure 4.11 – Analysis of the effect of GSK343 and/or AA in mES cells cultured in "Serum/LIF". ***(A)*** *Cell morphology in "Serum/LIF" conditions supplemented with chemical modulators or DMSO. Upon the application of AA there are less flattened differentiated-like cells at the colonies' periphery. Scale bar = 100 µm.* ***(B)*** *Percentage of Nanog:VNP positive cells in Nd mES cells cultured in "Serum/LIF" with GSK343 and/or AA for 48 hours. Mean values were depicted for each condition and error bars were calculated based on standard deviation from three experiments (n=3). Statistically significant difference with p-value ≤ 0.05 (t test) observed between DMSO and AA treated cells is denoted with (*).* ***(C)*** *Representative flow cytometry profile in Nd mES cells, in which there is a significant increase on Nanog:VNP upon AA exposure, visible by the shift to the right on both dark and light blue curves, AA and AA+GSK343, respectively. E14 cells were used as a control (depicted in filled grey).*

The percentage of Nanog:VNP cells was assessed through flow cytometry 48 hours upon GSK343 exposure (Figure 4.11.B and Figure 4.11.C). There were no significant differences between the percentage of Nanog:VNP cells for DMSO and GSK343 conditions (60.95% and 67.00%, respectively). Nevertheless, there is a slight increase in the number of Nanog:VNP positive cells cultured with GSK343, when compared to DMSO. This result is consistent with the observation that *Ezh2*-null iPS cells have increased levels of NANOG (Villasante *et al.* 2011). In these conditions flattened differentiated-like cells could be identified at the periphery of cell clusters and isolated between clusters (Figure 4.11.A).

In order to test the hypothesis regarding TET1 role on priming gene repression, mES cells were incubated for 48 hours with AA at a concentration of 1 mg/mL, previously determined for mES cells (Blaschke *et al.* 2014). Relatively to effects on the cell morphology, clusters were more tightly packed, with more round-shaped cells and a reduction in flattened differentiated-like cells at colonies' periphery (Figure 4.11.A), which is coincident with the significant increase of the percentage of Nanog:VNP

expressing cells upon AA exposure, when compared to DMSO (79.13% in the presence of AA and 60.95% with DMSO, $p < 0.05$) (Figure 4.11.B and Figure 4.11.C).

These results are in accordance with the already proved role of AA in promoting TET1 activity (Yin *et al.* 2013; Blaschke *et al.* 2014) and consequently NANOG expression (Ito *et al.* 2010; Wu *et al.* 2011), as it will be explained in detail later, in the end of section 4.3.1.

When mES cells were grown in AA+GSK343 condition, with 1 µM GSK343 and 1 mg/mL AA, there were observed morphological changes similar to AA condition. Cells exhibited less differentiated-like cells (Figure 4.11.A), corroborated by the marked increased percentage of Nanog:VNP positive cells (82.47% in the presence of AA+GSK343 and 60.95% with DMSO, $p < 0.05$) (Figure 4.11.B and Figure 4.11.C). Once again, there is a slight increase of Nanog:VNP when GSK343 is added to AA, compared to AA-only, probably due to the reasons mentioned above to justify GSK343 results.

4.2.2 Verification of GSK343 Effect

In order to verify if the small molecule inhibitor GSK343 was having the desired effect of reducing H3K27me3 levels, it was performed an intracellular staining - flow cytometry (IC-FC) analysis with an antibody against H3K27me3 (Figure 4.12).

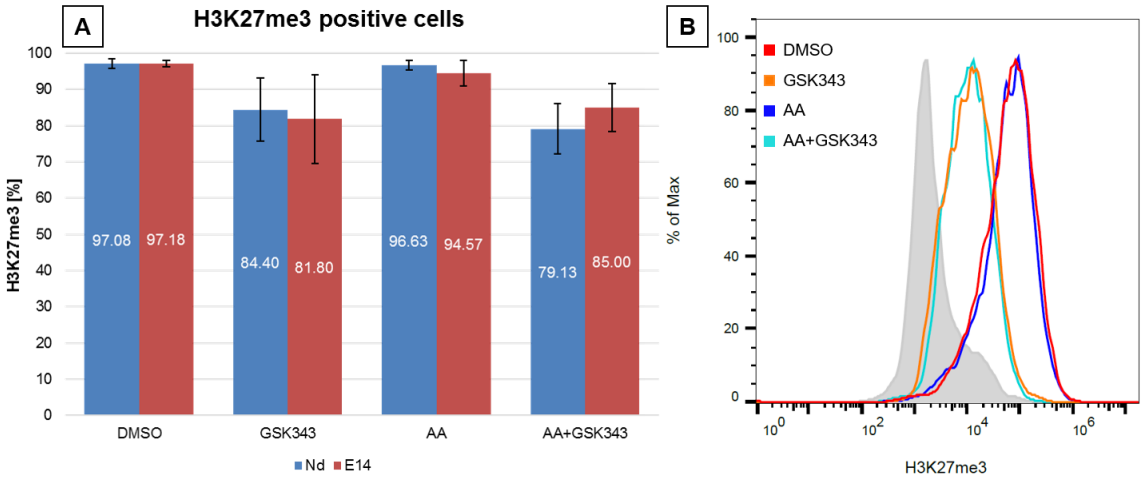


Figure 4.12 – H3K27me3 positive cells in mES cells grown in "Serum/LIF" supplemented with chemical modulators. (A) *Percentage of H3K27me3 positive cells with mean values depicted for each condition (DMSO, GSK343, AA, AA+GSK343), obtained by IC-FC for Nd and E14 mES cells, in blue and red respectively. Error bars were calculated based on standard deviation from three experiments (n=3).* ***(B)*** *Representative flow cytometry profile for Nd mES cells, using samples only marked with secondary antibodies as a negative fluorescent control (depicted in filled grey).*

This analysis showed a slight reduction on the percentage of H3K27me3 positive cells upon the application of GSK343 (alone or combined with AA), when compared to cells cultured with DMSO. Amongst a Nd population of cells, there were 97.08% H3K27me3 positive cells in the presence of DMSO, 84.40% in GSK343 condition and 79.13% in AA+GSK343 condition, whilst in the E14 population the average percentages of H3K27me3 positive cells were 97.18%, 81.80% and 85.00%, respectively. In AA condition, the percentage of H3K27me3 positive cells remained approximately the same

comparatively to DMSO: 96.63% for Nd and 94.57% for E14 cells. Interestingly, the error bars relatively to H3K27me3 were increased whenever GSK343 was added to cells, which suggests that GSK343 efficiency might vary (Figure 4.12.A and Figure 4.12.B).

There is only a slight reduction on the percentage of H3K27me3 expressing cells upon exposure to GSK343 and not an overall significant effect on the population. This might be due to the fact that GSK343 only acts on the inhibition of the *de novo* trimethylation of H3K27 and does not eliminate the methylation mark on histones already methylated, which would have caused a much more significant reduction in the global percentage of H3K27me3 positive cells. Thus, the decrease of H3K27me3 might depend on the normal remodelling of chromatin, which is apparently a slow process within 48 hours.

4.2.3 Analysis of Priming Gene Expression with GSK343 and/or AA

Single-cell analysis is required to understand the effect of NANOG on the regulation of priming gene expression through PRC2 and TET1. The reason underlying this need is the observed heterogeneity in individual mES cells, which makes so crucial to analyse one by one and also because "lineage priming" is an extremely rare event that would be completely disregarded and lost in averages.

Therefore, smFISH was performed in E14 mES cells treated with DMSO, GSK343, AA and AA+GSK343. The aim was to test our previously described model for NANOG repression of priming gene expression by affecting the activity of two participants, PRC2 and TET1, through the application of chemical modulators. In order to achieve this aim, I have looked into the differences in priming gene expression between low- and high-*Nanog* subpopulations, by smFISH, on these four distinct conditions. Two priming genes were selected: *Sox3*, a lineage-affiliated gene involved in neural lineage specification, and *Car2*, a sporadic gene, with no known role in embryonic development. A smFISH analysis with this probe combination, in "Serum/LIF", was already described in section 4.1.2. and revealed that *Car2* and *Sox3* expression are upregulated in low-*Nanog* cells when compared to high-*Nanog* cells.

DMSO

In DMSO condition, *Nanog* has an average expression of 80 transcripts/cell and its distribution between high and low states is 53.27% and 46.73%, respectively (Table A3 from Annex 1). *Car2* has an average expression of 92 transcripts/cell and *Sox3* of 14 transcripts/cell (Table 4.1.A). *Nanog*, *Car2* and *Sox3* present a long tailed distribution, suggesting a bursty transcription for these three genes (Figure 4.13.A). The threshold used for *Nanog* is 50 transcripts/cell and for priming genes corresponds to the top-5% expressing cells (for *Sox3* is 53 transcripts/cell and for *Car2* is 324 transcripts/cell).

Only 3.63% of high-*Nanog* cells express high levels of *Car2* transcripts, which is lower than the 7.32% of low-*Nanog* cells that are high-*Car2* (Table 4.2). The same happens with *Sox3*, with only 3.07% of high-*Nanog* cells expressing high levels of *Sox3* transcripts, compared to 7.64% of low-*Nanog* cells that are high-*Sox3* (Table 4.2). As it was confirmed in section 4.1.2, there is an upregulation of priming genes in low-*Nanog* cells compared to high-*Nanog* cells, suggesting the existence of "lineage priming" in the case of *Sox3*.

GSK343

When mES cells were treated with GSK343, there were no major changes in gene average expression: 106, 76 and 13 transcripts/cell for *Car2*, *Nanog* and *Sox3*, respectively (Table 4.1.B). Moreover, all genes continue to exhibit a long-tailed distribution (Figure 4.13.B).

Nevertheless, there is an increase in the percentage of high-*Nanog* cells expressing high levels of *Sox3* transcripts, when compared to DMSO (6.77% > 3.07%). Simultaneously, there is no difference in the percentage of low-*Nanog* cells that express high levels of *Sox3* (7.38%) when compared to DMSO (7.64%) (Table 4.2). These results are in accordance with the hypothesis that inhibition of EZH2 activity by GSK343 prevents the repression of lineage-affiliated gene promoters in a NANOG-regulated mechanism. This is expected to occur only in high-*Nanog* cells that have transited from a low-*Nanog* state within the 48 hours of GSK343 exposure. Hence, the limited number of high-*Nanog* cells that show increased lineage-affiliated gene expression, and the absence of effects in low-*Nanog* cells.

On the other hand, there is no significant change in *Car2* expression within low- or high-*Nanog* cells exposed to GSK343, when compared to DMSO. There are 4.00% of high-*Nanog* cells and 7.79% of low-*Nanog* cells that express high levels of *Car2* transcripts, which are similar values to the observed in DMSO (3.63% and 7.32%, respectively) (Table 4.2). These results indicate that PRC2 modulation is constrained to lineage-affiliated genes, such as *Sox3*, and is not applicable to sporadic genes, like *Car2*.

AA

In mES cells treated with AA, the first observation is that *Car2* average expression increases from 92 (in DMSO) to 155 transcripts/cell (Table 4.1.C). This is probably due to the AA-mediated activation of TET enzymes, which leads to DNA demethylation on *Car2* promoter, and consequent activation of transcription. However, there is a slight decrease on *Sox3* average expression from 14 (in DMSO) to 10 transcripts/cell, probably due to the higher number of cells with 0 mRNAs of *Sox3*.

Nanog average expression also increases from 80 (in DMSO) to 96 transcripts/cell (Table 4.1.C), correlating with the observed Nanog:VNP protein increase in Nd cells upon AA exposure. *Nanog* distribution changes from a long tailed distribution (in DMSO) to a more bell-shaped curve, vaguely similar to what is observed for *Nanog* in “2i” conditions (Abranches *et al.* 2014) (Figure 4.13.C). Indeed, *Nanog* distribution between high and low states is altered to 74.50% and 25.50%, respectively, which means that there is a higher number of high-*Nanog* cells when compared to DMSO (Table A3). Once again, these results are in agreement with the already proved role of AA in promoting TET1 activity (Yin *et al.* 2013; Blaschke *et al.* 2014) and consequently NANOG expression (Ito *et al.* 2010; Wu *et al.* 2011), as it will be explained in detail later, in the end of section 4.3.1.

Upon AA exposure, there is a slight decrease on the percentage of high-*Nanog* cells that express high levels of *Sox3* transcripts, when compared to DMSO (1.78% < 3.07%) (Table 4.2). Nevertheless, it is necessary to analyse more cells (N = 302, the lowest N in this set of experiments) to confirm these modifications in *Sox3* expression, which are not very pronounced. By contrast, there is no observable effect on the percentage of low-*Nanog* with high expression of *Sox3* (7.79%), in comparison to DMSO (7.64%) (Table 4.2).

The lower *Sox3* expression found in high-*Nanog* cells is in agreement with the proposed model in which AA activates TET1, that when bound to NANOG (Costa *et al.* 2013), will recruit PRC2 through the increased levels of 5hmC in NANOG-TET1 bound promoters. This "attraction" of PRC2 to the hypomethylated lineage-affiliated gene promoters will lead to H3K27 trimethylation and consequently cause the repression of lineage-affiliated gene expression (Wu *et al.* 2011). This hypothesis justifies why there is a decrease of *Sox3* expression in high-*Nanog* cells. This result is expected to only occur at high-*Nanog* cells that have transited from a low-*Nanog* state within 48 hours, which explains why there is no observable effect in low-*Nanog* cells and also the limited number of high-*Nanog* cells that show decreased lineage-affiliated gene expression.

Nevertheless, the same does not apply to *Car2* expression. There is a higher percentage of both low- and high-*Nanog* cells which express high levels of *Car2* transcripts (19.48% and 10.22%, respectively) when compared to DMSO (7.32% and 3.63%) (Table 4.2). This suggests that AA leads to a global increase on *Car2* expression, through DNA demethylation on *Car2* promoter, and does not lead to PRC2 recruitment, since there is no decrease of *Car2* expression. Thus, the 5hmC-mediated recruitment of PRC2 is only applicable to lineage-affiliated genes, such as *Sox3*, and not to sporadic genes like *Car2*.

AA+GSK343

Car2 expression, similar to what was observed in AA-treated cells, exhibits a general increase due to DNA demethylation on *Car2* promoter. *Car2* average expression is 123 mRNAs/cell, which is higher than the control (92 mRNAs/cells) but lower than the value obtained for AA-treated cells (155 mRNAs/cell) (Table 4.1.D). This result suggests that GSK343 might counteract the activity of TET1 as a transcriptional activator of *Car2*. Simultaneously, *Nanog* average expression is also higher (98 mRNAs/cell), when compared to the control (80 mRNAs/cell), similar to what was observed for AA-treated cells (96 mRNAs/cell) (Table 4.1.D).

In mES cells exposed to AA+GSK343, there is a slight increase on the percentage of high-*Nanog* cells expressing high levels of *Sox3* transcripts, when compared to DMSO (5.13% > 3.07%) (Table 4.2). Therefore, the effect of AA is completely reverted and PRC2 inhibition by GSK343 prevails, since the lineage-affiliated gene expression in high-*Nanog* cells is higher than the one of DMSO and similar to GSK343-treated mES cells (6.77%) (Table 4.2).

Simultaneously, there is no significant effect in low-*Nanog* cells with high levels of *Sox3* (7.20%), when compared to DMSO (7.64%) (Table 4.2). These observations are in agreement with the hypothesis that PRC2 acts "downstream" of TET1 (Wu *et al.* 2011).

Relatively to *Car2* expression, the effect of AA seems to be diminished due to GSK343 action, when compared to AA-treated cells: 11.20% < 19.48% for low-*Nanog* cells and 9.62% < 10.22% for high-*Nanog* cells expressing high levels of *Car2* transcripts (Table 4.2). Thus, TET1-mediated DNA demethylation on *Car2* promoter might be influenced by PRC2 action.

The statistical analysis for different conditions is summarized in Table 4.1 and mRNA distributions are represented in the form of histograms for each gene and condition in Figure 4.13.

***Table 4.1 – Statistical analysis of Car2, Nanog and Sox3 mRNA expression**. Mean, standard deviation (SD), median, minimum (Min), maximum (Max), Fano factor (FF), coefficient of variation (CV) and number of cells (N) are depicted for each condition and gene.*

	Condition	Gene	Mean	SD	Median	Min	Max	FF	CV	N
A	DMSO	Car2	91.77	102.08	56.00	0.00	827.00	113.54	1.11	672
		Nanog	80.11	76.68	52.00	0.00	470.00	73.39	0.96	672
		Sox3	14.29	17.79	8.00	0.00	123.00	22.13	1.24	672
B	GSK343	Car2	105.66	118.46	61.00	0.00	770.00	132.82	1.12	569
		Nanog	75.57	63.51	58.00	0.00	446.00	53.38	0.84	569
		Sox3	12.62	18.07	5.00	0.00	127.00	25.87	1.43	569
C	AA	Car2	155.31	151.55	96.50	3.00	975.00	147.88	0.98	302
		Nanog	95.56	62.28	86.00	0.00	413.00	40.59	0.65	302
		Sox3	9.91	18.36	3.00	0.00	155.00	34.02	1.85	302
D	AA+GSK343	Car2	122.90	131.11	72.00	1.00	638.00	139.86	1.07	437
		Nanog	98.26	79.08	72.00	1.00	440.00	63.65	0.80	437
		Sox3	14.15	16.90	8.00	0.00	285.64	20.19	1.19	437

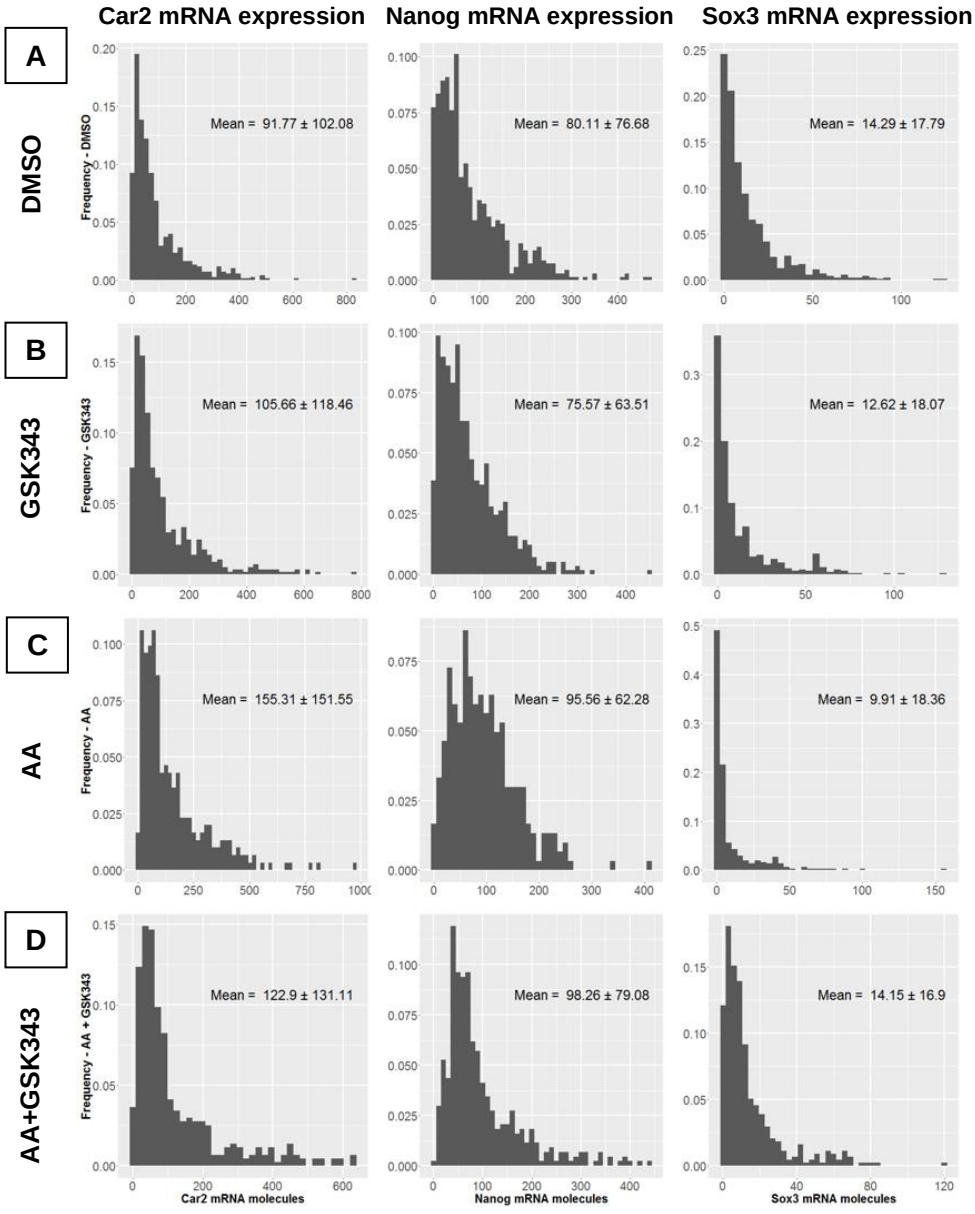


***Figure 4.13 – Histograms of the distribution of mRNA transcripts for Car2, Nanog and Sox3 for different conditions.** Mean ± standard deviation is shown for each gene and condition.*

In Table 4.2 is summarized the percentage of low- and high-*Nanog* cells that express high levels of *Car2* and *Sox3* transcripts. A more complete analysis is detailed in Table A3 and the correspondent scatterplots are in Figure A4, both from Annex A1.

Table 4.2 – Percentages of low- and high-Nanog cells that express high levels of Car2 (on the left) and Sox3 (on the right) transcripts for the different conditions. *The exposure of GSK343 increases the percentage of high-Nanog cells with high levels of Sox3 transcripts, while AA treatment decreases it. mES cells treated with AA+GSK343 exhibit an increased percentage of high-Nanog cells with high levels of Sox3 transcripts, almost the same as GSK343-treated cells. The same patterns are not observed for Car2 expression.*

	High-Car2 [%]		High-Sox3 [%]	
Condition	**Low-Nanog**	**High-Nanog**	**Low-Nanog**	**High-Nanog**
DMSO	7.32	3.63	7.64	3.07
GSK343	7.79	4.00	7.38	6.77
AA	19.48	10.22	7.79	1.78
AA+GSK343	11.20	9.62	7.20	5.13

The next subchapter 4.3 will cover two FACS-sortings that were performed upon bulk Nd population to separate into high- and low-NANOG subpopulations. The aim was to test if *de novo* trimethylation of H3K27 in priming gene promoters occurs from low- to high-NANOG state by exposing FACS-sorted cells to GSK343 and/or AA.

4.3 Sorting into Low and High-Nanog:VNP Cells

To test our hypothesis that the effects of GSK343 and/or AA occur only in mES cells that have transited from low- to high-NANOG state, mES cells subpopulations expressing high- or low-Nanog:VNP were isolated and then exposed to the chemical modulators, in parallel to the control DMSO, for the same 48 hours. To purify the two mES cell's subpopulations, FACS-sorting was performed on the Nd bulk population and low-Nanog:VNP (VNP_L) and high-Nanog:VNP (VNP_H) were isolated (Figure 4.14).

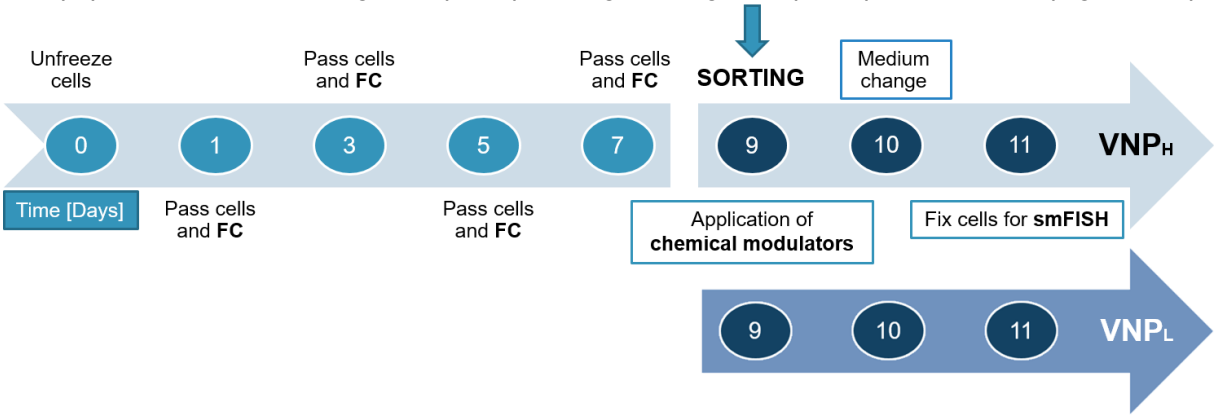


***Figure 4.14 – Scheme of the experiment procedure for the sorting of Nd mES cells into low-Nanog:VNP (VNP_L) and high-Nanog:VNP (VNP_H) subpopulations**. Shortly after the sorting, mES cells were exposed to the chemical modulators GSK343 and/or AA, in parallel to DMSO, for 48 hours and afterwards fixed for smFISH.*

4.3.1 Analysis of cell morphology and dynamics of Nanog:VNP

The sorted sub-populations were placed in culture ("Serum/LIF" conditions) and the morphology was analysed daily on an inverted bright field microscope (Figure 4.15). VNP_H cells, cultured in "Serum/LIF", resemble to ES cells cultured in "2i" conditions, with a notorious round-shaped format in tightly packed clusters, in comparison to VNP_L cells, which present more flattened differentiated-like cells at clusters' periphery or isolated between clusters. VNP_L cells exposed to AA exhibited less differentiated-like cells in comparison to the other VNP_L cells for different conditions (Figure 4.15).

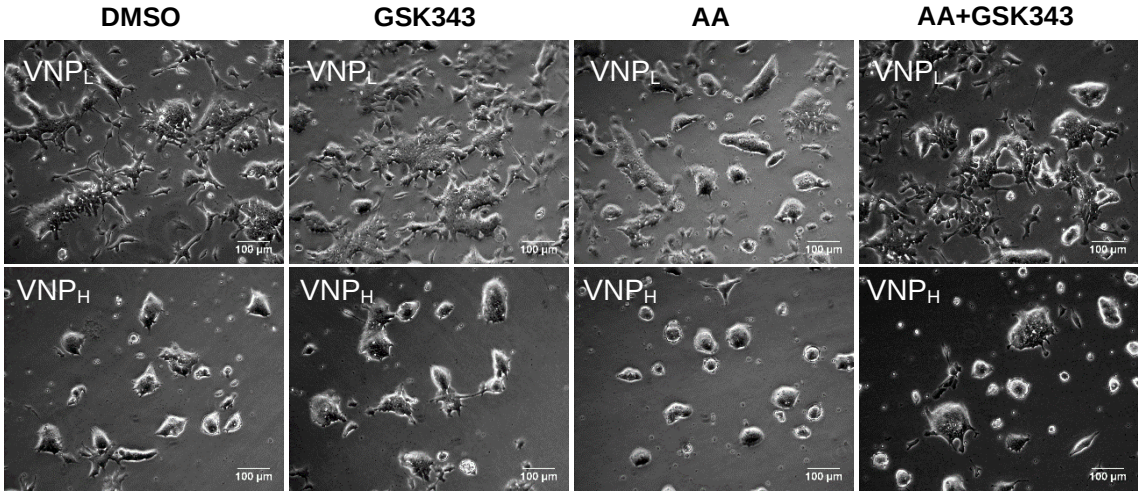
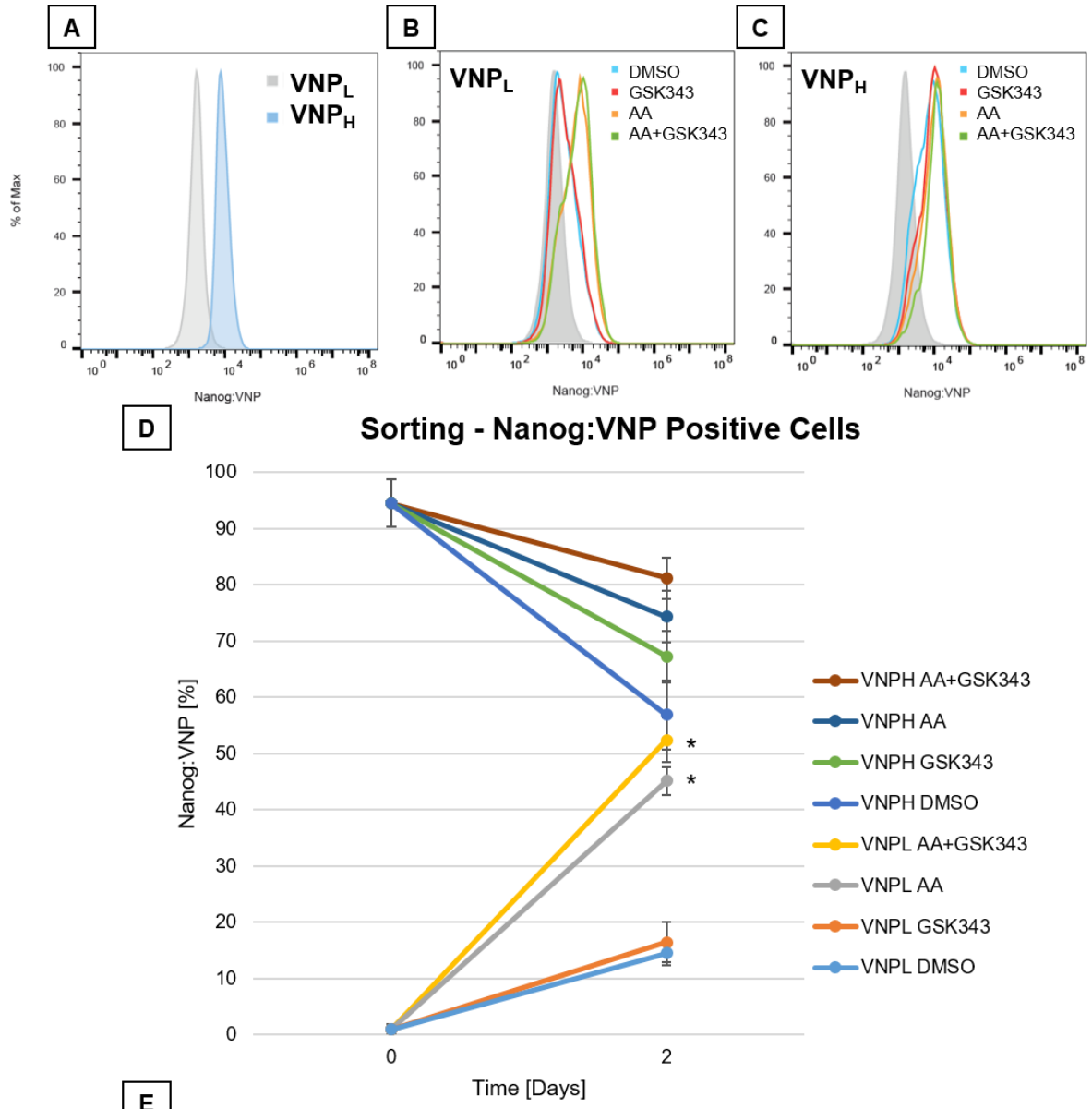


***Figure 4.15 – Morphology of the sorted cells with chemical modulators**. Bright field images of VNP_L cells are shown on the first line and VNP_H cells on the second. Images were taken 48 hours after exposure to chemical modulators or DMSO and before fixation for smFISH. Scale bar = 100 µm.*



E

	D0	D2			
		DMSO	GSK343	AA	AA+GSK343
VNP$_H$	94.5	56.85	67.25	74.35	81.20
VNP$_L$	0.93	14.45	16.45	45.15	52.40

Figure 4.16 – NANOG expression in FACS-sorted Nd mES cells. (A) *Representative histogram of FACS-sorted Nd VNP$_L$ and VNP$_H$ subpopulations immediately after sorting.* ***(B)*** *Representative flow cytometry profile of VNP$_L$ cells exposed to GSK343 and/or AA, in parallel to DMSO, for 48 hours. E14 cells were used as a control (depicted in filled grey).* ***(C)*** *VNP$_H$ cells.* ***(D)*** *Temporal evolution of the percentage of Nanog:VNP positive cells after replating FACS-sorted subpopulations in "Serum/LIF" supplemented with chemical modulators or DMSO during 48 hours. Error bars were calculated based on standard deviation from two experiments (n=2). Statistically significant difference with p-value ≤ 0.05 (t test) observed between DMSO and respective VNP$_L$ treated cells is denoted with (*).* ***(E)*** *Summary of the percentage of Nanog:VNP positive cells for the different conditions immediately after sorting (day 0 – D0) and 48 hours after exposure to chemical modulators (day 2 – D2) for VNP$_L$ and VNP$_H$ subpopulations.*

FACS-sorting efficiency was verified by analysing the percentage of Nanog:VNP on both VNP_L and VNP_H purified subpopulations: 0.93% and 94.50% (mean values from both sortings), respectively (Figure 4.16.A and Figure 4.16.E.). The results confirmed that highly purified subpopulations were obtained. Interestingly, there was always a final lower number of VNP_L sorted cells when compared to VNP_H cells, probably associated to increased cell death. The number of sorted cells, from an initial bulk population of approximately 10^8 Nd cells, using the trypan blue dye exclusion method for cell viability, was 7.11 x 10^6 for VNP_H and 4.49 x 10^6 VNP_L cells (mean values). This observation reveals that sorting is an inefficient method for the separation, with an increased cell death associated.

Despite the fact that the expression of lineage-affiliated genes (*Fgf5*, *Gata6* and *T-brachyury*) is upregulated in VNP_L cells, which implies that they are more predisposed to differentiation, the VNP_L subset still exhibits high levels of *Oct4* and *Sox2* expression (Abranches *et al.* 2013). Thus, VNP_L cells seem to remain in a pluripotent state and, once again, it is confirmed that they have the capacity to revert to a state of high-NANOG expression by the re-establishment of the typical (56.2 ± 8.0)% heterogeneity in self-renewal conditions (Chambers *et al.* 2007; Abranches *et al.* 2013).

Nanog:VNP results obtained for DMSO after two days of exposure were 14.45% for VNP_L and 56.85% for VNP_H (Figure 4.16.D and Figure 4.16.E), which are similar to those previously reported for "Serum/LIF" conditions (Abranches *et al.* 2013).

A previous study showed that VNP_L cells have a slower restoration of NANOG heterogeneity when compared to VNP_H cells, as VNP_L cells take approximately 4 days, whilst VNP_H cells only take 2 days, to reach approximately 50% of Nanog:VNP expression (Abranches *et al.* 2013). The same was verified in these two sortings, with the rate of conversion of Nanog:VNP expression being faster for VNP_H cells when compared to VNP_L cells in DMSO condition (Figure 4.16.D and Figure 4.16.E).

VNP_L cells treated with DMSO for 48 hours showed an increase of the Nanog:VNP percentage from 0.93% to 14.45% and with GSK343 to 16.45% (Figure 4.16.D and Figure 4.16.E). Since Nanog:VNP values are similar between DMSO and GSK343 (there is just a slight increase on GSK343 condition), as observed by the overlap in the flow cytometry profile (Figure 4.16.B), it is suggested that GSK343 might not influence the number of Nanog:VNP positive cells in VNP_L subpopulation.

Relatively to VNP_L cells cultured in presence of AA for 48 hours, the percentage of Nanog:VNP positive cells increased significantly from 0.93% to 45.15%. While in VNP_L cells cultured with AA and GSK343 increased to 52.40% (Figure 4.16.D and Figure 4.16.E). Thus, VNP_L cells exposed to AA (with or without GSK343) notably increase the conversion from low- to high-NANOG state, as observed by the shift to the right in the flow cytometry profile (Figure 4.16.B). This might correlate with the fact that there are less differentiated-like cells at clusters' periphery, which were probably low-NANOG (Figure 4.15). GSK343 addition to AA seems to not have a major influence on the number of Nanog:VNP expressing cells in VNP_L subpopulation (just a minor increase is observed when compared to AA).

Relatively to VNP_H cells, the exposure to AA sustained Nanog:VNP at higher percentages, inhibiting the conversion of cells from high- to low-NANOG state. Interestingly, contrarily to what was observed in VNP_L cells, GSK343 exhibited an effect on maintaining a high number of Nanog:VNP positive cells and it was accumulative with AA. Therefore, by increasing order of Nanog:VNP percentages, it was obtained 56.85%, 67.25%, 74.35% and 81.20% for DMSO, GSK343, AA and

AA+GSK343 treated cells, respectively (Figure 4.16.D and Figure 4.16.E). There is an overlap of curves in the flow cytometry profile for VNP_H cells because the differences between conditions are not very significant (Figure 4.16.C).

It is relevant to note that GSK343 has a more pronounced effect on maintaining high percentages of Nanog:VNP expressing cells in VNP_H subpopulation while it almost does not affect Nanog:VNP in VNP_L cells. Therefore, GSK343 mostly influences the conversion from high- to low-NANOG but not the contrary. Once again, this result is consistent with the observation that *Ezh2*-null iPS cells present an increased NANOG expression in comparison to control iPS cells. The researchers concluded that the higher NANOG expression was due to an expansion of high-NANOG subpopulation (Villasante *et al.* 2011), which is confirmed, here, by the inhibition of transition from high- to low-NANOG state caused by GSK343. Similar to what was observed in the previous study with *Ezh2*-null iPS cells, GSK343 "locks" VNP_H cells in a high-NANOG state of pristine pluripotency.

But why does GSK343 have a major effect in VNP_H rather than in VNP_L cells? It would be expected to also have a notorious effect in VNP_L cells since the levels of EZH2 and H3K27me3 present at the *Nanog* promoter are higher in low-NANOG cells when compared to high-NANOG cells (Villasante *et al.* 2011). A hypothesis to justify this result lies on a recent study which states that the addition of GSK343 to mES cell culture results in TET1 and 5hmC upregulation, mediated by a decrease of EZH2 binding and consequent reduction of H3K27 trimethylation on *Tet1* promoter, which prevents its repression. Moreover, the same researchers found that the *Tet1* gene is promoted by the ES cell pluripotency factors OCT3/4, NANOG and MYC (Neri *et al.* 2015).

Thus, it was hypothesized that, in a high-NANOG state, represented by VNP_H cells, there is more NANOG and, by GSK343 exposure, TET1 is promoted, leading to a positive feedback loop that would significantly increase NANOG expression (even more) and consequently TET1 (Neri *et al.* 2015). The reason why the positive feedback loop would be remarkably activated would be due to the increased NANOG expression in VNP_H cells, that would surpass a threshold, activating TET1 and consequently inducing more NANOG expression. By contrast, in VNP_L cells, this NANOG threshold would not be reached and the positive feedback loop would not be so accentuated, which would justify the observed results with GSK343.

By contrast, AA has a notorious effect in both VNP_L and VNP_H subpopulations, promoting the conversion from low- to high-NANOG in VNP_L cells and inhibiting the conversion from high- to low-NANOG in VNP_H cells. Once again, these results confirm the role of AA in the increase of NANOG expression. There is already a study about this direct correlation between AA and NANOG through the JAK/STAT pathway (Wu *et al.* 2014).

Nevertheless, the most probable and simplified hypothesis involves TET1: firstly, it is known that AA leads to a fast and global increase in 5hmC and this DNA demethylation is mediated by TET1 (Yin *et al.* 2013; Blaschke *et al.* 2014); secondly, TET1 promotes the transcriptional activation of the *Nanog* promoter by maintaining a hypomethylated state, proved at least by two independent studies with *Tet1* knockdown mES cells, by lentiviral shRNAs (Ito *et al.* 2010; Wu *et al.* 2011). Therefore, the inherent conclusion is that AA increases NANOG expression, as it was observed.

It is also worth mentioning that NANOG recruits TET1 to common genomic *loci* associated to pluripotency maintenance and that their physical association has a synergistic effect on improving the efficiency of iPS cell generation (Costa *et al.* 2013). Thus, it is rational to observe a remarkable higher number of high-NANOG cells, at expense of low-NANOG cells, whenever AA is added to either VNP_L or VNP_H cells, when compared to DMSO-treated cells.

4.3.2 Expected smFISH Results from VNP_L and VNP_H Cells

It is expected that the analysis by smFISH of VNP_L cells treated with GSK343 and/or AA will "highlight" the effects from low- to high-NANOG state, in which *de novo* trimethylation occurs. Thus, upon GSK343 exposure, we predict to have a higher *Sox3*, or other lineage-affiliated gene, expression in high-*Nanog* cells. By contrast, the addition of AA will lead to a decrease of *Sox3* expression in high-*Nanog* cells. Finally, the combination AA+GSK343 will increase *Sox3* expression in high-*Nanog* cells up to GSK343 levels or have similar effects to DMSO (Table 4.3).

It is also predicted that *Car2* expression would not be altered according to the *Sox3* pattern in VNP_L nor VNP_H subpopulations, upon GSK343 and/or AA exposure, because the proposed model of NANOG-TET1-PRC2 regulation does not seem to apply to sporadic genes.

Table 4.3 – Expected results after smFISH analysis of FACS-sorted VNP_L cells, after 48 hours of exposure to GSK343 and/or AA. *VNP_H cells will serve as a control and it is not expected to observe significant differences in Car2 expression caused by the proposed model of NANOG-TET1-PRC2 regulation. Low-Nanog cells, within VNP_L subpopulation, should also exhibit similar values to the control DMSO. Arrows are relatively to DMSO-treated cells.*

Low-NANOG → High-NANOG in VNP_L cells	**High-*Sox3* expression in high-*Nanog* cells**
GSK343	↑↑↑
AA	↓↓↓
AA+GSK343	↑ or =

5. Conclusions

Nanog Characterization and Lineage Priming

The heterogeneity in NANOG expression, observed by flow cytometry analysis of Nanog:VNP expression in Nd cells, is a hallmark of mES cells cultured in pluripotency conditions such as "Serum/LIF", and its cause has been linked to the occurrence of NANOG dynamic fluctuations in individual mES cells. Furthermore, NANOG heterogeneity is accompanied by an underlying variability at the *Nanog* mRNA distribution, revealing a bursty transcription through single-cell analysis by smFISH.

Priming genes comprise both lineage-affiliated genes, such as *Sox3* (a marker of neural commitment) and sporadic genes, like *Car2* (no known role in development). Priming gene expression is highly variable, reveals a bursty transcription and is upregulated in low-*Nanog* cells when compared to high-*Nanog* cells, in an uncoordinated manner.

It was shown that the expression of lineage-affiliated genes, such as *Sox3*, is increased in low-*Nanog* cells, which suggests the occurrence of "lineage priming" in pluripotent mES cells. Nevertheless, this "seeding of a particular fate on the way to commitment" (Martinez Arias *et al.* 2013) is a very rare event, since it only occurs to a small fraction of low-*Nanog* cells (around 9%), which corresponds to approximately 3% of the total mES cell population. In sum, *Nanog* stochastic fluctuations enable the existence of a transient low-*Nanog* state, which creates windows of opportunity for mES cells within pluripotency to explore lineage programmes, by expressing higher levels of lineage-affiliated genes, before definitive commitment. However, "lineage priming" is a reversible process, which allows low-*Nanog* cells to re-express *Nanog* and to maintain a pool of pristine pluripotent cells through self-renewal.

NANOG Regulation of Priming Gene Expression

In previous work from DHenrique Lab it was shown that priming genes are enriched for binding signatures in PRC2 components. Recently, it was found that TET1 might crosstalk with PRC2 in the regulation of shared genomic *loci* (Wu *et al.* 2011). This evidence led to the hypothesis that repression of priming gene expression in a high-NANOG state might involve PRC2 and TET1. Hence, GSK343 and AA, small molecules that interfere with PRC2 and TET1 activities respectively, were used to infer about NANOG regulation of priming gene expression. Their effect upon priming gene expression, in correlation with *Nanog*, was analysed quantitatively at the single-cell level by smFISH.

When mES cells were exposed to GSK343, for 48 hours, there was an increase of the expression of *Sox3*, a lineage-affiliated gene, in high-*Nanog* cells. This observation might indicate that GSK343, by blocking the EZH2-mediated *de novo* trimethylation of H3K27 which is predicted to occur from low- to high-*Nanog* state, prevented the repression of lineage-affiliated gene expression. By contrast, in mES cells treated with AA there was a decrease of *Sox3* expression in high-*Nanog* cells. This result suggests that AA, through the promotion of TET1-mediated DNA demethylation, "attracted" PRC2 to the hypomethylated promoters, which resulted in the repression of lineage-affiliated gene expression. When GSK343 and AA were applied simultaneously, there was a reversion of AA effect and thus, it was suggested that PRC2 might act "downstream" of TET1. The expression of *Car2*, a sporadic

gene with no known role in development, did not follow the same patterns of *Sox3* expression upon GSK343 and/or AA exposure.

In line with these results and based on the initial hypothesis, a model is proposed: NANOG recruits TET1 and both form a complex (Costa *et al.* 2013), which will increase 5hmC levels in lineage-affiliated promoters. In turn, PRC2 will be "attracted" to these hypomethylated promoters (Wu *et al.* 2011), leading to H3K27 trimethylation and consequent repression of lineage-affiliated gene expression in high-*Nanog* cells, which transited from a low-*Nanog* state. In the absence of NANOG, there is no formed complex with TET1, PRC2 is no longer recruited and, consequently, there is no H3K27me3 deposition, which prevents the repression of lineage-affiliated gene expression. This would justify why "lineage priming" occurs, the upregulation of lineage-affiliated gene expression in low-NANOG cells.

PRC2 seems to have no evident effect on *Car2* expression and AA causes its upregulation through TET1-mediated DNA demethylation. Thus, it is suggested that the proposed model for NANOG regulation is only applicable to lineage-affiliated genes, such as *Sox3*, being specific for "lineage priming", and not to sporadic genes. The observed *Car2* upregulation in low-*Nanog* cells might be due to chromatin neighbouring effects or regulation by another mechanism.

Besides, there were no effects on *Sox3* expression in low-*Nanog* cells because the *de novo* trimethylation of H3K27 is expected to occur from low- to high-NANOG state, although this will be confirmed by smFISH of FACS-sorted subpopulations.

Furthermore, the results for GSK343 and AA-treated cells were coincident with preliminary data obtained in the DHenrique Lab. However, the combination AA+GSK343 in the preliminary data showed similar levels of *Sox3* expression to DMSO, whilst in this analysis, the obtained values were closer to GSK343 results. In order to clarify this result, it is necessary to analyse more bulk cells and also the sorted VNP_L subpopulation treated with AA+GSK343.

This crosstalk between NANOG, TET1 and PRC2 allows lineage-affiliated genes to be in a "poised" state, ready to activate transcription as soon as cells transit to a low-NANOG state and undergo "lineage priming". According to our hypothesis, NANOG is the "master regulator" that imposes the order in the chaos of stochastic gene expression that characterizes pluripotency, by regulating the expression of lineage-affiliated genes. NANOG's absence predisposes cells to an increase of lineage-affiliated gene expression in the "lineage-primed" state by the ablation of NANOG-TET1 complex and consequent PRC2 dismissal. However, the beauty of this regulation lies beneath the fact that this is not a deterministic or irreversible process. Cells might revert their "decision" by re-expressing NANOG, maintaining a pool of pristine pluripotent cells and they might have the opportunity to exploit another cell-fate programme later, when they undergo "lineage priming" again. On the other hand, cells in a "lineage-primed" state might "choose" to proceed to an irreversible lineage commitment. But what makes a cell "decide" its fate specification or what makes it re-express NANOG? How many times does "lineage priming" have to occur before definitive commitment? Does this mechanism work for all lineage programmes? Indeed, the chaos is still an order to decipher.

Sorting into low and high-Nanog:VNP cells

Two sortings were successfully performed in bulk Nd cells, from which pure subpopulations of low-Nanog:VNP (VNP_L) and high-Nanog:VNP cells (VNP_H) were obtained. VNP_L cells exhibit a slower restoration of Nanog:VNP expression, when compared to VNP_H cells, in DMSO condition. Relatively to VNP_L cells, there is no significant change on the number of Nanog:VNP expressing cells upon GSK343 exposure, just a slight increase when compared to DMSO. Nevertheless, when AA or AA+GSK343 are added to VNP_L cells, there is a drastic increase of Nanog:VNP percentage of positive cells, by an efficient conversion from low- to high-NANOG state.

On the other hand, relatively to VNP_H cells, AA sustains Nanog:VNP expressing cells at a high-state, by preventing the transition between high- to low-NANOG. The most probable reason is composed of two arguments: firstly, AA leads to a fast and global increase in 5hmC mediated by TET1 (Yin *et al.* 2013; Blaschke *et al.* 2014); secondly, TET1 promotes the transcriptional activation of the *Nanog* promoter by maintaining a hypomethylated state (Ito *et al.* 2010; Wu *et al.* 2011). Therefore, AA increases NANOG expression.

Interestingly, in VNP_H cells, GSK343 seems to have a similar effect compared to AA and when both are combined there is an additive effect upon Nanog:VNP percentage of positive cells. Thus, GSK343 also inhibits the transition from high- to low-NANOG in VNP_H cells. This result is in line with the observation that *Ezh2*-null iPS cells exhibit higher levels of NANOG (Villasante *et al.* 2011) and also that GSK343 increases TET1 expression, leading to a consequent increase on NANOG (Neri *et al.* 2015).

It was not possible to achieve the last aim of this project, the smFISH analysis of FACS-sorted cells exposed to GSK343 and/or AA, due to time constraints and technical difficulties. Nevertheless, it is the next step and it will strengthen our model if the observed results and predictions are coincident.

Limitations of the Project

Even though acquisitions were done within 24 hours after sample mounting, sometimes the Cy5 signal photobleached, probably due to the inefficient enzymatic activity of catalase or glucose oxidase. Furthermore, it was detected more signal in the TMR channel than the expected. This “extra-signal” was identified as Alexa’s being incorrectly transmitted through the TMR filter (“bleedthrough”) and as auto-fluorescent components from mES cells. This observation led to the purchase of a new TMR filter, which was tested and is now being used.

The Cy5 signal photobleaching, the “extra-signal” in TMR and the existence of “hot pixels” negatively influenced the smFISH threshold analysis. In the case of Cy5 channel, the signal was weaker, thus difficult to detect, leading to an underestimation of the correct threshold value. On the other hand, the existence of “extra-signal” in the TMR channel made the identification of the correct threshold more difficult, resulting in an overestimation of the correct value. “Hot pixels” are high-intensity pixels which “masked” the true mRNA signal, making the identification of the correct threshold difficult.

When there was no clear threshold, the cell was identified as a “bad object”, being rejected from further statistical analysis. In other cases, where the threshold was slightly questionable, the method used to overcome these limitations was to find a good object (a cell with clear signal and defined threshold) within a certain position, detect the correct threshold range and apply a threshold within this

range to the other objects with a questionable threshold. Therefore, there was an undesirable subjectivity inherent to the analysis, dependent on the software's user, to identify the correct threshold for each cell in questionable cases. This doubtful classification of the thresholds can lead to unreliable and misleading data. Thus, it is extremely important to analyse many cells and to continue to optimize the smFISH signal acquisition in the new Zeiss Axio Observer system.

Moreover, it is difficult to define a threshold to distinguish between high- and low-gene states, as a reflection of the transcriptional activity. During this project, the thresholds used were inferred by visual observation of the histogram representation of the gene mRNA distribution. For instance, the *Nanog* threshold dictates the permissiveness for "lineage priming" versus pristine pluripotency.

As mentioned before, "lineage priming" is a very rare event, thus, it is necessary to analyse a large number of cells to obtain reliable data, which is not always possible due to the image acquisition limitations already described and time constraints. Besides, the number of cells within a position is highly variable and it can only reach a maximum of 20 cells. Simultaneously, within an experiment, a maximum of 130 positions can be taken, which is also highly dependent on the number of cells.

Another limitation of this project was the use of chemical modulators that are not specific for NANOG nor priming gene expression. The effects of GSK343 and AA are global to the modulation of the chromatin environment, through histone methylation (since GSK343 blocks the *de novo* trimethylation of H3K27) and TET1-promoted DNA demethylation (by AA).

It is very difficult to capture the dynamic mechanisms of ES cells, namely the reversibility of "lineage priming". In an ES cell population, even if individual cells change their *Nanog* expression, the population reaches an equilibrium over time. We capture this equilibrium when we perform static experiments, such as smFISH, to address gene expression. It is like we are taking multiple "snapshots" so that we could watch the whole film, and sometimes, if a scene is missing, the story might be interpreted differently.

Strengths of the Project

The mES cell heterogeneity requires a single-cell analysis, which is made possible by smFISH, at mRNA level. It is essential to study individual cells rather than population averages if we seek for a proper understanding of stem cell biology. Besides, "lineage priming" is a very rare event, thus, it is only possible to detect when considering many cells. Furthermore, NANOG is known to interact with PRC2 and TET1, whose activities were influenced by GSK343 and AA, small molecules with reversible effects, preferable for this type of analysis. Although less potent, if we have used deletion mutants it would have generated unwanted secondary effects, namely on cell's viability.

Nowadays it is necessary to quantify gene and protein expression, making use of statistical analysis to fully comprehend biological mechanisms. Thus, a quantitative single-cell method like smFISH should be routinely used and it is, so far, the best available tool to address the question that we proposed to answer. If this model proves to be true, NANOG regulation of lineage-affiliated gene expression will be unravelled. This will allow an improved comprehension of the mechanisms of "lineage priming" in mES cells, possibly useful for the optimization of cell differentiation, essential for regenerative medicine amongst other stem cell's applications.

6. Future Work

It is necessary to analyse the effect of GSK343 and/or AA upon NANOG regulation with more priming genes (besides *Sox3* and *Car2*), including lineage-affiliated genes (such as *T-brachyury*, a marker for mesoderm lineage) and sporadic genes (such as *Cld6*). This would give more confidence to our model, and would help to understand if this is a general mechanism for lineage-affiliated genes from different lineage programmes. It would be also interesting to confirm the smFISH results by using a TET1 inhibitor (we would expect the same result as GSK343) or a PRC2 agonist (same result as AA).

It is essential to analyse the sorted VNP_L and VNP_H fixed cells by smFISH. The smFISH analysis of VNP_L cells will "highlight" the effects on priming gene expression (with VNP_H cells used as a control). It is expected that these results will provide new insights of the effects of GSK343 and/or AA in the cells that transited from low- to high-NANOG state, during 48 hours. Instead of having to analyse the global population of cells, with heterogeneous expression of NANOG, the sorting enables us to focus on the cells where histone *de novo* trimethylation might occur.

Moreover, it should be performed a functional test on VNP_L and VNP_H subpopulations treated with GSK343 and/or AA, to see if they can still form chimaeras and teratomas with the three germ layers, as well as the alkaline phosphatase assay, to infer about the self-renewal potential.

During my master project, I also fixed cells treated with GSKJ4, a small molecule inhibitor of histone demethylases UTX and JMJD3 (thereby blocking H3K27me3 demethylation). These fixed cells should also be analysed by smFISH, specifically to focus on cells expressing high levels of priming genes within low- and high-*Nanog* states. Preliminary data from DHenrique Lab has already showed that GSKJ4 exposure leads to a decrease of *Sox3* expression in low-*Nanog* cells.

It would be also interesting to perform a similar smFISH analysis in mES cells treated with these chemical modulators but in "2i" conditions, to study how would priming gene expression can be affected in a "ground-state" pluripotency, in which there is a higher percentage of high-*Nanog* cells.

Another challenge would be to mathematically define the probability density function from the histograms of the genes analysed, to better characterize the distribution type (not only by visual analysis and statistical parameters) and for threshold determination.

It is extremely difficult to capture the complex dynamics of "lineage priming" and to infer about correlations between *Nanog* and priming gene expression based only in the smFISH "snapshots". A time-lapse video to track Nd cells, with the Nanog:VNP reporter and, simultaneously, an antibody for simultaneous priming gene detection would be interesting to know more about the protein dynamics of "lineage priming" and to confirm its reversibility.

In fact, the correlation between Nanog mRNA and protein is still poorly understood. Technological advances, such as smFISH coupled to immunofluorescence, will make the simultaneous study of mRNA and protein possible, providing new insights on the meaning of the smFISH data.

Finally, it is essential to continue to optimize the signal acquisition on the new system Zeiss Axio Observer, for smFISH experiments. There are still few technical limitations that must be overcome to collect more reliable data. Moreover, other techniques such as single-cell RNA sequencing can be used, in order to obtain the whole transcriptome in single cells.

7. References

Abranches, E. et al., 2014. Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency. *Development (Cambridge, England)*, 141(14), pp.2770–9.

Abranches, E., Bekman, E. & Henrique, D., 2013. Generation and Characterization of a Novel Mouse Embryonic Stem Cell Line with a Dynamic Reporter of Nanog Expression. *PLoS ONE*, 8(3), pp.1–12.

Acampora, D., Di Giovannantonio, L.G. & Simeone, A., 2013. Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development (Cambridge, England)*, 140(1), pp.43–55.

Aloia, L., Di Stefano, B. & Di Croce, L., 2013. Polycomb complexes in stem cells and embryonic development. *Development*, 140(12), pp.2525–2534.

Avilion, A.A., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & Development*, 17(1), pp.126–140.

Azuara, V. et al., 2006. Chromatin signatures of pluripotent cell lines. *Nature cell biology*, 8(5), pp.532–8.

Barth, T.K. & Imhof, A., 2010. Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences*, 35(11), pp.618–626.

Batish, M., Raj, A. & Tyagi, S., 2011. Single Molecule Imaging of RNA In Situ. In *Methods*. pp. 3–13.

Bedzhov, I. et al., 2014. Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1657), pp.20130538–20130538.

Berger, S.L. et al., 2009. An operational definition of epigenetics An operational definition of epigenetics. , pp.781–783.

Bernstein, B.E. et al., 2006. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell*, 125(2), pp.315–326.

Blaschke, K. et al., 2014. Vitamin C induces Tet-dependent DNA demethylation in ESCs to promote a blastocyst-like state. , 500(7461), pp.222–226.

Boroviak, T., Nichols, J. & Nichols, J., 2014. The birth of embryonic pluripotency.

Boyer, L.A. et al., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441(7091), pp.349–353.

Bradley, a et al., 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309(5965), pp.255–256.

Brookes, E. et al., 2012. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell*, 10(2), pp.157–170.

Chambers, I. et al., 2003. Functional expression cloning ot nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5), pp.643–655.

Chambers, I. et al., 2007. Nanog safeguards pluripotency and mediates germline development. *Nature*, 450(7173), pp.1230–4.

Costa, Y. et al., 2013. NANOG-dependent function of TET1 and TET2 in establishment of

pluripotency. *Nature*, 495(7441), pp.370–4.

Elgin, S.C.R. & Grewal, S.I.S., 2003. Heterochromatin: silence is golden. *Current biology : CB*, 13(23), pp.R895–R898.

Elowitz, M.B. et al., 2002. Stochastic gene expression in a single cell. *Science*, 297(5584), pp.1183–1186.

Esteban, M.A. et al., 2010. Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. *Cell Stem Cell*, 6(1), pp.71–79.

Evans, M.J. & Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292, pp.154–156.

Gagliardi, A. et al., 2013. A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *The EMBO journal*, 32(16), pp.2231–47.

Gilbert, S., 2003. *Developmental Biology*,

Guedes, A.M. V., Henrique, D. & Abranches, E., 2016. Dissecting Transcriptional Heterogeneity in Pluripotency: Single Cell Analysis of Mouse Embryonic Stem Cells. In *Methods in Molecular Biology*. pp. 101–119.

Hackett, J.A. & Azim Surani, M., 2014. Regulatory principles of pluripotency: From the ground state up. *Cell Stem Cell*, 15(4), pp.416–430.

Heard, E., 2004. Recent advances in X-chromosome inactivation. *Current Opinion in Cell Biology*, 16(3), pp.247–255.

Henrique, D. et al., 2015. Neuromesodermal progenitors and the making of the spinal cord. *Development (Cambridge, England)*, 142(17), pp.2864–2875.

Ito, S. et al., 2010. Role of Tet proteins in 5mC to 5hmC conversion , ES cell self- renewal , and ICM specification. , 466(7310), pp.1129–1133.

Kohli, R.M. & Zhang, Y., 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, 502(7472), pp.472–9.

Lee, T.I. et al., 2006. Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell*, 125(2), pp.301–313.

Leitch, H.G. et al., 2013. Naïve pluripotency is associated with global DNA hypomethylation. , 20(3), pp.311–316.

Lindskog, S., 1997. Structure and mechanism of carbonic anhydrase. *Pharmacology & therapeutics*, 74(1), pp.1–20.

Marks, H. & Stunnenberg, H.G., 2014. Transcription regulation and chromatin structure in the pluripotent ground state. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(3), pp.129–137.

Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 78(12), pp.7634–7638.

Martinez Arias, A., Nichols, J. & Schröter, C., 2013. A molecular basis for developmental plasticity in early mammalian embryos. *Development*, 140(17), pp.3499–510.

Masui, S. et al., 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse

embryonic stem cells. *Nat Cell Biol*, 9(6), pp.625-U26.

Mitsui, K. et al., 2003. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell*, 113, pp.631–642.

Miyanari, Y. & Torres-Padilla, M.-E., 2012. Control of ground-state pluripotency by allelic regulation of Nanog. *Nature*, 483(7390), pp.470–473.

Morris, S.A. et al., 2010. Origin and formation of the first two distinct cell types of the inner cell mass in the mouse embryo. *Proc Natl Acad Sci U S A*, 107(14), pp.6364–6369.

Navarro, P. et al., 2012. OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells. *Embo J*, 31(24), pp.4547–4562.

Neri, F. et al., 2013. Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome biology*, 14(8), p.R91.

Neri, F. et al., 2015. TET1 is controlled by pluripotency-associated factors in ESCs and downmodulated by PRC2 in differentiated cells and tissues. *Nucleic acids research*, 43(14), pp.6814–6826.

Nichols, J. et al., 1998. Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell*, 95, pp.379–391.

Nichols, J. & Smith, A., 2009. Naive and Primed Pluripotent States. *Cell Stem Cell*, 4(6), pp.487–492.

Niwa, H. et al., 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development*, 12(13), pp.2048–60.

Niwa, H., Smith, A.G. & Miyazaki, J., 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*, 24(4), pp.372–376.

Osorno, R. et al., 2012. The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development*, 139(13), pp.2288–2298.

Osorno, R. & Chambers, I., 2011. Transcription factor heterogeneity and epiblast pluripotency. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 366(July), pp.2230–2237.

Pasini, D. et al., 2007. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol*, 27(10), pp.3769–3779.

Photometrics, CoolSNAP HQ Monochrome Data Sheet.

Pina, C. et al., 2012. Inferring rules of lineage commitment in haematopoiesis. *Nature Publishing Group*, 14(3), pp.287–294.

Raj, A. et al., 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*, 4(10), pp.1707–1719.

Raj, A. & van Oudenaarden, A., 2009. Single Molecule Approaches to Stochastic Gene Expression. *Annual review of biophysics*, (25), pp.255–270.

Raj, A. & Tyagi, S., 2010. *Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes.* 1st ed., Elsevier Inc.

Saiz, N. & Plusa, B., 2013. Early cell fate decisions in the mouse embryo. *Reproduction*, 145, pp.R65-80.

Sarkar, A. & Hochedlinger, K., 2013. The Sox Family of Transcription Factors: Versatile Regulators of

Stem and Progenitor Cell Fate. *Cell Stem Cell*, 12(1), pp.15–30.

Schrode, N. et al., 2013. Anatomy of a blastocyst: Cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo. *Genesis*, 51(4), pp.219–233.

Schübeler, D., 2015. Function and information content of DNA methylation. *Nature*, 517(7534), pp.321–326.

Schwarz, B.A. et al., 2014. Nanog is dispensable for the generation of induced pluripotent stem cells. *Current Biology*, 24(3), pp.347–350.

Sharova, L. V et al., 2009. Database of mRNA Half-Life of 19977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells Supplementary data. *DNA research*, 16(1), p.S1.

Shen, X. et al., 2008. EZH1 Mediates Methylation on Histone H3 Lysine 27 and Complements EZH2 in Maintaining Stem Cell Identity and Executing Pluripotency. *Molecular Cell*, 32(4), pp.491–502.

Silva, J. et al., 2009. Nanog Is the Gateway to the Pluripotent Ground State. *Cell*, 138(4), pp.722–737.

Silva, J. & Smith, A., 2008. Capturing Pluripotency. *Cell*, 132(4), pp.532–536.

Singh, A.M. et al., 2007. A Heterogeneous Expression Pattern for Nanog in Embryonic Stem Cells. *Stem Cells*, 25(10), pp.2534–2542.

Sui, X. et al., 2012. Crosstalk Between DNA and Histones: Tet's New Role in Embryonic Stem Cells. *Current Genomics*, 13(8), pp.603–608.

Surface, L.E., Thornton, S.R. & Boyer, L.A., 2010. Polycomb group proteins set the stage for early lineage commitment. *Cell Stem Cell*, 7(3), pp.288–298.

Takahashi, K. & Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), pp.663–676.

Takeda, K. et al., 1997. Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), pp.3801–4.

Torres-Padilla, M.-E.E. & Chambers, I., 2014. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage [review]. *Development (Cambridge, England)*, 141 VN-(11), pp.2173–2181.

Tosolini, M. & Jouneau, A., 2015. From Naive to Primed Pluripotency: In Vitro Conversion of Mouse Embryonic Stem Cells in Epiblast Stem Cells. In *Methods in Molecular Biology*. pp. 209–216.

Verma, S.K. et al., 2012. Identification of potent, selective, cell-Active inhibitors of the histone lysine methyltransferase EZH2. *ACS Medicinal Chemistry Letters*, 3(12), pp.1091–1096.

Villasante, A. et al., 2011. Epigenetic regulation of Nanog expression by Ezh2 in pluripotent stem cells. *Cell Cycle*, 10(9), pp.1488–1498.

Williams, R.L. et al., 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*, 336(6200), pp.684–687.

Wu, H. et al., 2010. Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes. , 329(5990), pp.444–448.

Wu, H. et al., 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, 473(7347), pp.389–93.

Wu, H. et al., 2014. Vitamin C Enhances Nanog Expression Via Activation of the JAK / STAT Signaling Pathway. *Stem Cells*, 32, pp.166–176.

Yin, R. et al., 2013. Ascorbic acid enhances tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *Journal of the American Chemical Society*, 135(28), pp.10396–10403.

Ying, Q.-L. et al., 2008. The ground state of embryonic stem cell self-renewal. *Nature*, 453(September 2016), pp.519–23.

Zeiss, 2013. ZEISS Axiocam 506 mono Product Information.

Zhao, S. et al., 2004. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Molecular and Cellular Neuroscience*, 27(3), pp.332–342.

Biosearch Technologies. *Stellaris RNA FISH*. Available at: https://www.biosearchtech.com/support/education/stellaris-rna-fish (consulted at 15.02.2016).

Bitbucket. *Rajlabimagetools/Home*. Available at: https://bitbucket.org/arjunrajlaboratory/rajlabimagetools/wiki/Home (consulted at 15.02.2016).

McFee, C. *Noise sources in a CCD*. Available at: http://www.mssl.ucl.ac.uk/www_detector/optheory/darkcurrent.html (consulted at 01.08.2016).

McDonald, J. H., 2014. *Handbook of Biological Statistics – Spearman rank correlation.* Available at: http://www.biostathandbook.com/spearman.html (consulted at 18.08.2016).

Premium Beat, 2016. *What is a Hot Pixel and How Can You Remove One?* Available at: http://www.premiumbeat.com/blog/what-is-a-hot-pixel-and-how-can-you-remove-one/ (consulted at 01.08.2016).

QSI, 2013. *Understanding CCD read noise*. Available at: http://qsimaging.com/ccd_noise.html (consulted at 01.08.2016).

8. Annex A

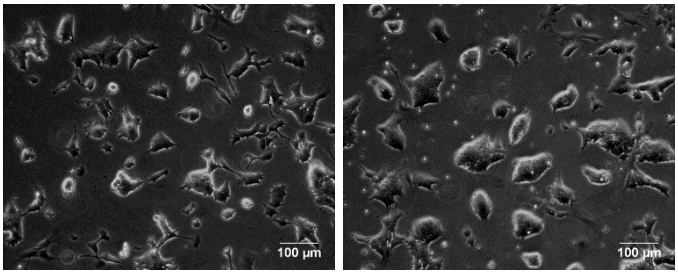


***Figure A1 – Cell morphology of E14 (on the left) and Nd (on the right) mES cells**. mES cells were cultured in "Serum/LIF" conditions and these bright field images were taken before a new passage. Scale bar = 100 µm.*

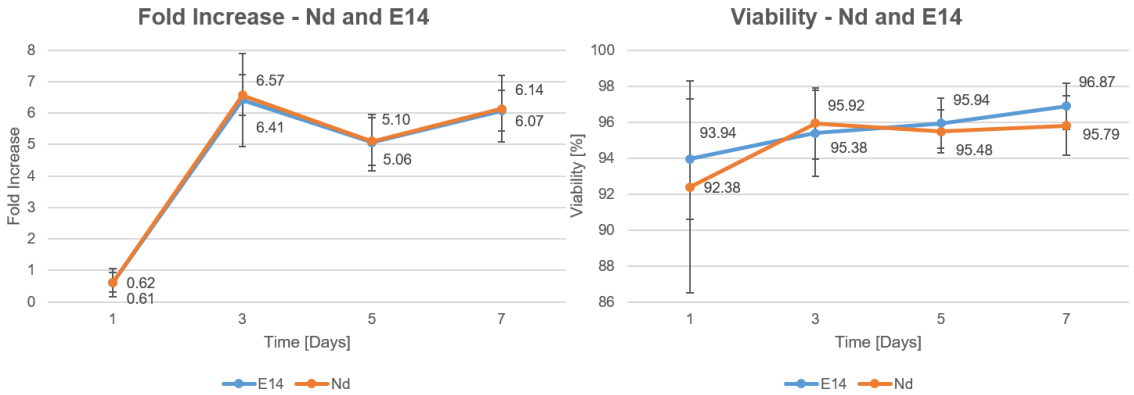


***Figure A2 – Fold increase and viability calculated for Nd and E14 mES cells**. The fold increase and viability are lower in the first day, due to the fragile condition of the cells and recent adaptation to the new culture medium. In the next days, it shows a fold increase varying from 4 to 8, within the normal range calculated for these mES cells. Viability was also measured and its values are within the normal range, from 86% to 98%. Mean values were depicted for each day and error bars were calculated based on standard deviation from four experiments (n=4).*

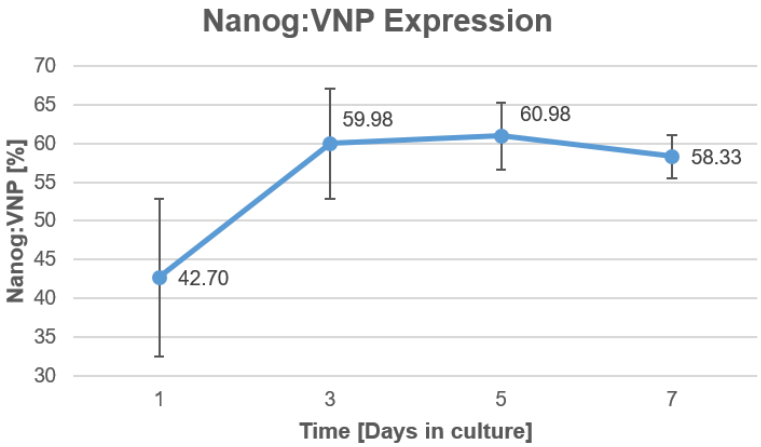


***Figure A3 – Analysis of the Nanog:VNP reporter expression in Nd cells by flow cytometry.** Nanog:VNP values, measured every 48 hours, are within the normal range of (56.2 ± 8.0)%. Mean values are depicted for each day. Error bars were calculated based on standard deviation from four experiments (n=4).*

***Table A1 – Threshold for each gene analysed by smFISH**. The threshold value for Nanog and Sox2 was determined by visual observation of the histogram of mRNA distribution. For Otx2, Car2 and Sox3 was the top-5% expressing cells through a quantile of 95%. There are two values for Car2 and Sox3, the first is for the priming analysis, in the first subchapter of the results, and the second is for the exposure of chemical modulator's, in the second subchapter (based on DMSO analysis). Different thresholds were used because cells were fixed on different days and also the culture conditions were different (DMSO was used as control for chemical modulator's analysis). Ideally, it should be always used the same threshold.*

Gene	Threshold
Nanog – Pluripotency	50
Sox2 – Pluripotency	50 – For pluripotency analysis
Otx2 – Epiblast transition	45 – For priming analysis
Car2	305 – For priming analysis 324 – For chemical modulator's analysis
Sox3 – "Lineage Priming"	41 – For priming analysis 53 – For chemical modulator's analysis

Table A2 – Summary of data analysis obtained by smFISH for Car2-Nanog-Sox3 and Nanog-Otx2-Sox3 probe combinations. *For each gene state combination is presented the number of cells (Ncells), the percentage of cells with that specific combination within the total and relatively to Nanog. In the columns of Nanog- and Nanog+ is the percentage of low-Nanog and high-Nanog cells, respectively, expressing high levels of Car2, Sox3 or both transcripts. Finally, the last column has the predicted percentage of low-Nanog cells with high expression of Car2 and Sox3 if they were independent events (product between the percentage of low-Nanog cells expressing high-Car2 and high-Sox3).*

Car and Sox3 Analysis						Nanog-			Nanog+			Nanog-
Nanog	Car2	Sox3	Ncells	Total	Nanog	Car2+	Sox3+	Car2+Sox3+	Car2+	Sox3+	Car2+Sox3+	Car2+Sox3+
-	-	-	120	29.41	86.33							Predicted
-	-	+	8	1.96	5.76		8.63					
-	+	+	4	0.98	2.88	7.91		2.88				0.68
-	+	-	7	1.72	5.04							
+	-	-	251	61.52	93.31							
+	-	+	8	1.96	2.97					3.35		
+	+	+	1	0.25	0.37				3.72		0.37	
+	+	-	9	2.21	3.35							
Otx2 and Sox3 Analysis						**Nanog-**			**Nanog+**			**Nanog-**
Nanog	**Otx2**	**Sox3**	**Ncells**	**Total**	**Nanog**	**Otx2+**	**Sox3+**	**Otx2+Sox3+**	**Otx2+**	**Sox3+**	**Otx2+Sox3+**	**Otx2+Sox3+**
-	-	-	305	44.85	87.64							Predicted
-	-	+	25	3.68	7.18		7.47					
-	+	+	1	0.15	0.29	5.17		0.29				0.39
-	+	-	17	2.50	4.89							
+	-	-	309	45.44	93.07							
+	-	+	7	1.03	2.11					2.71		
+	+	+	2	0.29	0.60				4.82		0.60	
+	+	-	14	2.06	4.22							

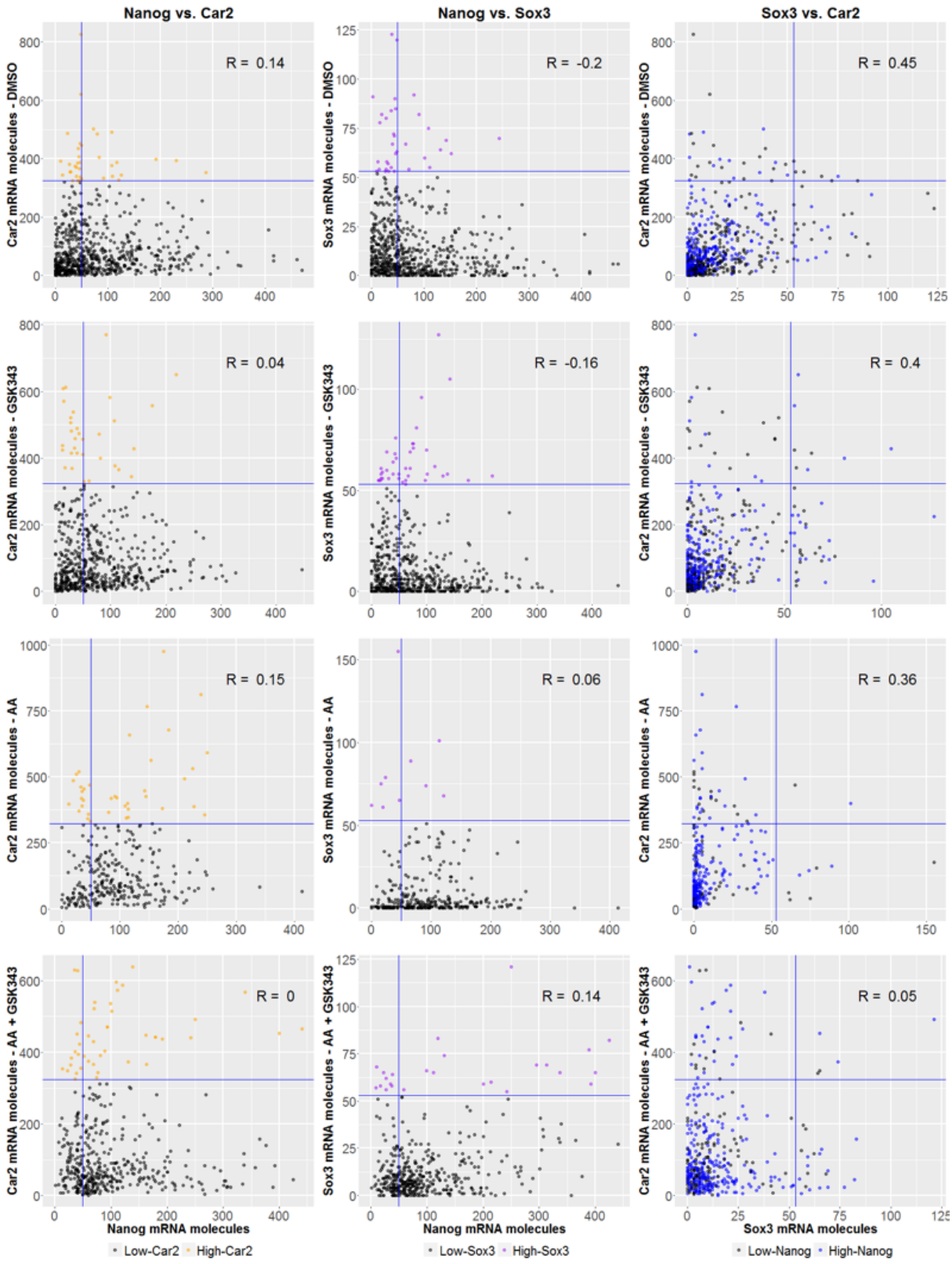


Figure A4 – Correlation scatterplots for the three possible combinations between Nanog, Sox3 and Car2 within the different conditions. *Each dot represents a cell and the cells expressing high levels of Car2 are coloured in orange, Sox3 in purple and Nanog in blue. R is the Spearman correlation coefficient and it is depicted for each combination.*

Table A3 – Summary of data analysis obtained by smFISH upon exposure to chemical modulators. *For each gene state combination is presented the number of cells (Ncells), the percentage of cells with that specific combination within the total and relatively to Nanog. In the columns of Nanog- and Nanog+ is the percentage of low-Nanog and high-Nanog cells, respectively, expressing high levels of Car2, Sox3 or both transcripts.*

DMSO Condition						Nanog-			Nanog+		
Nanog	Car2	Sox3	Ncells	Total	Nanog	Car2+	Sox3+	Car2+Sox3+	Car2+	Sox3+	Car2+Sox3+
-	-	-	273	40.63	86.94						
-	-	+	18	2.68	5.73		7.64				
-	+	+	6	0.89	1.91	7.32		1.91			
-	+	-	17	2.53	5.41						
+	-	-	335	49.85	93.58						
+	-	+	10	1.49	2.79					3.07	
+	+	+	1	0.15	0.28				3.63		0.28
+	+	-	12	1.79	3.35						
GSK343 Condition						**Nanog-**			**Nanog+**		
Nanog	Car2	Sox3	Ncells	Total	Nanog	Car2+	Sox3+	Car2+Sox3+	Car2+	Sox3+	Car2+Sox3+
-	-	-	210	36.91	86.07						
-	-	+	15	2.64	6.15		7.38				
-	+	+	3	0.53	1.23	7.79		1.23			
-	+	-	16	2.81	6.56						
+	-	-	295	51.85	90.77						
+	-	+	17	2.99	5.23					6.77	
+	+	+	5	0.88	1.54				4.00		1.54
+	+	-	8	1.41	2.46						
AA Condition						**Nanog-**			**Nanog+**		
Nanog	Car2	Sox3	Ncells	Total	Nanog	Car2+	Sox3+	Car2+Sox3+	Car2+	Sox3+	Car2+Sox3+
-	-	-	57	18.87	74.03						
-	-	+	5	1.66	6.49		7.79				
-	+	+	1	0.33	1.30	19.48		1.30			
-	+	-	14	4.64	18.18						
+	-	-	199	65.89	88.44						
+	-	+	3	0.99	1.33					1.78	
+	+	+	1	0.33	0.44				10.22		0.44
+	+	-	22	7.28	9.78						
AA+GSK343 Condition						**Nanog-**			**Nanog+**		
Nanog	Car2	Sox3	Ncells	Total	Nanog	Car2+	Sox3+	Car2+Sox3+	Car2+	Sox3+	Car2+Sox3+
-	-	-	104	23.80	83.20						
-	-	+	7	1.60	5.60		7.20				
-	+	+	2	0.46	1.60	11.20		1.60			
-	+	-	12	2.75	9.60						
+	-	-	269	61.56	86.22						
+	-	+	13	2.97	4.17					5.13	
+	+	+	3	0.69	0.96				9.62		0.96
+	+	-	27	6.18	8.65						