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

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: Bio/18

Induced Pluripotent Stem Cells: An Innovative Tool to Dissect Ovarian Cancer Pathogenesis

Pietro Lo Riso

European Institute of Oncology, Milan

Matricola n. R09841

Supervisor: Prof. Giuseppe Testa

European Institute of Oncology, Milan

University of Milan, Milan

Anno accademico 2014-2015

1. TABLE OF CONTENTS

2. FIGURES AND TABLES INDEX

3. LIST OF ABBREVIATIONS

5hmC	5'-hydroxymethylcytosine
AS	Ascite
CGH	Comparative genome hybridization
ChIP	Chromatin immunoprecipitation
CNV	Copy number variation
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
ddH2O	Bi-distilled water
DSB	Double strand break
EDM	Epithelial Digestion medium
EF1a	Elongation Factor 1 alpha
EOC	Epithelial Ovarian Cancer
EPI	Epithelial cells culturing medium
ESC	Embryonic stem cells
FACS	Fluorescence activated cell sorting
FBS	Fetal Bovine Serum
FI	Fimbria
FIGO	International Federation of Gynecology and Obstetrics
FRT	Flippase recognition target
FRTE	Female reproductive tract epithelium
FSC	Forward scatter

FW	Forward
GBM	Glioblastoma
gDNA	Genomic DNA
GFP	Green fluorescent protein
gRNA	Guide RNA
HGSOC	High grade serous ovarian cancer
hiPSC	Human iPSC
iPSC	Induced Pluripotent Stem Cells
IRES	Internal ribosome entry site
ITR	Inverted terminal repeat
LMP	Low malignant potential tumors
LTR	Long terminal repeat
LV	Lentiviral vector
MEFs	Mouse embryonic fibroblast
NHEJ	Nonhomologous end joining
NOMe	Nucleosome occupancy and DNA methylation
NS	Neural stem
NuFF	Newborn Foreskin Fibroblasts
OC	Ovarian cancer
OSE	Ovarian Surface Epithelium
PanIN	Pancreatic intraepithelial neoplasia
PARP	Poly(ADP)ribose
PcG	Polycomb

PDAC	Pancreatic ductal adenocarcinoma
PDX	Patient-derived xenograft
PGK	Phosphoglycerate kinase
PNET	Primary neuroectodermal tumor
PTMs	Post-translational modifications
RT	Room temperature
rtTA	Reverse tetracycline transactivator
RV	Reverse
SCNT	Somatic cell nuclear transfer
SNV	Single nucleotide variation
SOC	Serous ovarian cancer
SSC	Side scatter
STEMCCA	Stem cell cassette
STIC	Serous intraepithelial carcinomas
TAE	Tris-acetate-EDTA
TAL	Transcription activator-like
tracrRNA	Trans-activating CRISPR RNA
TrxG	Trithorax
TSS	Transcription start site

4. ABSTRACT

Ovarian cancer (OC) has one of the highest death-to-incidence ratios among all tumor types, which points to the need for novel therapeutic and prognostic strategies. Indeed, the absence of relevant tumor cell lines that can recapitulate disease histopathology highlights an acute need for new model systems to study this pathology. In particular, it is still unclear whether the most common and aggressive form of this disease, high grade serous ovarian cancer (HGSOC), could arise from in the ovarian surface epithelium (OSE), as initially thought, or might be arising from the fimbrial epithelium. Here I addressed these issues in two complementary ways based on induced pluripotent stem cells: i) the modeling of Ovarian Cancer by somatic cell reprogramming to pluripotency of tumor cells; ii) the molecular characterization of HGSOC and its putative cells of origin. Somatic cell reprogramming, by erasing tumor-associated epigenetic marks while preserving the underlying genetic mutations, would allow for the first time the precise dissection of genetic and epigenetic contribution to this disease, through the differentiation of OC-iPSC into disease-relevant cell types. I demonstrated the feasibility of OC reprogramming through a non-integrative platform, showing that OC-derived iPSC are closely similar to human ESC, and proving their tumoral origin by whole exome sequencing. Moreover, I showed that independent iPSC clones derived from the same tumor upon trilineage differentiation *in vivo* show differential tumorigenic potential. For a more precise dissection of this phenotype, I set up a differentiation protocol

that allows differentiation of pluripotent cells into mesodermal progenitors, that are precursors of both fimbria and OSE. To isolate a pure population of these cells, I resorted to CRISPR/Cas9 to integrate a selection cassette in the MIXL1 locus. By this approach, I was able to show correct gene targeting at the intended site, allowing also for selection of mesodermal progenitors upon differentiation of normal iPSC. The same approach translated to OC-derived iPSC would allow to study the effects of genetic mutations deprived of tumor-associated epigenetic marks during differentiation, both at the stage of mesodermal progenitors and in cells directed towards the female reproductive epithelium *in vivo*. The second approach relies on the identification of specific molecular features of fimbria and ovarian surface epithelium, the two putative cells of origin of HGSOC. On this side, I offer a first glimpse on molecular features of HGSOC cancer and normal gynecological tissues. I could show that specific DNA methylation signatures of fimbrial epithelial cells and ovarian surface epithelium cells are partially retained in tumor samples and stratify HGSOC samples according to the putative cell of origin of this tumor. Moreover, I show for the first time a description of histone modifications in primary HGSOC, concentrating on marks of activation/repression sitting on promoter regions (H3K4me3 and H3K27me3, respectively) and marks that characterize active/closed-poised enhancers (H3K4me1, H3K27ac and H3K27me3).

5. INTRODUCTION

5.1 Ovarian Cancer

Ovarian cancer (OC) is the fifth cause of cancer-related death for women, and the most lethal malignancy among the gynecological ones (Bowtell 2010). This disease is usually diagnosed at a late stage, being mostly asymptomatic or with vague symptoms that can be attributed to other gastrointestinal or reproductive system diseases.

The standard treatment is mainly based on extensive surgery and treatment with *cis*-platin (or carboplatin), that was introduced in the clinical practice in the late 70s, more recently in combination with taxanes. Since then, no major advances in care of these patients were obtained, made exception of reduced side-effects. Indeed, the percentage of 5-years disease-free survival of patients treated with surgery and chemotherapy is still below 40% (Vaughan et al. 2011) (Figure 1). A subset of OC carrying BRCA1/2 mutations, accounting for less than 10% of all OC, is eligible for treatment with poly(ADP)ribose (PARP) inhibitors (Bryant et al 2005, Audeh et al. 2010).

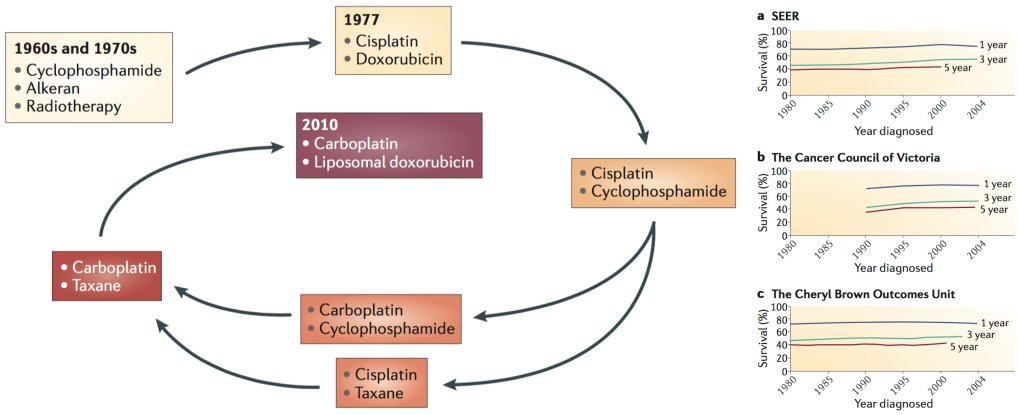

Figure 1. Current treatments of Ovarian Cancer have not ameliorated patient's care.

Left panel: the timeline of ovarian cancer treatments since 1960s to date. Right panel: disease-free survival curves showing no major improvements since 1980s (adapted from Vaughan et al. 2011).

The term "Ovarian Cancer" refers to a heterogeneous group of neoplasms rather than a single type of tumor. OC might arise from three different cell types: i) epithelial cells; ii) sex cord/stromal cells; iii) germ cells. About 40% of all tumors belonging to this group are originally non-epithelial, and usually do not progress to the malignant stage, accounting in the end for only about 10% of ovarian tumors (Karst and Drapkin 2010). Indeed, 90% of these tumors arise from epithelial cells (Epithelial Ovarian Cancer, EOC), constituting the most predominant form of the disease. EOC *per se* is a very heterogeneous group of tumors that can be classified into eight different subtypes, according to the World Health Organization: serous, endometrioid, mucinous, clear cell, transitional cell, squamous cell, mixed epithelial, and undifferentiated (Tavassoli and Devilee 2003). Within each subtype, tumors are further

described as either benign, malignant, or borderline (low malignant potential tumors, LMP) and, depending upon tumor subtype, classified as low- or high-grade.

Upon diagnosis of malignancy, ovarian tumors are surgically staged, according to the International Federation of Gynecology and Obstetrics (FIGO), to determine how far they have extended beyond the ovary. Stage I tumors are confined to the ovary or fallopian tubes. Stage II tumors extend from ovaries and/or fallopian tube to adjacent pelvic structures. Stage III tumors are characterized by metastasis to the peritoneum and/or to regional lymph nodes. Stage IV tumors metastasize to distant sites (Prat et al. 2015).

Serous ovarian cancer (SOC) is the most common form of this tumor and is classified based on histolopathology and mutational patterns into Type I and Type II tumors (Vang et al. 2009).

Type I SOC evolve by slow transformation of LMP. It comprises low grade serous tumors and serous borderline tumors and is characterized by frequent mutations in BRAF and KRAS (Singer et al 2003) and devoid of TP53 mutations (Wong et al 2010).

An opposed behaviour can be found in Type II SOC that comprises mainly high grade serous ovarian cancer (HGSOC). Indeed, TP53 mutations are almost ubiquitous (Ahmed et al 2010) while BRAF and KRAS mutations are usually not present (Wong et al 2010). Moreover, mutations in DNA repair-genes BRCA1 and BRCA2 have been reported (Geisler et al 2002, Hilton et al 2002), conferring to this type of tumors high chromosomal instability and hence more

sensitivity to platin- and PARP inhibitors-based treatments (Bowtell et al. 2010) (Figure 2).

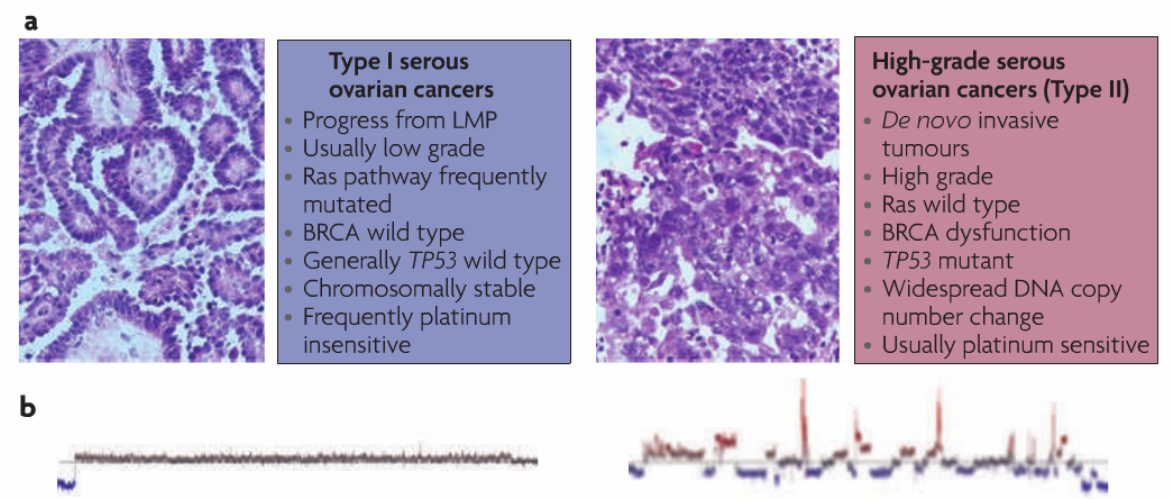

Figure 2. Classification of serous ovarian cancers.

a. Histopathological and mutational features of type I and type II SOC; b. DNA copy number plot of typical Type I and Type II SOC (Bowtell et al. 2010).

These symmetrical features underline the unlikelihood that high grade serous ovarian cancer are temporal progressions of low grade serous tumors, but rather a *de novo* neoplasm characterized by his own typical genetic and epigenetic aberrations.

5.1.1 Ovarian cancer cell of origin

The identification of ovarian cancer cell of origin is still a matter of debate. Traditionally, since these tumors are diagnosed when they already have invaded most of the abdominal portion of patients, including the ovaries, they have been attributed an ovarian origin. Growing evidence is suggesting that

HGSOC might be instead derived from epithelial cells of the distal portion of the tuba, namely the fimbrial epithelium, that is located in close proximity to the ovary (Figure 3).

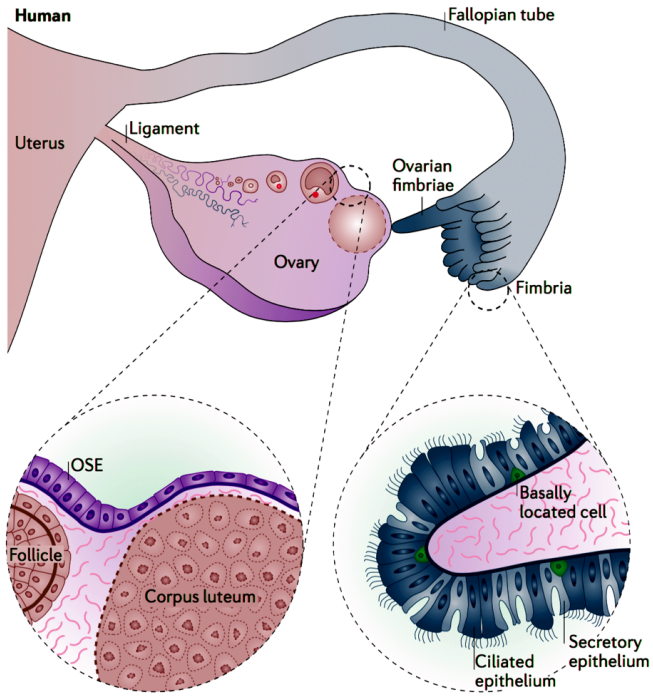

Figure 3. Anatomy of the female reproductive tract

Schematic representation of the female reproductive tract showing the close proximity of the ovarian surface epithelium (OSE) and the epithelium from the distal portion of the fimbria (adapted from Ng and Barker, 2015).

Studies with matched serous tubal intraephitelial carcinomas (STIC) and HGSOC revealed the presence of the same TP53 mutations in matched samples, suggesting that these tumors could be clonal evolution of pre-neoplastic lesions in the fimbria (Kuhn et al. 2012). This “p53 signature”, that might be even affecting resident stem cells in the fimbria, unveils a scenario of fimbria epithelial/stem cells that acquire TP53 mutations and, given the close proximity to the ovarian surface epithelium (OSE), shed from their location and get trapped in inclusion cysts in the ovary (Ng and Barker, 2015).

Very few examples of pre-neoplastic lesions in the OSE have been reported for HGSOC (Roland et al 2003). This can be attributed either to the fimbrial origin of this tumor, but also to the more permissive environment that ovary constitutes for the growth of tumors, as highlighted by the high frequency of ovarian metastases originated from gastrointestinal, breast and lung cancers (Young 2007). This latter aspect would reconcile with the fact that usually in patients carrying very spread HGSOC, STIC are microscopic in size. Most probably, the fimbrial environment is less permissive to invasiveness and growth than the ovary. When STIC cells are captured in the ovarian stroma, they find a favourable environment for growth and development, with mechanisms that are still poorly defined (Parrott et al. 2001, Schauer et al. 2011). So, tumors arising from OSE, being plunged in an optimal environment for growth, might be progressing at such rate that it is hard to capture pre-neoplastic or early lesions in this site. An additional controversial aspect is related to the presence of traits of the Muellerian ducts (from which the fallopian tube is derived, but not the ovary) in ovarian cancer, especially in low grade and borderline tumors. Also in this case, this is either attributed to cells shedding from the fimbrial epithelium and included into the ovary, or to metaplasia of the inclusion cysts of the ovary, generated from the OSE, into "fimbrial-like" cells, possibly for either a high plasticity of putative resident stem cells (Szotek et al 2008, Bowen et al. 2009), or related to the common mesodermal origin of these two organs. This might explain the preferential Muellerian differentiation of OSE cells during metaplasia and neoplastic

transformation (Auersperg 2013).

Given the supporting evidence of both theories, a more unifying vision can be envisaged, identifying in both tissues the plausible origin of a subset of HGSOC. The identification of signatures that allow to identify for each patient's tumor the precise cell of origin, might allow for a better design of studies aimed at the identification of critical pathways that contribute to ovarian cancer pathogenesis.

5.1.2. Current models for high grade serous ovarian cancer

Despite several cell lines are available to model high grade serous EOC, these are poorly characterized in terms of histopathological features and site of origin (Vaughan et al. 2010). As recently shown in a comparative genomic and expression study (Domcke et al. 2013), the cell lines that are mostly used in laboratories are quite dissimilar to primary HGSOC samples in terms of the amount of genetic mutations, copy number alterations and in terms of gene expression patterns (Figure 4).

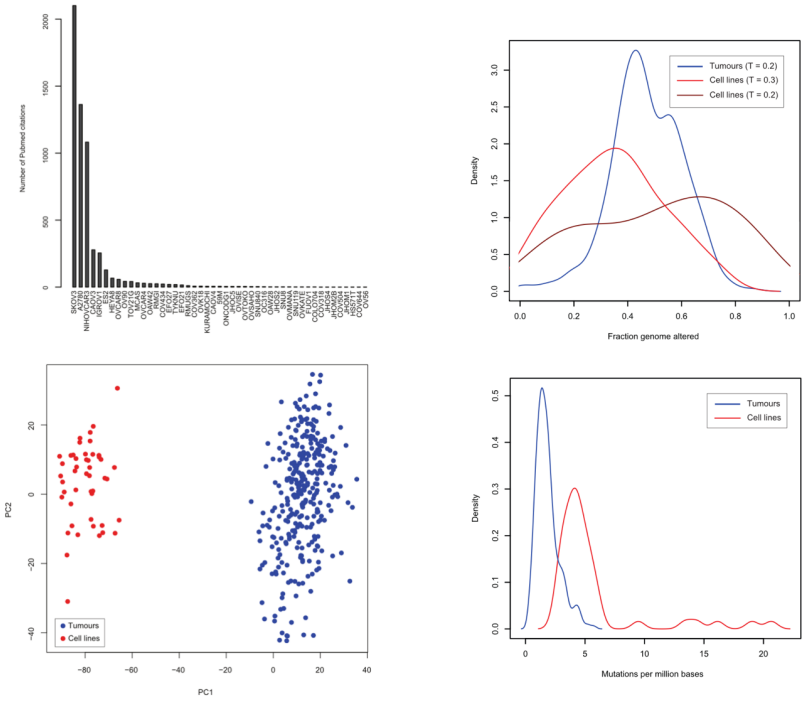

Figure 4. Most commonly used cell lines do not recapitulate HGSOC phenotypes.

Top left panel: number of publications employing the depicted cells lines. Bottom left panel: principal component analysis showing the unrelatedness of ovarian cancer cell lines with primary tumor samples at the level of gene expression. Right panels: fraction of altered genome (top) and number of mutations per million bases (bottom) of cancer cell lines and primary HGSOC (from Domcke et al. 2013).

While this analysis suggests that some undervalued cell lines might be instead more useful to model this tumor subtype, highlighting on the contrary that the most frequently used ones may prove useless to model HGSOC, such as in the case of SKOV3 cells being more closely related to endometrioid cancers, there is a strong consensus that new well characterized cellular models should be established.

Animal models can be useful to reveal new insights on HGSOC.

Patient-derived xenografts (PDX) have been extensively used to propagate *in vivo* primary tumor cells, that quickly undergo senescence upon culturing.

They have proved to be useful models to assess drug response *in vivo* and to recapitulate tumors' heterogeneity (Scott et al. 2013). The limitation of this approach lies in the use of immunocompromised mice, probably lowering the impact of tumor microenvironment on tumor growth. Moreover, it has been shown that at the first passage transplanted tumors lose markers of human stroma and vasculature (Hylander et al. 2013), suggesting that host vascular and stromal system sustain the growth of engrafted tumors. This can obviously affect studies aimed at targeting tumor microenvironment.

Genetically modified animals instead have been used to address mainly the the problem of the origin of HGSOC. Perets and colleauges used a mouse model conditionally deleting Brca, Tp53 and Pten specifically in fallopian tube epithelium and not in the ovarian surface epithelium by means of a Pax8-Cre (Perets et al. 2013). These mice develop high grade serous ovarian cancer, going through STIC, a pre-neoplastic lesion of the fallopian tube. In accordance with these results, Kim and colleagues (Kim et al. 2012) showed that specific deletion of Pten and Dicer in Anti Muellerian Hormone Receptor 2 (Amhr2) expressing cells (general female reproductive tract epithelial cells) causes emergence of HGSOC from the fallopian tube only, even after removal of the ovaries. The same group, though, showed that an additional gain of function mutation of Tp53 ($p53^{R172H}$) results in emergence of HGSOC both from the

fallopian tube and the ovary, even in the presence of a wild-type Dicer, suggesting that both organs can be suited to the development of this kind of tumor (Kim et al. 2015).

Additional work from Flesken-Nikitin and colleagues (Flesken-Nikitin et al. 2013), identified a tumor-prone stem cell niche in the hilum, the junctional area between OSE, distal fallopian tube epithelium and mesothelium. This area, however, is not present in the human specie (Ng and Barker 2015), suggesting that this finding might be relegated to the murine setting.

These controversial results indicate that genetically modified mice can be useful to study peculiar mutations, but fail to recapitulate the whole complexity of HGSOC.

5.2 Epigenetic aberrations

5.2.1. DNA methylation

DNA methylation is a covalent modification occurring at cytosines of 5'-CG-3' dinucleotides (CpG), that is deposited early during development.

This mark is deposited and maintained by DNA methyltransferases DNMT1, DNMT3A, DNMT3B. In normal conditions, CpG islands, stretches of CpG sequences usually associated with promoters, are hypomethylated in the context of surrounding methylation, a condition that is usually associated with

active gene expression. Further modification of methylcytosine by TET enzymes results in the deposition of 5'-hydroxymethylcytosine (5hmC), that has been proposed as an initial step toward de-methylation. Despite this, stable presence of this mark has been reported in active gene bodies (Branco, Ficz and Reik 2012) and seems to have a role in areas of plastic nucleosome remodeling during differentiation (Teif et al 2014).

Aberrant DNA methylation occurs commonly in tumors and is considered to be one of the earliest molecular changes in carcinogenesis (Baylin and Jones 2011, Baylin and Ohm 2006). (Figure 5)

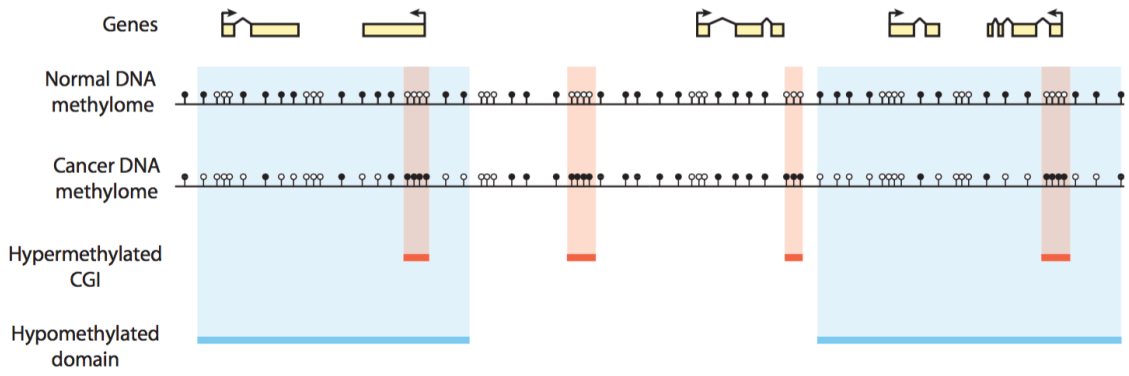

Figure 5. Aberrant DNA methylation in cancer.

Schematic representation of methylation aberrations in cancer. Frequently, TSS-associated CpG islands become hypermethylated in the context of larger hypomethylated domains (Reddington, Sproul and Meehan 2010).

Candidate gene and whole-genome studies have identified methylation signatures that may serve as biomarkers for HGSOC characterization including classification (Barton et al. 2008), progression (Wei et al. 2002) and response to therapy (Wei et al. 2006). Hypermethylation of DNA occurs mainly at the

promoter level of well known tumor suppressor genes, such as BRCA1 (Baldwin et al 2000), p16 (Katsaros et al 2004) and MLH1 (Zhang et al 2008), while hypomethylation has been shown to be occurring in the promoter of oncogenes, such as BORIS (Woloszynska-Read et al 2007), CLDN3 and CLDN4 (Honda et al. 2006 and 2007), and in repeated regions of the genome such as pericentromeric Sat2 DNA at chromosome 1 (Widschwendter et al. 2004), associated with poor prognosis. It has been suggested that this phenomenon increased susceptibility to genomic instability and re-activation of retro-transposons (Esteller 2008).

5.2.2. Histone modifications

DNA is wrapped around globular protein complexes known as histones to form the fundamental repeating unit of chromatin, the nucleosome. This is constituted by 146 base pairs of DNA packed around a histone octamer, which is composed by two copies each of histones H2A, H2B, H3 and H4, with the addition of one copy of histone H1 that tops this structure and controls higher order chromatin compaction. The degree of compaction governs the accessibility of DNA, hence the tight regulation of this property plays a crucial role in dictating and propagating distinct patterns of gene expression. The fine adjustment of DNA accessibility to transcriptional effectors is achieved by post-translational modifications (PTMs) of histone tails that extend out of the nucleosome core. Several modifications have been described, including

methylation, acetylation, phosphorylation, poly(ADP)-ribosylation and ubiquitilation. These modifications can play different roles: i) they can serve as docking sites for non-histone proteins, that have conserved domains able to recognize such modifications, tailoring the recruitment of proteins to specific genomic loci; ii) they can alter the ionic interaction between histones and wrapped DNA regulating chromatin compaction and permissiveness to binding of transcription factors, such as in the case of histone acetylation; iii) they can mediate the establishment, maintenance and heritability of the transcriptional landscape, such as in the case of Polycomb and Trithorax group of proteins-mediated histone modifications (Kouzarides 2007, Orkin and Hochedlinger 2011, Laugesen and Helin 2014, Steffen and Ringrose 2014). Thus, every cell of a given organism is characterized by a particular gene expression pattern also as the result of the tight interplay between transcription factors and different histone modifications.

Among the best characterized histone modifications, there is methylation of H3K4 and H3K27 that is catalyzed by the Trithorax (TrxG) and Polycomb (PcG) groups of proteins, respectively. In particular, trimethylation (me3) of these lysines is associated to active (Byrd and Shearn 2003) and repressed genes (Kirmizis et al 2004), respectively. These groups of proteins were first identified in *Drosophila melanogaster* as regulators of the spatio-temporal expression of *Hox* genes (Lewis 1978), and their action was found to be regulating a plethora of other genes, including transcription factors involved in cell fate establishment and maintenance. Their role has been shown to be

extensively conserved also in mammals (Bracken, Dietrich et al. 2006).

Genome-wide distribution of these two epigenetic marks have highlighted four fundamental chromatin states determined by their tight interplay (Pan et al 2007, Zhao et al 2007, Mikkelsen et al 2007): i) a repressed state, characterized by the presence of H3K27me3 at the promoter of genes; ii) an active state, characterized by H3K4me3 at promoter regions of genes; iii) a bivalent state, with concomitant presence of both H3K27me3 and H3K4me3 peaks especially at the promoter of developmental related genes, which is permissive for low mRNA transcription and eager to activation/repression according to lineage specification (Bernstein et al. 2006); iv) a mute state, in which both marks are absent, and RNA polymerase II is not bound.

Alterations of the normal equilibrium between these two marks have been associated with many tumors. For example, EZH2, the catalytic subunit of Polycomb repressive complex 2 that mediates the deposition of H3K27me3, is frequently overexpressed in metastatic prostate (Varambally et al. 2002), breast (Kleer et al. 2003, Raaphorst et al. 2003), and bladder cancer (Arisan et al. 2005), and can promote cancer progression through the silencing of the p14 and p16 (Ink4A/ARF locus) (Bracken et al. 2007). Lysine Methyltransferases (KMT2 or MLL), members of TcG and the key regulator of H3K4 methylation, are frequently mutated in a variety of cancers (Kandoth et al. 2013).

In the case of HGSOC, the role of histone modifications is still unclear. High expression of EZH2 was found to be correlated with advanced stage, poor

survival (Rao et al. 2010), and cisplatin resistance (Hu et al. 2010). This suggests a role of repressive histone marks in sustaining late progression of ovarian cancer. An interesting finding showed that bivalently marked genes in ovarian cancer can favour malignant progression and confer chemoresistance to tumors (Chapman-Rothe et al. 2013). Being bivalent domains characteristic of pluripotent stem cells (Pan et al. 2007, Xhao et al. 2007), this might suggest the emergence of stem-like properties at advanced stages of this disease.

5.3. Inducing Pluripotency in differentiated cell types

Cell potency is defined as the capability of a defined cell to differentiate into other cell types. This attribute is progressively lost in the transition from the fertilized egg (zygote) to more committed cell types (pluripotent cells, progenitors). Terminally differentiated cells lose the capability to self renew and to become a different cell type.

This model, exemplified by Waddington's "epigenetic landscape" (Waddington CH, The strategy of the genes, 1957), which endows a view of differentiation as an irreversible process of progressive specification, was challenged during the last 60 years by seminal work of various research groups which progressively demonstrated that it is indeed possible to: a) revert more differentiated cell states in less differentiated ones; b) transition from one differentiated cell state to another.

In 1950s, seminal work by Briggs and King (Briggs and King 1952) and

Gurdon (Gurdon et al 1958) pioneered the establishment of the technique of somatic cell nuclear transfer (SCNT). This technique allows the insertion in an enucleated egg of a somatic cell nucleus of choice to generate totipotent cells. By this approach they showed that albeit the nucleus was derived from a differentiated cell, it was still able to give rise to all cells that constitute an entire organism (i.e. it is genetically totipotent). This finding was further confirmed by cloning of mammals by the same approach (Wilmut et al 1997, Hochedlinger and Jaenisch 2002, Eggan et al. 2004).

These data suggested that it was indeed possible to erase/rewind the transcriptional program of terminally differentiated cells to establish a new landscape of expression that was typical of the zygote.

Further confirmation of this finding came later on since the establishment of embryonic stem cells (ESC) cultures from mouse (Evans and Kaufman 1981, Martin 1981) and human blastocysts (Thomson et al. 1998). These cells when fused to differentiated cells could reactivate the expression of pluripotency markers in the somatic nuclei (Tada et al. 2001, Cowan et al. 2005), highlighting the existence of factors that could drive the re-expression of genes associated with the pluripotent state.

Also, the finding that lineage-associated transcription factors could drive the "transdifferentiation" of cells into other cell types, with paradigmatic examples described for MyoD in driving the conversion of fibroblasts into myocytes (Davis et al. 1987) and C/EBPα in driving the conversion from lymphocytes to

macrophages (Xie et al. 2004, Laiosa et al. 2006), pointed to the reversibility of lineage determination.

5.3.1. Transcription-factor mediated reprogramming

In 2006 Takahashi and Yamanaka (Takahashi and Yamanaka 2006) proved that mouse embryonic and adult tail tip fibroblasts could be reprogrammed to the pluripotent state by ectopic expression of four transcription factors (namely Oct3/4, Sox2, Klf4 and c-Myc). By screening 21 factors in several combinations, infecting MEFs or tail-tip fibroblasts from a *Fbx15*$^{\beta geo/\beta geo}$ mice, with retroviruses (RV) encoding for these transcription factors, these cells could be converted to an ESC-like state and propagated in ESC culture conditions. Induced Pluripotent Stem Cells (iPSC) expressed pluripotency-related genes and could be differentiated into all three germ layers both *in vitro* and *in vivo*. Even if these cells could not result in viable chimeras upon injection into blastocysts and showed differences with blastocysts-derived ESC, findings that were later attributed to the choice of Fbx15 as pluripotency surrogate instead of a more strictly pluripotency associated gene, such as Nanog (Okita et al. 2007), it was the first demonstration that transcription factors can drive the conversion to the pluripotent state in differentiated cells.

In addition to this, this method for the first time allowed to overcome the limitations related to other techniques aimed at the induction of pluripotency.

Indeed, in the case of SCNT, both the technical difficulty of this technique and

the low efficiency related to the number of "premium-quality" oocytes to be used represent great hurdles for the large scale application of this technique. At the same time, generation of human ESC from blastocysts undergoes several implications from the bioethical point of view, and cannot always be used.

The advent of transcription factor-mediated reprogramming, that can be easily achieved by expressing a limited number of transcription factors in target cells, truly scaled up the possibility to derive pluripotent stem cells for disease-modeling, *in vitro* studies and future application for regenerative medicine. Since 2006 this approach has been translated to different species including humans (Takahashi et al. 2007, Yu et al. 2007, Park et al. 2008) and rhesus monkeys (Liu et al 2008), and to different target cells, such as neural stem/progenitor cells (Kim et al. 2008, Eminli et al. 2008), melanocytes (Utikal et al. 2009), mature lymphocytes (Hanna et al. 2008), adipocytes (Qu et al. 2012) and many others, showing that this process is universal and can be easily translated to the cell type of interest.

5.3.2. Optimization of the reprogramming procedure

Albeit it is possible to translate transcription factor-mediated reprogramming to a number of different cell types and species, the efficiency of this process for human cells is very low (≤0.01%). Since 2006, many efforts have been made in order to optimize the procedure.

5.3.1.1. Integration-based systems

The original protocol described by Yamanaka and colleagues was based on the retroviral delivery of the reprogramming factors. This procedure is affected by the following drawbacks: a) a suboptimal efficiency of delivery, which affects the efficiency of reprogramming (only cycling cells will be stably infected (Lewis and Emerman 1994), only a fraction of the cells will be infected by all vectors); b) multiple integrations in the genome of the cells, which might affect genome stability; c) an integration bias towards transcription start sites that might result in insertional mutagenesis (Mitchell et al 2004); d) a position-effect with variegated levels of expression (Yee and Zaia 2001); e) a stochastic silencing of the expression cassette due to the presence of CpG islands in the RV sequence, which can be transient or stable, influencing the differentiation outcome (Yee et al. 2001, Ramos-Mejia et al. 2012).

In order to reduce the number of integrations and obtain the stable co-expression of the four Yamanaka factors, mono-/bi-cistronic lentiviral vectors have been derived (Carey et al 2009, Sommer et al. 2009). In particular, the STEM-Cell Cassette (STEMCCA) described by Mostoslavsky's group, makes use of 2A self-cleaving peptides, that allow cleavage of a single fusion peptide to give rise to independent proteins (Donnelly et al 2001), and of an internal ribosome entry site (IRES) that allows translation from two separated cistrons, under the transcriptional control of a reverse tetracycline transactivator (rtTA); this favors sustained expression upon doxycycline administration. In this case,

a single copy of the vector was sufficient in order to drive reprogramming of MEFs, reducing the potential effects of insertional mutagenesis. Moreover, by removal of doxycycline, the vector is silent, reducing deleterious effects on differentiation.

An evolution of this cassette was described in 2010, when the STEMCCA vector was modified in order to be expressed from the Elongation Factor 1 alpha (EF1α) constitutive promoter and to carry LoxP sites in the LTRs in order to allow for excision of the single copy integrated provirus upon Cre recombinase administration (Somers et al 2010). In this case, vector-free iPSC can be generated, even though a single LoxP site is left in the genome as a result of Cre-mediated recombination.

The compromise between the generation of footprint-free iPSC and stable expression was reached by the use of PiggyBac transposons (Woltjen et al. 2009). This system makes use of transposases that insert/excise the Inverted Terminal Repeat (ITR)-flanked cassette in a "cut and paste" fashion, without altering the locus where the integration has occurred.

Still, this reprogramming/excision method requires long and tedious screenings for excised clones to be used for downstream application. In addition, the presence of multiple copies of the genome of repeat-flanked cassettes can cause intra/inter-chromosomal rearrangements upon recombinase delivery. Moreover, the transient-delivery of the excising enzyme (usually by plasmid transfection) requires a quick selection method for transfected cells, and might

occasionally result in the random integration of the plasmid by non-homologous end joining (NHEJ).

5.3.1.2. Non-integrative methods

In order to avoid integration *tout-court* during reprogramming many different approaches have been developed.

In 2008, Hochedlinger's lab showed that it was possible to reprogram mouse fibroblasts to pluripotent cells without vector integration using adenoviral vectors encoding the four Yamanaka factors, albeit with three fold magnitude lower efficiency (0.0001% compared to 0.1% with retroviral vectors)(Stadtfeld et al 2008). Since the transgenes were delivered as separated vectors and get rapidly diluted in cycling cells, this may account for such low efficiency.

The latter inconvenient was encountered also when transfecting monocistronic plasmids (Okita et al. 2008) or nonviral minicircle vectors (Jia et al. 2010)

Another approach relied on the use of Epstein Barr-derived episomal vectors (Yu et al. 2009). In this case seven factors in three individual plasmids were used (the four Yamanaka factors plus Lin28, Nanog and the SV40 Large T antigen) to reprogram human fibroblasts, even if at very low efficiency. Even though it is possible to isolate clones which diluted the episomes, these account for one third of all clones and screening is still required to isolate factors-free iPSC. Moreover, the immunogenicity of EBNA1 protein (Münz et al. 2000),

required for the replication of these constructs, might hamper the application of this approach to the regenerative medicine setting, weren't it be eliminated from the cells.

The very low efficiencies of these approaches based on DNA-based delivery were surpassed when Rossi's group published an innovative approach based on the transfection of synthetic modified mRNAs encoding for the four Yamanaka factors plus LIN28 (Warren et al. 2010). The 5-methylcytidine and pseudouridine modifications, in combination with the supplemented B18R protein, by suppressing most of the cellular interferonic response against exogenous RNA, increased the stability of the mRNA in the cell and cell survival. By daily transfections they obtained vector-free human iPSC with efficiencies up to 1.5% in less than 3 weeks. Similar results were obtained by the expression of microRNAs belonging to the 302/367 cluster from lentiviral vectors (Anokye-Danso et al. 2011), or of miR-200c together with miR-302 and -369 family by multiple transfections (Miyoshi et al. 2011).

Despite being the most efficient methods to date, they require daily transfections of mRNA/microRNA in order to achieve sufficient expression of the transcription factors to drive reprogramming.

A more recent approach relied on a single transfection of a self-replicative RNA derived from the Venezuelan Encephalitis Equine Virus RNA replicon (Yoshioka et al. 2013). In this case, efficiencies were variable (~0.01% with human adult fibroblasts and up to 1.9% in human newborn foreskin fibroblasts)

but in most cases the self-replicative construct could not be detected in iPSC.

5.3.2. Reprogramming of cancer cells to the pluripotent state

Given the strict interplay between genetic and epigenetic aberrations in cancer, there is a strong need for an approach that would be able to dissect epigenetic from genetic contribution to cancer pathogenesis. Moreover, current models lack a complete recapitulation of cancer phenotype. In particular, established cell lines and tumor xenograft models provide a much higher fit with advanced tumor status. More information on the early stages of tumor development can be provided by engineered mouse models. These systems, on the other hand, can only provide information on a limited amount of predefined effectors at the same time, lacking to recapitulate the intra-patient heterogeneity of primary tumor samples. As an additional level of complexity, primary samples cannot be propagated indefinitely in culture, so multilayered analysis on fresh samples is still a demanding issue.

iPSC could be a powerful tool to overcome these hurdles.

The reprogramming process entails a multistep epigenetic resetting (Papp and Plath 2013), that allows the establishment of a transcriptional landscape compatible with pluripotency. This process could be exploited to reset tumor-associated epigenetic marks, without disturbing the underlying genetic lesions. Cancer-derived iPSC (c-iPSC) would indefinitely expandable and could be differentiated in cells derived from all three germ layers. Since a plethora of

differentiation protocols towards cancer-relevant lineages is already available (Cheng et al. 2012, Ye et al. 2013, Sampaziotis et al. 2015), by the same approach it would be possible to study the contribution of genetic lesions to early stages of cancer development in different tumor settings. Moreover, the well-known phenomenon of "epigenetic memory" of the parental tissue, in this setting could become an opportunity to study the effects of retained tumor-associated epigenetic marks on early tumor pathogenesis (Figure 6)

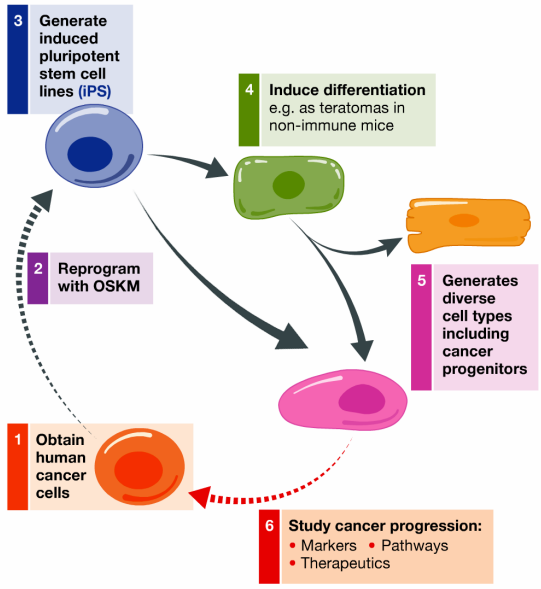

Figure 6. iPSC-based tumor modeling.

Schematic representation of cancer modeling through iPSC (adapted from Kim and Zaret 2015).

5.3.2.1. Evidence for c-iPSC as tools to model cancer pathogenesis

Seminal work by Jaenisch's lab in 2004 proved the feasibility of reprogramming tumor cells into pluripotent stem cells. By converting embryonal carcinoma cells (Blelloch et al 2004), and well-characterized mouse

cancer cell lines (Hochedlinger et al. 2004), the authors could show that:

- the pluripotent state can suppress temporarily the tumor phenotype, resulting in normal development of the blastocyst. The subsequent establishment of tumor-derived ES cell lines was tumor type-dependent;
- not all tumors could be reprogrammed and give rise to chimeras (only RAS-inducible, Ink4A-Arf$^{-/-}$ melanoma cells could give rise to ES cells and contribute to most tissues during differentiation);
- reprogrammed melanoma cells could contribute to development up to E9.5 in tetraploid complementation assays, thereafter the effect of mutations disrupted the normal development of the embryo;
- generated melanoma-iPSC chimeras were tumor prone and developed melanomas and rhabdomyosarcomas, that have overlapping pathways.

In contrast with more recent findings on the antagonistic role of WT p53 during reprogramming (Zhao et al 2008, Hong et al. 2009, Li et al. 2009), p53-mutated breast cancer cell lines did not give rise to ES lines. This can be attributed to the high degree of aneuploidy of the cell line (Kuperwasser et al. 2000).

Since primary cancer cells and some cancer cell lines could not be reprogrammed, many questions were still left open: it was unclear whether different mutations could affect the reprogramming procedure, or whether different cancer types could not be induced to the pluripotent state. In addition to this, still there was no demonstration of human tumors being reprogrammed. With the advent of iPSC, overcoming ethical issues on the use of oocytes and

technical complexity, it was possible to partially answer to these questions.

Carette and colleagues (Carette et al. 2010), showed that, upon reprogramming of human blast crisis chronic myelogenous leukemia (CML) dependent on the expression of the BCR-ABL fusion oncogene, these cells became resistant to treatment with imatinib, an allosteric inhibitor of BCR-ABL, despite the expression of the fusion protein. This resistance is lost upon differentiation of iPSC into hematopoietic lineages only, suggesting that cell lineage-associated expression can modulate the activity of oncogenes. Further characterization of this model (Kumano et al. 2012) showed that despite imatinib was active in these cells, as assessed by the reduced phosphorylation of STAT5 and CRKL down-regulation, still iPSC could compensate for this effect and survive the treatment. The same resistance occurs in more immature hematopoietic lineages but not in differentiated ones, a trait that the authors link to subpopulations of putative leukemic stem cells, responsible for treatment resistance in patients.

Miyoshi and colleagues applied the same approach to gastrointestinal tumor cell lines (Miyoshi et al., 2010), showing that upon reprogramming and differentiation, the tumorigenicity was abolished, an effect that the authors attributed to the re-activation of tumor suppressor genes, such as P16, and to changes in the transcriptional landscape during the process. These features can be attributed also to the differentiation process that was not directed towards specific progenitors of the gastrointestinal tract. This results in a

methylation/expression profile that is a mean of all cell types present in the bulk population of differentiating cells. Also, tumorigenicity *in vivo* could be hampered by differentiation in cell types that are non-permissive to tumor transformation in the genetic context of those cancer cells.

This last hypothesis was confirmed by more recent work in the context of glioblastoma cells (GBM) (Stricker et al. 2013). In 2 out of 14 human GBM samples, they could establish stable iPSC lines by using piggybac transposable vectors expressing OCT4 and KLF4, since SOX2 and MYC were already expressed. These iPSC showed loss of methylation upon reprogramming of tumor suppressors CDKN1C and TES. By differentiating iPSC into NS cells and transplanting the latter *in vivo*, they could not observe any obvious difference from the parental GBM in terms of tumorigenicity, despite the reversal of the de/methylation only of a minority of sites. Instead, when differentiating these cells towards the mesodermal lineage, specifically into chondrocytes, the transplantation of these cells *in vivo* showed suppression of the tumorigenicity, with scoring only of benign masses, if any. Thus, despite the presence of mutations that are known to be causal to GBM and other tumors, the tumorigenicity of the reprogrammed cells can only be revealed in a lineage-dependent manner.

The last piece of information to date that we can derive from reprogramming of cancer cells was provided by Zaret's group by reprogramming pancreatic cells to iPSC-like cells (Kim et al. 2013). Kim and colleagues, showed that

differentiation by teratoma formation assay of iPSC-like cells derived from pancreatic ductal adenocarcinoma (PDAC), resulted in the recapitulation of early lesions of this disease. Indeed, areas of the teratoma developed into early stage tumoral lesions, namely pancreatic intraepithelial neoplasia (PanIN). By harvesting teratoma lesions at later time points after transplantation resulted in the scoring of PDAC lesions, which had evolved from PanIN. Despite being tested with the only iPSC line carrying the original lesions of the tumor, named "iPSC-like" being dependent on the continuous expression of the exogenous reprogramming transgenes, this is the first demonstration that reprogramming tumors to a more undifferentiated state allows to study the evolution of primary tumor samples through differentiation.

Despite these seminal experiments that underlined the power of this approach, many questions are still left open. To date, cancer-derived iPSC have been characterized at the chromosomal level for genetic retention of tumor-associated mutations. Despite being a proof that these cells are truly derived from tumor cells and not from normal tumor-infiltrating cells, still this characterization was bound to known mutations already described in the literature or at its best to CNV detected through CGH arrays. Single nucleotide characterization of tumors and iPSC could give information on the genetic composition of tumor subclones. Moreover, since the reprogramming process has been shown to induce mutations and chromosomal rearrangements in a fraction of cases (Mayshar et al. 2010, Taapken et al. 2011, Pasi et al. 2011), this phenomenon should be controlled in order to exclude effects of newly

acquired mutations.

To this regard, the best reprogramming procedure would be endowed with the following characteristics:

- low or no impact on the genome and transcription of the cells: ideally integration-free or excisable integrating systems would be the best choice;
- oncogene-free: methods which allow to avoid the use of c-MYC and/or KLF4 would be preferable.

5.3.2.2. Genetic engineering of cancer-derived pluripotent cells

Undirected differentiation of pluripotent cells might not provide the best setting for the scoring of the tumorigenic outcome of reset cancer cells.

Teratoma formation assays are quick and easy to perform but require in depth analysis of the whole teratoma to score the phenotype of interest. Moreover, the effect of mutations on different lineages cannot be excluded.

In vitro differentiation of embryoid bodies can be considered an alternative, but methods for solid enrichment in developmental precursors of interest must be existing or be set up. Indeed, the presence of contaminant cells could hamper the molecular characterization of the differentiating progeny.

To this regard, in the absence of surface markers that could allow specific isolation of cells of interest by antibody-based technologies, gene editing could be useful to engineer iPSC to carry efficient selection cassettes.

The groundbreaking discovery of CRISPR/Cas9-based genome editing (Mali et al. 2013, Cong et al. 2013), building on previous approaches based on meganucleases (Smith et al 2006) zinc finger nucleases (Bibikova et al 2001, Hochemeyer et al 2009), TAL nucleases (Miller et al. 2011, Hochemeyer et al. 2011), allows seamless experimental design and targeted modification in pluripotent cells (Mali et al. 2013, Hou et al. 2013).

This technology takes advantage of an acquired immunity system in *Streptococcus pyogenes* used to recognize and degrade exogenous bacteriophage DNA. A guide RNA (crRNA) transcribed from the CRISPR (clustered regularly interspaced short palindromic repeats) locus, containing pieces of DNA from previous infections, is loaded with a second RNA (tracrRNA) within the Cas9 protein and "guided" by homology to the target DNA. Here the Cas9 protein delivers a double strand break (DSB) allowing for degradation of the exogenous molecule. This process has been exploited by combining tracrRNA and custrom crRNA into a single molecule that can be expressed from a plasmid and allows site specific delivery of a DSB. This can be repaired either by nonhomologous end joining (NHEJ), which is error prone and causes insertion or deletion of nucleotides during repair, or by homologous recombination in the presence of a donor DNA molecule. This molecule can contain a selection cassette flanked by homology regions to the intended site, allowing for engineering of virtually any sequence in the genome, as the guide RNA is responsible for specificity.

By this approach and the appropriate choice of promoters it would be possible to target a selection cassette (either with a fluorescent protein or an antibiotic resistance or both) that allows selection of the intended differentiating cells.

Another approach could be used to create isogenic cancer-derived iPSC in which the intended mutation is inserted or reverted. The idea would be to verify in the same genetic background the role of the gene of interest during tumor development.

5.4 Aim of the thesis

The aim of this work is to identify pathogenetic mechanisms underlying ovarian cancer, and in particular HGSOC, through the use of two complimentary approaches based on induced pluripotent stem cells. The first one is aimed at the dissection of the epigenetic vs. genetic contribution to this disease through the reprogramming-induced epigenetic resetting of primary tumors and differentiation of cancer-derived iPSC into disease-relevant lineages. The second one is aimed at the stratification of HGSOC samples, by identifying for each tumor its cell of origin, either the fimbrial epithelium or the ovarian surface epithelium, through a deeper molecular characterization of a unique cohort of primary tumors and normal samples. These two approaches would allow the precise identification of disease-relevant pathways, shedding light on the molecular mechanisms underlying this disease and leading to improved care and therapy of OC-affected patients.

6. MATERIALS AND METHODS

6.1. Tumor biopsy and normal samples dissociation and culturing

Human solid tumors, ascetic cells, fimbrial and ovarian biopsies were provided by the IEO biobank.

Tumor masses were cut into small pieces and resuspended in EDM solution (Table 2.1) and kept at 37°C.

Epithelial cells Digestion Medium (EDM)		
Reagent	**Volume**	**Final Concentration**
Stock for 1L of medium		
Ham's F12	500 mL	
DMEM	500 mL	
Glutamine	10 mL	1%
Penicillin/Streptomycin	10 mL	1%
Insulin	1 mL	1 µg/mL
Hydrocortisone	2 mL	0.2 µg/mL
Supplements for 100 mL of medium (to be used by 48 hours)		
EGF	10 µl	10 ng/mL
Collagenase IA	400 µl	200 U/mL
Hyaluronidase	1 mL	100 U/mL

Table 1. Epithelial digestion medium composition.

Every 30 minutes cells were resuspended until cell aggregates and single cells were released from the tissue. Cells were pelleted at 1500 rpm for 5 minutes and resuspended in ACK lysing solution (Lonza) for 2', in order to lyse red cells. Cells were then washed in PBS and centrifuged at 500g for 3 minutes and plated on Collagen I coated flasks (BD Biosciences) in EPI medium (Table 1).

Epithelial cells culturing medium (EPI)		
Reagent	**Volume**	**Final Concentration**
Stock for 1L of medium		
Ham's F12	500 mL	
DMEM	430 mL	
FBS (NA)	10 mL	1%
Glutamine	10 mL	1%
Penicillin/Streptomycin	10 mL	1%
Gentamycin	2 mL	0.2%
Amphotericin	2 mL	0.2%
Transferrin	10 mL	10 µg/mL
Insulin	1 mL	1 µg/mL
Hydrocortisone	10 mL	1 µg/mL
Hepes pH 7,5	10 mL	10 mM
Ascorbic Acid	1 mL	50 µM
Sodium Selenite	25 µl	15 nM
Etanolamine	6 µl	0.1 mM
Cholera Toxin	1 mL	50 ng/mL
Supplements for 50 mL of medium (to be used by 24 hours)		
EGF	5 µl	10 ng/mL
BPE	165 µl	35 µg/mL
T3	50 µl	10 nM
β-Estradiol	50 µl	10 nM

Table 2. Epithelial culturing medium composition.

In the case of ascites fluid, cells were pelleted at 500g for 3 minutes, red blood cells were lysed by ACK solution, and derived epithelial cells were plated as for solid tumors.

Fimbriae and Ovaries were incubated in Dispase 1 mg/mL for 30' at 37°C. Epithelial cells from the distal portion of the fimbria and the surface of the ovary were then scraped with a scalpel, pelleted at 500g for 3 minutes, red

blood cells were lysed by ACK solution, and derived epithelial cells were plated as for solid tumors.

Cells were passaged by trypsin 0.05% and expanded 1:2 or 1:3 according to confluency and growth rate of the cells.

6.2. Reprogramming by lentiviral vectors

The STEMCCA-OKSM 3rd generation lentiviral vector (LV) was provided by Gustavo Mostoslavsky (Somers et al 2010).

6.2.1. Vector production

Vector stocks were prepared by calcium phosphate transfection and concentrated by ultracentrifugation. Vectors were produced by transfection of human embryonic kidney 293T cell line (containing the mutant gene of SV40 Large T Antigen), because these cells are optimal DNA recipients in transfection procedures and the backbone of the vector constructs contains the SV40 origin of replication. 9x10^6 293T cells were seeded in 15 cm dishes and incubated in IMDM 10% FBS, Penicillin and Streptomycin (25U/mL each) for 24 hours before transfection. One hour before transfection medium was replaced.

To produce LVs, for each dish a plasmid DNA mix was prepared with 9 µg of pMD2- VSV-G, 12.5 µg of pCMVΔR9-D64V, 6.25 µg of pCMV-REV, 36 µg of STEMCCA transfer construct.

The plasmid solution was made up to a final volume of 1125 µL with 0.1X TE/dH2O (2:1) in a 15 mL polypropylene tube. Finally, 125 µl of 2.5 M CaCl2 were added and solution mixed. DNA precipitate was formed by dropwise addition of 1300 µl 2X HBS solution (281 mM NaCl, 100mM HEPES, 1.5 mM Na2HPO4, pH 7.12, 0.22 µM filtered) to the 1300 µl DNA-TE-CaCl2 mixture while vortexing at full speed and immediately added to 293T cells supernatant. Cells were incubated at 37°C. for other 14-16 hours after transfection and afterward medium was replaced with 16 mL of fresh or 1 mM Na butyrate (for donor vectors) containing medium. 30 hours after medium changing, supernatant was collected, filtered through 0.22 µm pore nitrocellulose filter and ultracentrifuged at 20000 rpm in SW32Ti rotor (Optima L-60 preparative Ultracentrifuge; Beckman) for 2 hours at RT. Pellets containing the vector were resuspended in a volume of sterile PBS representing 1/500 of the starting medium volume, pooled and rotate on a wheel at RT for 1 hour. The concentrated vector preparation was then divided into small aliquots (15 µl) and stored at -80°C.

6.2.2. Viral transduction of epithelial cells

DAC68, 12-Cit-1, 12_O_313, 12_O_1 and 12_O_333 were subjected to reprogramming by the STEMCCA LV. The evening before infection, target cells were plated at a density of $7.5x10^5$ cells per well of a six-well plate and incubated o/n at 37°C, 5% CO_2. The day of infection, 1 mL of freshly prepared

EPI medium was applied to cells and 15 µL of 500x concentrated STEMCCA virus were added in the presence of Polybrene at a final concentration of 8 µg/mL. Cells were incubated o/n and the medium was replaced the day after. At day 2 post-infection cells were harvested and replated on mytomycin-C inactivated MEF-coated 15 cm dishes. The seeding density of MEFs was of 5 millions of cells per 15 cm dish. At day 5 post-infection EPI medium was replaced by hiPSC medium (Table 2), which was daily added to cells until iPSC-like colonies appeared.

hiPSC medium			
Product	**Cat. Nr.**	**Producer**	**Final mix (500 ml)**
D-MEM/F-12 (1X) liquid 1:1	21331-046	Life Technologies	400 mL
Knockout™ Serum Replacement	10828-028	Life Technologies	100 mL
Sodium Pyruvate MEM 100 mM, liquid	11360-039	Life Technologies	5 mL
MEM Non Essential Amino Acids (100X), liquid without L-Glutamine.	11140-035	Life Technologies	5 mL
Penicillin-Streptomycin, liquid	15140-122	Life Technologies	5 mL
L-Glutamine 200 mM (100X), liquid	25030-024	Life Technologies	5 mL
2-Mercaptoethanol, 50 mM (1000X)	31350-010	Life Technologies	0,2 mL
FGF-basic, AA 10-155 Recombinant (final concentration 10ng/ml)	PHG0021	Life Technologies	0,05 mL

Table 3. Human iPSC medium composition.

6.3. Reprogramming by mRNA/miRNA

Samples 12_O_313 and 12_O_333 were reprogrammed using mRNA/miRNA Reprogramming Kit (Stemgent). Briefly, $5x10^6$ million newborn foreskin

fibroblast cells (NuFF) (Stemgent) were plated onto a T-75 flask with 25 mL of Pluriton medium (Stemgent) and cultured over 8 days. Conditioned medium was collected daily and used as reprogramming medium for daily transfections. Next, target cells were plated at two different densities (5-7.5x10^5 per 3.5 cm well) onto Matrigel (Corning)-coated plates (diluted 1:40 in DMEM/F12 1:1). Medium supplemented with B18R, an inhibitor of IFN-alpha cell response, was supplied to cells two hours prior to transfection. Cells were transfected with miRNAs on day 1 and 5 and daily since day 2 with the mRNA cocktail (OCT4, SOX2, KLF4, c-MYC, LIN28 and SOX2) along with nuclear GFP mRNA (Stemgent) for 12 (12_O_333 sample) to 15 days (12_O_313 sample). iPSC-like colonies were mechanically picked and expanded.

6.4. iPSC culturing

Early passage iPSC were cultured on hESC-qualified Matrigel (Corning) with mTESR1 medium (StemCell Technologies) and passaged when at ~70% confluency. The passaging procedure required pre-treatment of cells for 2 minutes with Dispase at 37°C, mechanical picking of undifferentiated cell clumps under a microscope and transfer to a new Matrigel-coated plate in mTESR1 medium. After 4-5 passages, when the population of iPSC was pure, cells were enzymatically passaged. Briefly, medium was removed, cells were washed in DPBS 1x and incubated at 37°C with Accutase for 3-4 minutes. Cells were single-cell dissociated, harvested in a conical tube with mTESR medium and pelleted at 125g for 3 minutes. After removal of the supernatant, cells were

resuspended in mTESR1 medium supplemented with 10µM Rock inhibitor (Sigma-Aldrich) and plated onto Matrigel-coated plates.

6.5. Staining of iPSC

6.5.1 Immunofluorescence

iPSC were dissociated by Accutase, plated on Matrigel coated coverslips and cultured in mTESR1 until the desired dimension of colonies.

Coverslips were incubated in paraformaldehyde 4% for 20 min on ice, washed three times in DPBS, blocked in 10% FBS, 0.1% Triton X-100 in DPBS for 30 min at room temperature.

Staining was performed in DPBS supplemented with 10% FBS with the following antibodies: anti-OCT3/4 (Santa Cruz, sc-5279, 1:250 dilution) and anti-NANOG (Everest Biotech, EB06860, 1:100 dilution). Coverslips were stained o/n, washed three times in PBS and the secondary conjugated antibody was added in PBS, 10% FBS. After 1 hour of incubation, coverslips were washed three times with PBS and mounted with Vectashield mounting medium with DAPI (H-1500) on slides. After an o/n, slides were visualized through a widefield microscope.

6.5.1. TRA-1-60 live staining

TRA-1-60 Dylight 488 was diluted in fresh medium at 2.5 µg/mL concentration and added to live iPSC and left to bind for 30’ at 37°C. Two washes with cell

culture medium were performed and cells imaged with an EVOS FL (Thermo Fisher).

6.5.2 Alkaline Phosphatase Staining

A solution made of Citrate, Acetone and Formaldehyde was used to fix cells for 45 seconds after a wash in DPBS. Cells were rinsed in ddH_2O for 30 seconds and the staining solution (Sodium Nitrite:FRV alkaline solution 1:1, Sigma) was added to cells for 30 minutes. Cells were rinsed in ddH2O and imaged for the presence of a red precipitate.

6.6. Teratoma assay

Approximately 4 million iPSC per sample were injected subcutaneously in the flank of NOD SCID IL2RGnull mice. Teratomas were taken from the mice, washed rapidly in PBS and fixed o/n in 4% formalin solution. The day after they were placed in histological cassettes and processed by an automatized tissue processor. After processing, the samples were included in paraffin blocks and cut with a Leica microtome in 4 µm thick sections that were attached on glass slides and stained with haematoxilin and eosin to proceed to the identification of cell lineages derived from ectoderm, mesoderm and endoderm and the degree of differentiation.

6.7. Immunohistochemistry on FFPE sections

The following antibodies were used to stain tumor or teratomas sections:

Target protein	Cat. Nr.	Producer	Dilution	Antigen Retrieval
S100	Z0311	Dako	1:600	none
Desmin	M0760	Dako	1:20	EDTA 0.25mM pH 8, 30 min
Pan-CK	M0821	Dako	1:400	Proteinase K, 5 min
NCAM	sc-7326	Santa Cruz	1:250	Tris-Sodium Citrate 10mM pH 6, 30 min
WT1	M3561	Dako	1:200	EDTA 0.25mM pH 8, 30 min
GFAP	Z0334	Dako	1:400	none
PAX8	10336-1-AP	Proteintech	1:500	Tris-Sodium Citrate 10mM pH 6, 30 min
OCT3/4	2750	Cell Signaling	1:200	Tris-Sodium Citrate 10mM pH 6, 30 min
NANOG	4903	Cell Signaling	1:800	Tris-Sodium Citrate 10mM pH 6, 30 min

Table 4. List of antibodies used in IHC.

Paraffin embedded sections were cut as previously described. Sections were washed twice in Bioclear (Bio Optica) for 15 minutes each to remove paraffin and hydrated in a descending alcoholic scale with 99%, 95% and 70% ethanol (the 99% ethanol step was repeated twice, 5 minutes for each step) and then washed in ddH2O. Depending on the antigen different unmasking were performed (Table 2.4); endogenous peroxidase was inhibited with 3% hydrogen peroxide (H_2O_2) for 5 minutes. Sections were then incubated in blocking buffer (2% BSA in TBST) for 20 minutes; primary antibodies were diluted in blocking buffer and incubated for 1 hour at room temperature. After washing, sections were incubated with the secondary antibodies for 30 minutes

at RT. Counterstaining was performed with hematoxylin (10 s); sections were then de-hydrated in 95%, 99% ethanol and Bioclear, 10 minutes each) and glasses were mounted using Eukitt (Bio Optica). Signals were revealed using Dako EnVision+ Kit.

6.8. Isolation of RNA from cells in culture

12_O_313 and 12_O_333 cells were processed with the RNeasy mini kit (Qiagen). Cells were harvested by trypsin 0.05% and pelleted at 500 g in DPBS. Dry cell pellets were resuspended in RLT buffer (a highly denaturing buffer containing guanidine isothiocyanate) supplemented with 143 mM beta-mercaptoethanol to inactivate RNases and stored at -80°C until the day of extraction. Lysed cells were homogenized by passing through an insulin syringe and and ethanol added to the lysate to allow loading onto the purification column. A step of digestion of DNA was carried out by the addition on column of RNAse-free DNAse (Qiagen). Two rounds of washing eliminated all contaminants, while the RNA was eluted in RNase-free water (Qiagen). This procedure allows to isolate RNA molecules bigger than 200 nucleotides, so that smaller ribosomal RNAs are depleted.

6.9. Isolation of gDNA from cells in culture

Cells were harvested by trypsin 0.05% (tumors and normal fimbria and ose) or by Accutase (iPSC) and pelleted at 500 g or 125g, respectively, in DPBS. Dry cell pellets were stored at -80°C until the day of extraction.

Using Qiagen DNeasy® blood and tissue kit, cell pellets were lysed in proteinase K and RNA was depleted by RNAse (Qiagen). The lysate was loaded onto the column. DNA is bound onto the column while contaminants passed through. Remaining contaminants and enzyme inhibitors were removed by two wash steps, DNA was eluted in ddH_2O and stored at 4°C until use.

6.10. Differentiation of iPSC in mesodermal progenitors

6.10.1. Embryoid Bodies

3000 hESCs per well were aggregated by centrifugation to form spin EBs in serum-free APEL medium. Cells were treated or not with 50 ng/mL BMP4 and 20 ng/mL ACTIVIN A (both from R&D Systems) for 6 days, when FACS analysis for GFP reactivation was performed.

6.10.1. Adhesion culture

iPSC clone B4 was harvested by Accutase, pelleted in mTESR at 125g for 3', replated in mTESR complemented with 10µM Rock inhibitor on Matrigel-coated plates at a density of $5x10^4/cm^2$. On the following day Stemdiff Mesoderm Induction Medium (StemCell Technologies) was applied to cells.

At day 4, FACS analysis for the expression of Brachyury T and GFP was performed.

6.11. Generation of gene targeting constructs

6.11.1. Donor construct

The donor construct was designed with the Benchling software (https://benchling.com) and synthesized by GeneArt (Thermo Fisher).

6.11.2. Guide RNAs constructs

gRNAs were designed on the MIXL1 locus by using the online tool from Zhang's lab (crispr.mit.edu).

gRNA sense 1: 5'-

TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGAGCGC

ACGGGACTCGGCTG - 3'

gRNA antisense 1: 5'-

GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCAGCCG

AGTCCCGTGCGCTC -3'

gRNA sense 2:

5'-

TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGTGCG

CTCCAGTTTGCCGA -3'

gRNA antisense 2: 5'-

GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCGGCA

AACTGGAGCGCACC -3'

gRNA sense and antisense for each couple were cloned in the gRNA cloning vector (Addgene #41824) by the Gibson assembly cloning kit (NEB), according to manufacturer's instructions.

6.11.3. Cas9 and Cas9D10A constructs

Cas9-GFP (#44719) and Cas9D10A-GFP (#44720) were obtained from Addgene.

6.12. Gene targeting experiments

6.12.1. Electroporation of iPSC

iPSC were grown in 15 cm-dishes. For the electroporation setup, cells were grown up to 70-80% confluency (10^7 cells approximately), pre-treated with 10 uM Rock Inhibitor Y-27632 (Sigma-Aldrich, Y0503) for 4 hours, harvested by Accutase, resuspended in either ice cold PBS or Gene Pulser Electroporation Buffer (Biorad), and transfected with 50 or 100 µg of pCAS_GFP plasmid at 250V and 500 µF. Cells were then resuspended in mTESR medium supplemented with 10 µM Rock Inhibitor and plated on Matrigel coated 15 cm plates. Cells were analyzed for GFP-expression 48 hours post transfection.

For the gene targeting experiment, PBS was used with the same electroporation conditions. The combination of plasmids used in condition A, B, C is depicted in Table 5.

Plasmid	Condition A	Condition B	Condition C
gRNA 1	25 µg	X	12.5 µg
gRNA 2	X	25 µg	12.5 µg
Cas9-GFP	25 µg	25 µg	X
Cas9D10A-GFP	X	X	25 µg
Donor construct	50 µg	50 µg	50 µg

Table 5. Transfection conditions for the gene targeting experiments.

Cells were then resuspended in mTESR medium supplemented with 10 µM Rock Inhibitor and plated on Matrigel coated 15 cm plates. G418 (200µg/mL) was used from 48 hours to 15 days after transfection select resistant clones, that were picked and expanded for analysis.

6.12.2. Polymerase Chain Reaction (PCR) on the 5' of the integration

PCR reactions were carried out in a total volume of 25 µL and 200 ng of gDNA were used. For the Master Mix, 12.5 µL of Amplitaq Gold 360 Master mix (with an hot-start *Taq* polymerase), 0.5 µM of each primer, and 10% 360 GC-buffer. GC-enhancer was used as the target amplicon was 63% GC rich. The reaction conditions used were the following: 10 minutes of initial denaturation at 95°C, 35 cycles consisting of 15 seconds denaturation at 95°C, 30 seconds of annealing at 56°C, 3 minutes extension at 72°C, and after the last cycle a final 7 minutes extension at 72°C. PCR products were run on 1% agarose gels

containing 0.5 µg/mL of ethidium bromide in Tris-Acetate-EDTA (TAE) buffer. The expected size of the band was 2739 bp.

Primer FW: 5'-GGTATCTTTATTGGTGGGCC-3' (mapping outside the homology arm).

Primer RV: 5'-TATGTTTCAGGTTCAGGGGG-3' (mapping on the SV40 poly A).

6.12.4. Digestion of the PCR amplicon

PCR products were purified with the QiaQuick PCR Purification kit (Qiagen). Samples were eluted in 21 µL of ddH_2O and processed for digestion.

For digestion 1 µL of BamHI and NaeI restriction enzyme (NEB) were applied to the eluted DNA together with 2.5 µL of Cutsmart Buffer. The reaction was incubated at 37°C for 4 hours and run on a 3% agarose gel in TAE buffer. Staining post-run was performed by incubating the gel with a 0.5 µg/mL ethidium bromide/TAE solution for 30 minutes.

6.13. FACS analysis

Tumors, iPSC and differentiating iPSC were harvested as described for passaging.

For surface antigens, FcR blocking reagent was added and cells incubated for 15' at 4°C. Directly conjugated antibodies were added according to manufacturer's instructions in 100 µL total volume and allowed to bind for 45'

at 4°C in the dark. Cells were then washed with staining solution at 250 g for 3' and resuspended in 100 µL of the same buffer for analysis.

For intracellular staining, cells were fixed and permeabilized by resuspending them in 100 µL of Fix/Perm solution (BD) and incubating for 20 minutes on ice. Cells were washed twice with 1x Perm/Wash solution (BD) and incubated with the appropriate amount of antibody according to manufacturer's instructions in 100 µL total volume, for 30 minutes on ice in the dark. Cells were washed again two times in Perm/Wash solution and resuspended in 200 µL of the same solution for analysis.

Cells were run through either the BD FACSCantoII or the BD influx machines for acquisition and data were analysed with Flowjo Vx (Treestar).

Target	Fluorochrome	Code	Supplier	Dilution
CD44	PerCP-Cy5.5	560531	BD Biosciences	1:20
CD133	APC	293C3	MACS	1:10
Brachyury T	APC	IC2085A	R&D Systems	1:10
TRA-1-60	Alexa-Fluor 647	560850	BD Biosciences	1:20

Table 6. List of the antibodies used for FACS analysis.

6.14. High-throughput experiments

6.14.1. DNA methylation analysis

gDNA extraction was performed as described in paragraph 2.9. Processing of the samples by the Illumina 450k Beadchip kit was performed by our collaborator Gilles Gasparoni in Joern Walter's lab. Briefly, starting from

gDNA bisulfite conversion was carried out with the EZ DNA methylation kit (Zymo research) according to Illumina's recommended incubations. Converted DNA is subjected to a pre-amplification step, fragmentation of DNA and hybridization to a 450K BeadChip, that contains allele specific probes that recognize either an uracil (that is generated by the conversion of unmethylated cytosines) or a cytosine (when cytosines are methylated they are not converted by bisulphite treatment). By single base extension fluorescent nuclotides are inserted and signals were imaged with a Hiscan system (Illumina).

6.14.2. Bioinformatic analysis of DNA methylation

Raw data were analyzed by the RnBeads R package (Assenov et al. 2014). Briefly, the package performs a quality control of the signals in the chip, removes probes with a detection p-value < 0.05 using the Greedycut algorithm, normalizes the signal by using the SWAN normalization method, generates beta-value of methylation for each probe and summarizes them as being part of promoters, entire genes or CpG islands. For unsupervised clustering, I considered the top 1000 variable sites, using correlation-based distances for hierarchical clustering. For differential methylation analysis, I used the limma method and I used the combined ranking to select the top promoters for clustering. The combined ranking is generated by the pipeline taking into account: i) the difference in mean methylation levels of the two groups being compared, b) the quotient in mean methylation and c) the FDR-adjusted p-

value. For each site the worst of those three rankings (i.e. the highest) is established as the combined rank of the comparison.

6.14.3. Whole exome sequencing analysis

gDNAs extracted from tumors, iPSC and blood, were processed by the sequencing facility of the IFOM/IEO Campus which performed the following part of the protocol; the starting amount of DNA was 10 ng. The overhangs of the DNA fragments were converted into phosphorylated blunt ends, using T4 DNA polymerase, E. coli DNA polymerase I large fragment (Klenow polymerase), and T4 polynucleotide kinase. The 3' to 5' exonuclease activity of these enzymes removed 3' overhangs and the polymerase activity filled in the 5' overhangs. A single ‘A’ nucleotide was added to the to the 3' end of the blunt phosphorylated DNA fragments, using the polymerase activity of Klenow fragment (3' to 5' exo minus). This prepared the DNA fragments for ligation to the adapters, which have a single ‘T’ base overhang at their 3' end. Adapters were ligated to the ends of the DNA fragments, preparing them to be hybridized to a flow cell. DNA was run on a TAE 2% agarose gel to remove excess adaptors and selects a size range of templates; a gel slice containing the material in the 300±50 bp range was cut from the gel and purified with QIAquick Gel Extraction Kit (Qiagen) according to manufacturer instructions. Processed DNA was subjected to exon enrichment by the TruSeq exome enrichment kit (Illumina) according to manufacturer’s instructions. Briefly, gDNA was incubated with capture probes of exonic regions. Then streptavidin beads were

used to purify the captured regions and a second round of enrichment was performed. Finally, the selected adapter-modified DNA fragments were enriched by PCR amplification. The exon-enriched DNA library was diluted to 16 pM and used for cluster generation and sequencing on an Illumina HiSeq machine to obtain 70 million (tumors) or 35 million (iPSC and blood) paired-end reads, 100 bp length.

6.14.4. Bioinformatic analysis of whole exome sequencing

Quality control of the raw reads was performed using the Fastqc tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Each lane of sequencing data underwent alignment to the hg19 assembly using BWA (Burrows-Wheeler Aligner) algorithm (Li et al. 2009) resulting in sorted sequence alignment/mapping file (SAM) format that was converted to binary format (BAM) using SAMtools (Li et al. 2009).

For SNV variant detection, high confidence somatic variant calling was performed on the GATK 2.3.4 post-filtered and processed BAM files using two callers i) VarScan2 (Koboldt et al. 2012) with its default setting for the reads coverage while the p-value threshold set to 0.05 and ii) Mutect (Cibulskis et al. 2013) with default coverage. Only high confidence somatic variants were considered identified by both the methods where no evidence in the matched germline sample was included.

For somatic copy number variation (CNV) analysis, CNV calling was

performed with the Control-FREEC tool (Boeva et al. 2011) with window size 500 and step size 250. Control frequency signals were used as reference to call CNV in tumor and iPSC samples.

For correlation analysis, the correlation heatmap was built based on standard deviation and taking in account of the top most dispersed somatic mutations (called with GATK) in tumors, iPSC and blood. The top 8000 variants were considered with a mutation frequency threshold of 0.4.

6.14.5. Chromatin immunoprecipitation coupled to deep-sequencing (ChIP-seq)

Cells were cross-linked with 1% formaldehyde in DPBS for 10 minutes at RT. Fixation was stopped by quenching by the addition of 0.125 M glycine for 5 minutes at RT. Cells were washed two times in DPBS and harvested in SDS buffer (100 mM NaCl, 50 mM Tris HCl pH 8.0, 5 mM EDTA pH 8, 0.2% NaN_3, 0.5% SDS) with protease inhibitors (2µg/mL aprotinin, 5 µg/mL leupeptin). This lysate was stored at -80°C until the day of sonication.

The lysate was thawed in a beaker full of water and centrifuged at 400 g for 6 minutes. The lysate was resuspended in 130 µl of SDS Buffer:Triton Dilution buffer 2:1 (Triton Dilution Buffer: 100 mM Tris-HCl, pH 8.5, 5mM EDTA, pH 8.0, 0.2% NaN_3, 5% Triton X-100), and sonicated by Covaris S220 focused ultrasonicator in an AFA 130 µL microcuvette (conditions: 105 peak power, 5.0 duty factor, 200 cycles/burst, 3 minutes) to obtain a sonicated chromatin in the order of ~250 bp. The sonicated chromatin was centrifuged at maximum

speed for 30 minutes at 4°C, and the protein content was quantified by Bradford assay using BSA (NEB) to derive a standard curve. 2 µL of sonicated chromatin were diluted in 800 µL of water plus 200 µL of Biorad Protein Assay and the absorbance at 595 nm was measured by a spectrophotometer. 50 to 60 µg were used for each IP and diluted in 1 mL of SDS Buffer:Triton Dilution buffer 2:1. From this solution, 10 µL were taken to be used as total control (1% input).

1 µg of primary antibody was added to the chromatin and incubated overnight at 4°C on a rotating wheel. ChIP complexes were collected by incubating them with 35 µL of protein G dynabeads (Thermo Fisher) (pre-equilibrated with SDS Buffer:Triton Dilution buffer 2:1) for two hours at 4°C on a rotating wheel. ChIP-bead complexes were put on a Dynamag magnet (Thermo Fisher), the supernatant was removed and beads were washed three times with ice-cold 150 mM Wash Buffer (1% Triton X-100, 0.1% SDS, 150 mM NaCl, 2mM EDTA pH 8.0, 20mM Tris-HCl pH 8.0) and one time with ice-cold 500 mM Wash Buffer (1% Triton X-100, 0.1% SDS, 500 mM NaCl, 2mM EDTA pH 8.0, 20mM Tris-HCl pH 8.0). ChIP-bead complexes and the 1% input were incubated in 120 µL of 1% SDS, 0.1M $NaHCO_3$ at > 1300 rpm at 65°C overnight to reverse crosslink between proteins and DNA. DNA was purified with the QiaQuick PCR purification kit (as in paragraph 6.12.4) and eluted in 43 µL of ddH_2O.

For the sequencing, samples were quantified by Qubit 2.0 fluorometer and given to the sequencing facility of the IFOM/IEO Campus which performed

the following part of the protocol, as in paragraph 6.14.7.; the starting amount of DNA was 10 ng. Diluted libraries were used for cluster generation and sequencing on a HiSeq 2000 instrument (Illumina) following manufacturer's protocol. ChIPseq was performed in single end, 50 bp with coverage of 30x for all the IPs (H3K4me1, H3K4me3, H3K27ac) while the H3K27me3 and inputs were sequenced at 60x.

Histone mark	Code	Producer	Quantity used
H3K27me3	9733	Cell Signaling	1 µg/IP
H3K27ac	AB4729	Abcam	
H3K4me1	AB8895	Abcam	
H3K4me3	AB8580	Abcam	

Table 7. List of the antibodies used in ChIP experiments.

6.14.6. Bioinformatic analyses for ChIP-seq

The ChIP-seq data quality was checked for quality control with FASTQC tool and the alignment of the samples to the hg19 reference genome were made with BowTie (Langmead et al. 2009. The peak calling was performed with MACS 2.0.9 (Zhang et al. 2008) with p-value set as 10^{-5} and other default parameters. The usual peaks output of MACS2 for H3K4me1, H3K4me3 and H3K27ac were taken into account while broad peak output files were considered for H3K27me3. The annotation for the output peak files was done with the HOMER tool (Heinz 2010).

6.14.7. RNA sequencing (RNAseq)

RNA Samples were processed with the TruSeq Stranded Total RNA Library Prep Kit (Illumina). The starting amount of RNA was 1 µg per sample, as quantified by Agilent RNA 600 Nano kit (RNA integrity number: 0.9-1).

After as step of bead-mediated ribosomal RNA depletion (rRNA removal beads, ribo-zero kit), the RNA was also fragmented using divalent cations under elevated temperature and primed for cDNA synthesis with random hexamers. The primed and cleaved RNA fragments were reverse into first strand cDNA using reverse transcriptase and random primers. The RNA template was removed and a replacement strand was synthesized with DNA Polymerase I to generate double-strand (ds) cDNA. Beads (AMPure XP beads) are used to separate the ds cDNA from the second strand reaction mix. Overhangs resulting from fragmentation were converted into blunt ends using an End Repair Mix: the 3' to 5' exonuclease activity of this mix removes the 3' overhangs and the polymerase activity fills in the 5' overhangs. A single ‘A’ nucleotide was added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single ‘T’ nucleotide on the 3' end of the adapter provided a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera (concatenated template) formation. Multiple indexing adapters were ligated to the ends of the ds cDNA, preparing them for hybridization onto a flow cell. PCR was used to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the

library. Fragments with only one or no adapters on their ends are by-products of inefficiencies in the ligation reaction. Neither species can be used to make clusters, as fragments without any adapters cannot hybridize to surface-bound primers in the flow cell, and fragments with an adapter on only one end can hybridize to surface bound primers but cannot form clusters. The PCR was performed with a PCR primer cocktail that anneals to the ends of the adapters. The sequencing was performed with an Illumina HiSeq 2000, with paired end 50 bp reads to achieve a coverage of 35x.

6.14.8. Bioinformatic analysis of RNAseq

The Salmon tool (Zhang et al. 2015) was used to perform transcript-level quantification that has a streaming inference method. It is a lightweight based algorithm that allows mappings of reads to transcript positions without performing a base-to-base alignment of the read to the transcript. The transcriptome index was build on hg19.

7. RESULTS

7.1 Reprogramming of Ovarian Cancer Cells

Since the discovery of the transcription factor-induced reprogramming, many different cell types have been used to generate iPSCs. Anyway, it is still unknown whether all cells derived from tumor tissues can be reprogrammed to pluripotency and if, among all patient-derived tumors of the same type, all mutations can be compatible with the pluripotent state.

Lentiviral vectors encoding the "Yamanaka factors", namely OCT4, KLF4, SOX2 and c-MYC, have been extensively used to achieve reprogramming. Nevertheless, genetic instability is a hallmark of tumor cells and it is therefore important to avoid high copy number of integrations that might affect the tumor phenotype by insertional mutagenesis.

To assess whether ovarian cancer cells can be amenable to reprogramming with reduced impact on the genome, I made use of the well established human stem cell cassette-containing lentiviral vector (STEMCCA LV) (Somers et al. 2010). This vector allows reprogramming of fibroblasts from a single copy integrated in the genome and expresses all four reprogramming factors from a polycistronic vector. Moreover, the presence of LoxP sites in the LTR of the LV allows for excision of the reprogramming transgenes by Cre-mediated recombination (Figure 7).

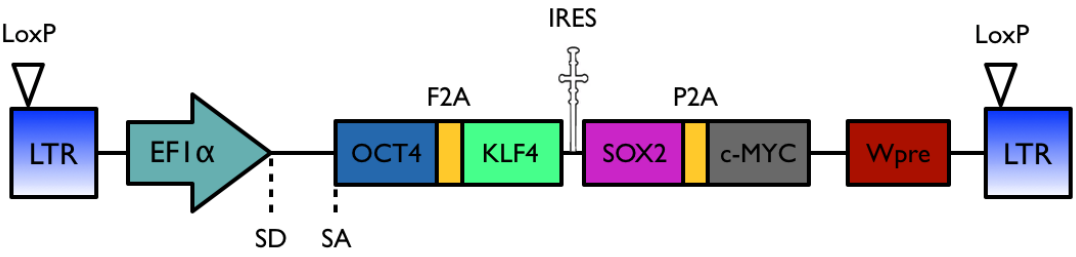

Figure 7. Schematic representation of the STEMCCA vector.

LTR: Long Terminal Repeat; EF1α: Elongation Factor 1 Alpha promoter; SD: Splice Donor; SA: Splice Acceptor; OCT4: Octamer-binding Transcription factor 4; KLF4: Kruppel-Like Factor 4; SOX2: Sex determining region-box 2; Wpre: Woodchuck hepatitis virus Post transcriptional Regulatory Element; F2A: foot-and-mouth disease virus 2A peptide; P2A: porcine teschovirus 2A peptide; IRES: Internal Ribosome Entry Site.

I subjected to reprogramming a set of samples comprising normal cells from the fimbria, tumor cells derived from ascites, and three primary samples of OC characterized by different histopathology (Table 8). Following the protocol illustrated in Figure 8a, I was able to successfully reprogram the sample derived from the fimbria (DAC68) and the sample derived from ascites (12-Cit-1). These cells stained positive for Alkaline Phosphatase (AlkPh) and TRA-1-60 (Figure 8b), *bona fide* markers of pluripotency.

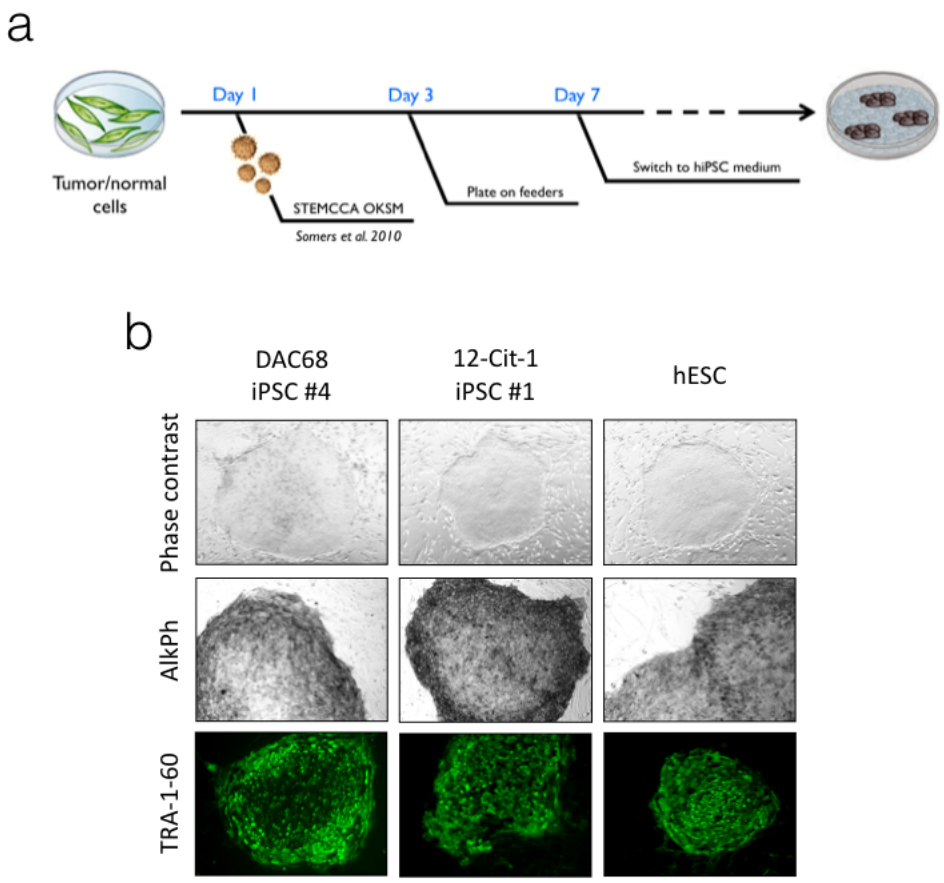

Figure 8. Generation of iPSC from gynecological samples.

a. Schematic representation of the reprogramming procedure; b. Representative iPSC colonies compared to hESC in terms of morphology (top panels, phase contrast pictures) and expression of pluripotency markers such as Alkaline phosphatase (middle panels) and TRA-1-60 (low panels).

Solid tumors-derived cells were refractory to reprogramming and the overall efficiency was very low, with > 30 days necessary to obtain iPSC-like colonies (Table 8)

Sample	Hystopathology	# of stable lines	Overall Efficiency
DAC68	fimbrial epithelium (normal)	5	0.005%
12-Cit-1	ascites (from HGSOC)	2	0.002%
12-O-313	undifferentiated OC	0	0%
12-O-1	HGSOC	0	0%
12-O-333	low grade SOC	0	0%

Table 8. Overall efficiency of LV-driven reprogramming.

Increasing the multiplicity of infection (MOI) of the vector would increase the reprogramming efficiency but would also increase the risk of insertional mutagenesis, therefore I decided to switch the reprogramming platform to a non-integrative system. Our lab has recently set up a new method based on the daily transfection of mRNA and miRNA to reprogram patient's fibroblasts to the pluripotent state (Adamo, Atashpaz, Germain et al., 2015). By this approach it is possible to obtain vector-free, feeder-free iPSC in less than 15 days without integration of the transgenes. Moreover, contrary to the standard culturing of iPSC on feeder cells (usually Mytomicin-C inactivated fibroblasts necessary for providing nutrients and growth factors to the cells), this protocol allows the direct generation of feeder-free pluripotent cells, avoiding cell type contamination in downstream assays (e.g. next generation sequencing).

I decided to reprogram two samples for which I was not able to obtain any iPSC colony despite two months of post-infection culturing, namely 12-O-313 (grade

4 undifferentiated OC, hereafter 313) and 12-O-333 (low grade SOC, hereafter 333).

By this approach I was able to generate stable iPSC colonies from both tumor samples with increased efficiency, and to confirm the expression of defined pluripotency markers (Figure 9).

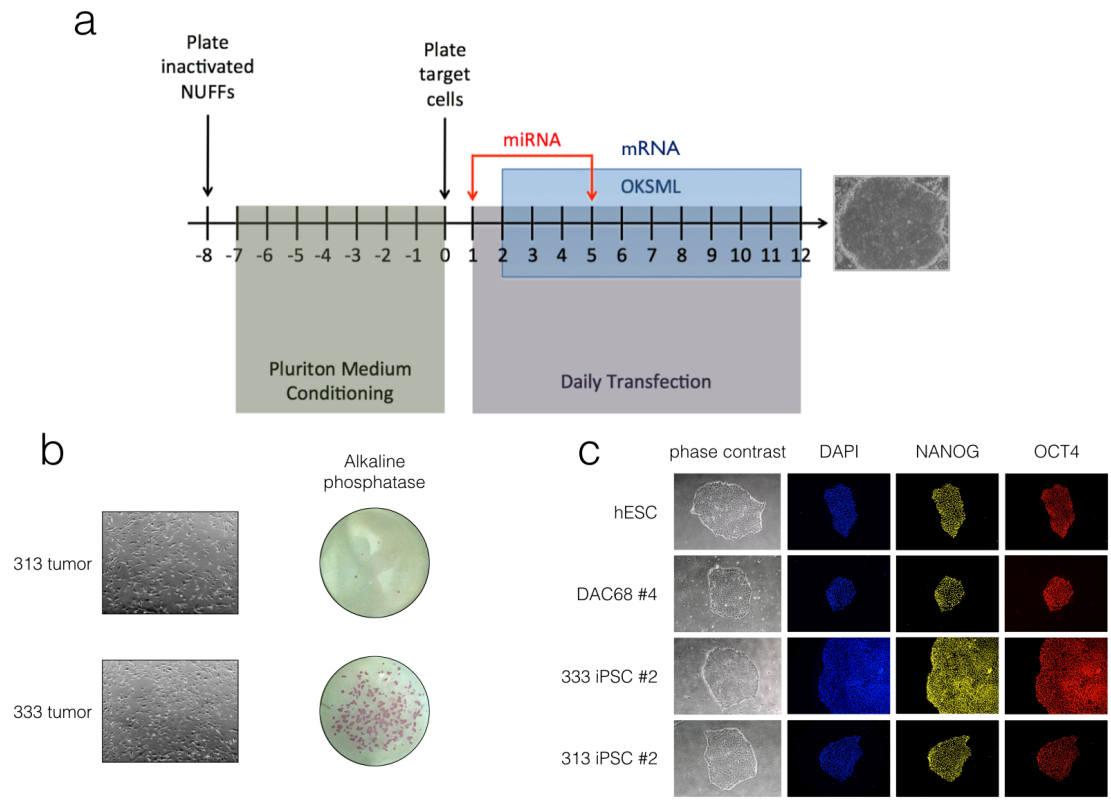

Figure 9. Generation of vector-free iPSC from OC.

A. Schematic representation of the reprogramming procedure by mRNA/miRNA transfection. NUFFs: newborn foreskin fibroblasts; OKSML: OCT4, KLF4, SOX2, c-MYC, LIN28. B. Phase contrast pictures of tumor cells (left panels) at the moment of transfection and alkaline phosphatase staining (right panels) at the end of the procedure, after picking of iPSC-like cells. C. Representative immunofluorescence analysis of OC-derived iPSC. hESC and fimbria derived iPSC (DAC68 #4) were used as positive controls. Nuclei were counterstained with DAPI.

This experiment shows that, at least in this limited set of samples, it is possible to reprogram OC cells to iPSCs, albeit with different efficiencies (313 tumor: 0.01%; 333 tumor: 0.4%).

To understand whether this difference could be ascribed to the presence of cells expressing OCT4 or NANOG already present in the tumor, I performed immunohistochemistry staining on the formalin-fixed paraffin-embedded (FFPE) samples. These two tumors show low or absent expression of either of these factors (Figure 10).

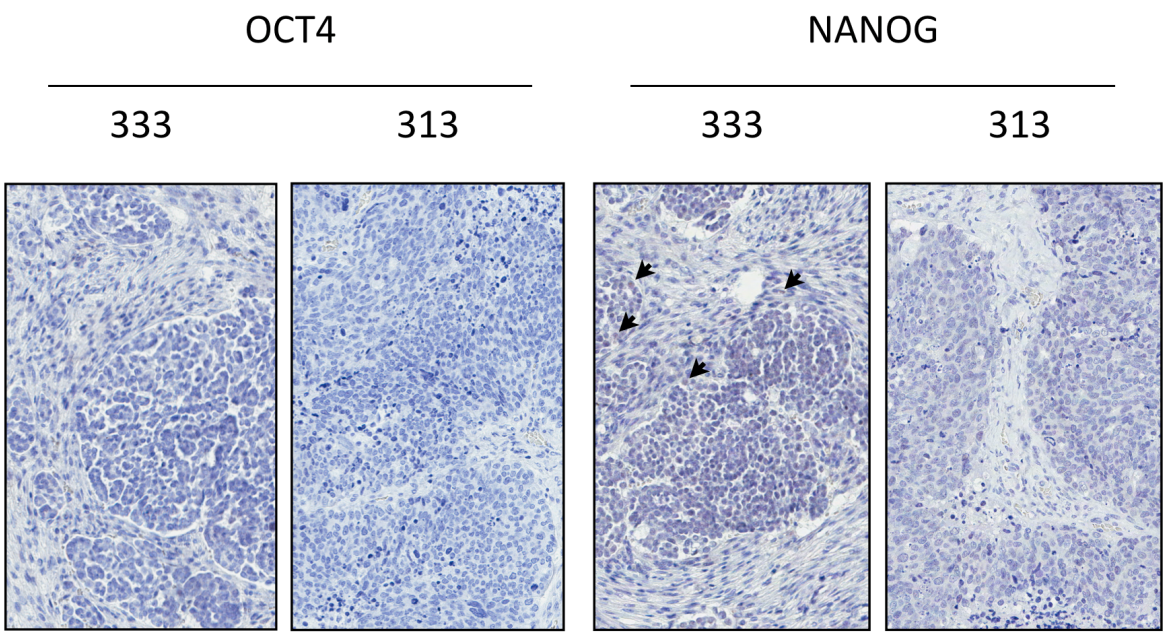

Figure 10. Parental tumors do not show expression of OCT4 and NANOG.

IHC staining for OCT4 and NANOG in 313 (undifferentiated OC) and 333 (low grade SOC) FFPE sections. Haematoxylin is used as counterstaining. Black arrows indicate weakly positive cells.

In addition, in order to exclude that the higher efficiency was due to the presence of putative cancer stem cells in the starting culture (i.e., prior to

reprogramming), I performed FACS analysis for CD44 and CD133, two well described markers of putative stem cells in ovarian cancer (Kryczek et al 2012, Zhang et al 2008).

As assessed by FACS analysis, there was no expression of these markers in the two samples, proving that the putative stem cell compartment, as defined previously, was not responsible for the differences in reprogramming efficiency (Figure 11).

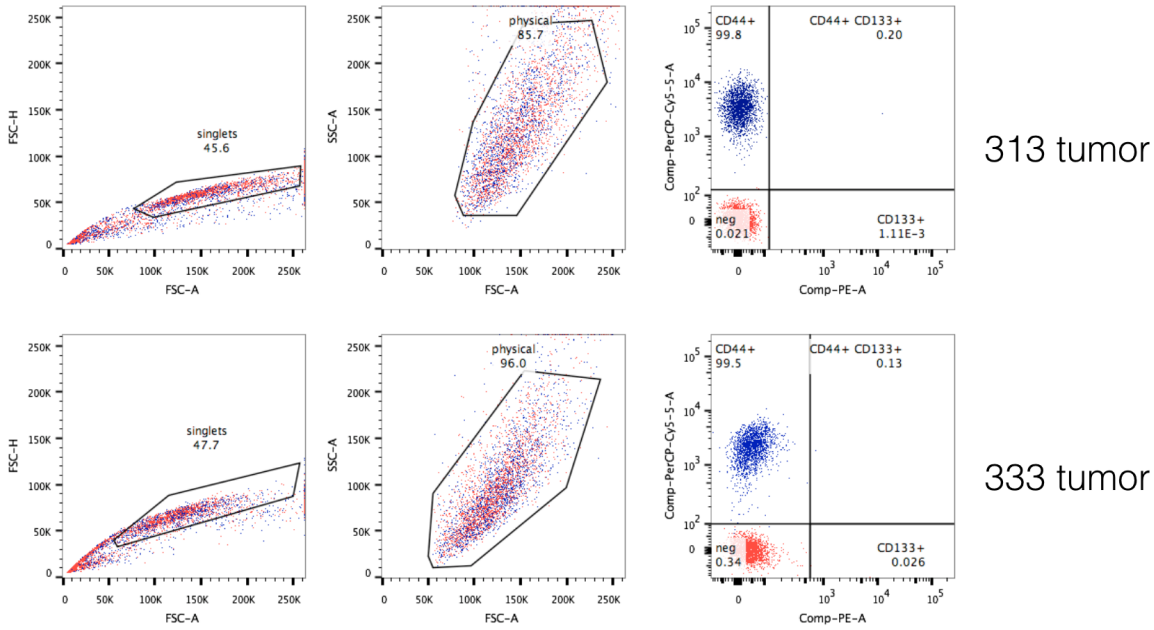

Figure 11. Parental tumors show comparable expression of putative stem cell markers. FACS analysis for the expression of CD44 and CD133 in 313 and 333 tumor cells. Left panels: gating for single dissociated cells; Middle panels: selection of living cells based on physical parameters. Right panels: Staining for CD44 (y-axis) and CD133 (x-axis). Red dots: unstained control; Blue dots: stained sample.

Also, by RNAseq I assessed the expression stem cell-related genes were not expressed in the original cell culture established for the two samples (Figure 12)

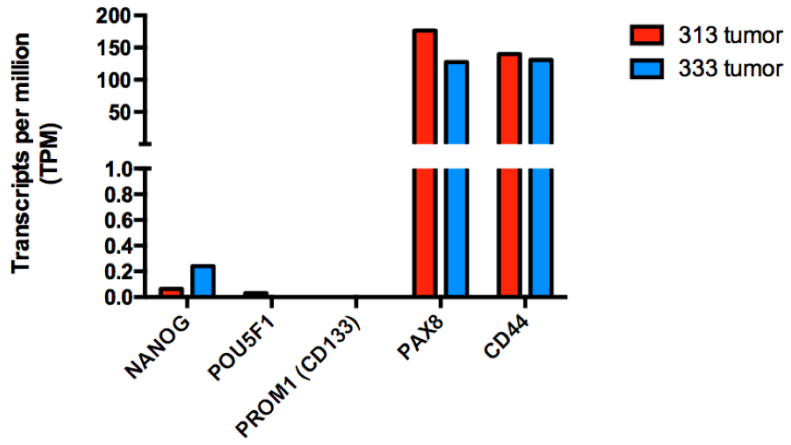

Figure 12. 313 and 333 tumors do not express pluripotency related genes.

Expression levels of stem cell-related genes in 313 and 333. CD44 and PAX8 were included as positive controls (highly expressed genes).

I went on verifying that these cells were reprogrammed to pluripotency at the DNA methylation level. I obtained genome wide DNA methylation profiles of these two tumor samples and the fimbria together with their respective iPSCs. Unsupervised hierarchical clustering based on all normalized methylation values shows that, regardless of the parental cells from which they were generated, iPSCs cluster away from parental samples, indicating consistent DNA methylation re-setting (Figure 13a). By including previously published DNA methylation profiles of hESC lines as a reference (Ziller et al., 2011),

OC-derived iPSCs cluster together with hESC, as shown both by unsupervised clustering of the top 1000 variable sites (Figure 13a) and by overall Pearson and Spearman correlation (Figure 13b). Moreover, by looking at POU5F1 and NANOG promoter regions, OC-derived iPSCs show consistent de-methylation of these regions, with a pattern that closely mirrors that of hESCs (Figure 13c), indicating functional reprogramming of these cells.

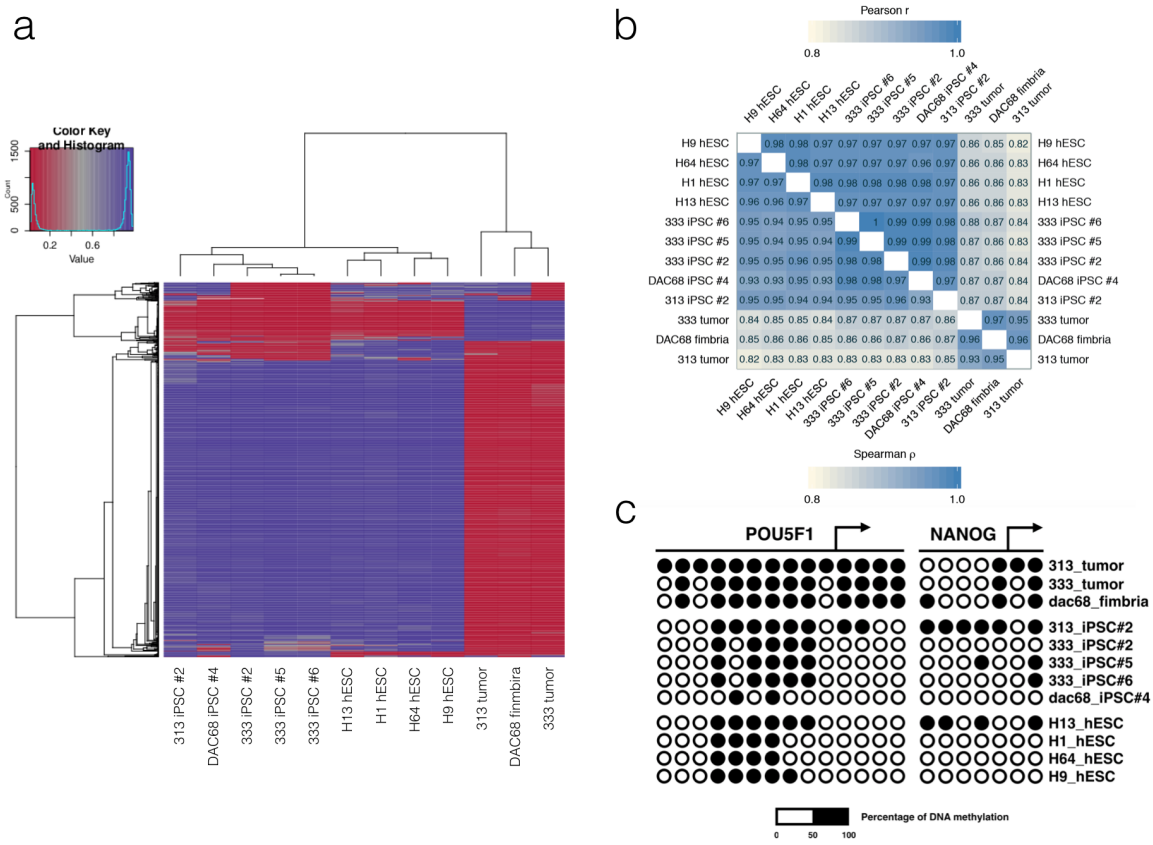

Figure 13. iPSC are closely similar to human ESC.

a. Hierarchical clustering of the top 1000 variable sites of hESC, iPSC and parental tissues from which they were derived; b. Pearson and Spearman correlation half matrices for the samples in a.; c. Methylation profile of POU5F1 (OCT4) and NANOG promoters in parental cells, iPSC and hESC. Black dots: CpG with a methylation beta value higher than 0.5; white dots: CpG with a methylation beta value lower than 0.5.

7.2 Teratoma formation assay

In order to prove the pluripotency of these cells, I subjected iPSCs to *in vivo* teratoma formation assay. By subcutaneous injection of iPSC/ESC, it is possible to observe the outgrowth of a benign mass that is composed of differentiated cells belonging to ectoderm, mesoderm and endoderm and thereby proving that these cells are equivalent to inner cell mass cells.

In this setting, this approach can be used to evaluate the contribution of somatic mutations to tumor transformation. The occurrence of a teratocarcinoma, characterized by transformed cells within the mass, would be an indication of a major role played by genetic mutations regardless of epigenetic resetting.

I injected subcutaneously two independent iPSCs clones derived from the 333 tumor (namely clone #2 and clone #6) into NOD SCID IL2RGnull mice (NSG) and isolated teratomas from mice. Strikingly, while clone #6 gave rise to a fully differentiated teratoma, without malignant cells throughout the whole mass, clone #2 gave rise to both terminally differentiated cells from all three germ layers and to malignant areas. Further characterization of this tumor by IHC and histological analysis revealed that it was a primitive neuroectodermal tumor, a medulloepithelioma, characterized by the expression of cytoplasmic WT1, focal GFAP and S100, and NCAM. In addition, the malignant areas expressed PAX8, that was not described as a neuroectodermal marker, but that was already present in the parental tumor (Figure 14).

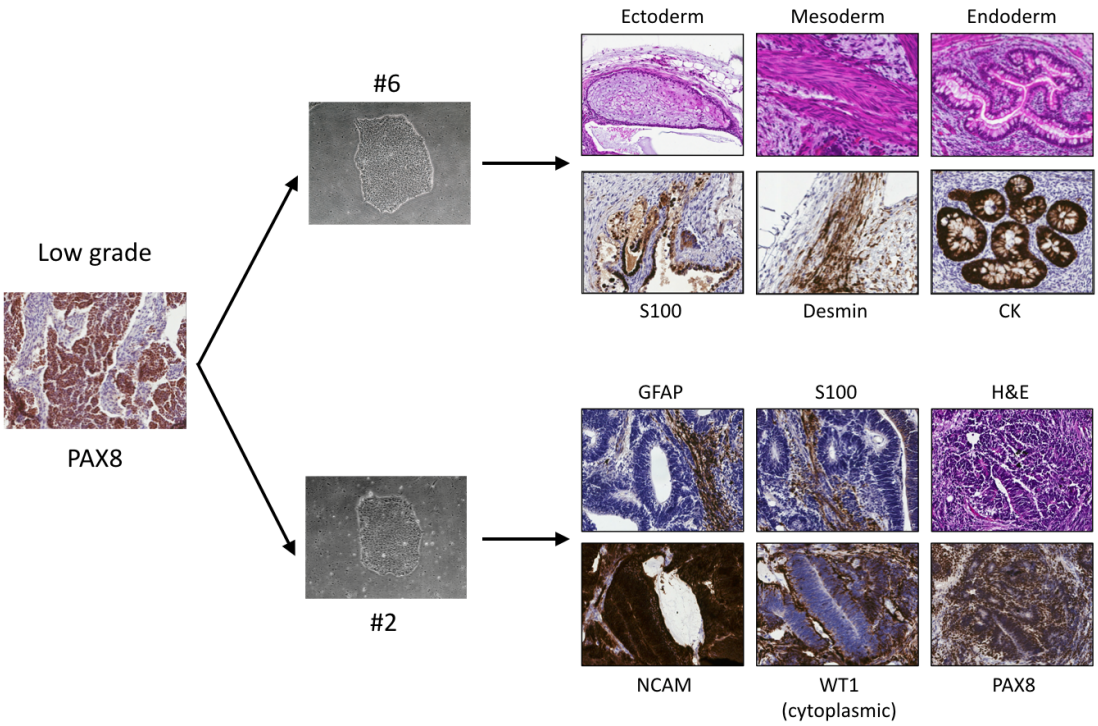

Figure 14. Differential tumorigenic potential in vivo of 333-derived iPSC.

IHC analysis of teratomas derived from two independent iPSC clones derived from the 333 low grade SOC. Left panel: PAX8 staining of the 333 tumor. Middle panels: phase contrast pictures of iPSC clones in culture. Top right panels: trilineage differentiation of iPSC, as defined by histological analysis (top panels) and marker expression (lower panels). Bottom right panels: depiction of neuroectodermal transformation as assessed by histological analysis (H&E) and expression of defining markers (lower panels).

7.3. Genetic analysis of OC-iPSC

The different efficiency of reprogramming for these two tumor samples could be explained also by the presence of contaminant cells in the culture at the moment of reprogramming. Despite the presence of cholera toxin in the epithelial culturing medium, that boosts the growth of epithelial cells while inhibiting fibroblasts proliferation, the fraction of mesenchymal tumor-

associated cells can vary from patient to patient. These cells could be reprogrammed more easily due to the absence of underlying mutations that could hamper the reprogramming process. This occurrence might also reflect the different outcome of *in vivo* differentiation, as clone #2 might be derived from a tumoral cell and therefore give rise to malignant areas within the teratoma, while clone #6 might be derived from a tumor-associated normal cell. To address this point, together with our bioinformaticians, we performed Whole Exome Sequencing analysis on parental tumors and two iPSC for each tumor. Genomic DNA extracted from the peripheral blood cells of the patients was used as a reference.

7.3.1. Single nucleotide variants (SNV) analysis

We used Varscan 2 (Koboldt et al., 2012) in order to call mutations over the human genome reference hg19 assembly in our samples. As a first level of analysis, we used the frequency of SNVs in each sample to compute a correlation value to check if iPSC were more similar to tumors or to normal samples. As shown in Figure 3.9, each iPSC showed a higher correlation with its respective tumor (Figure 15), indicating that iPSC were more closely related to tumor samples.

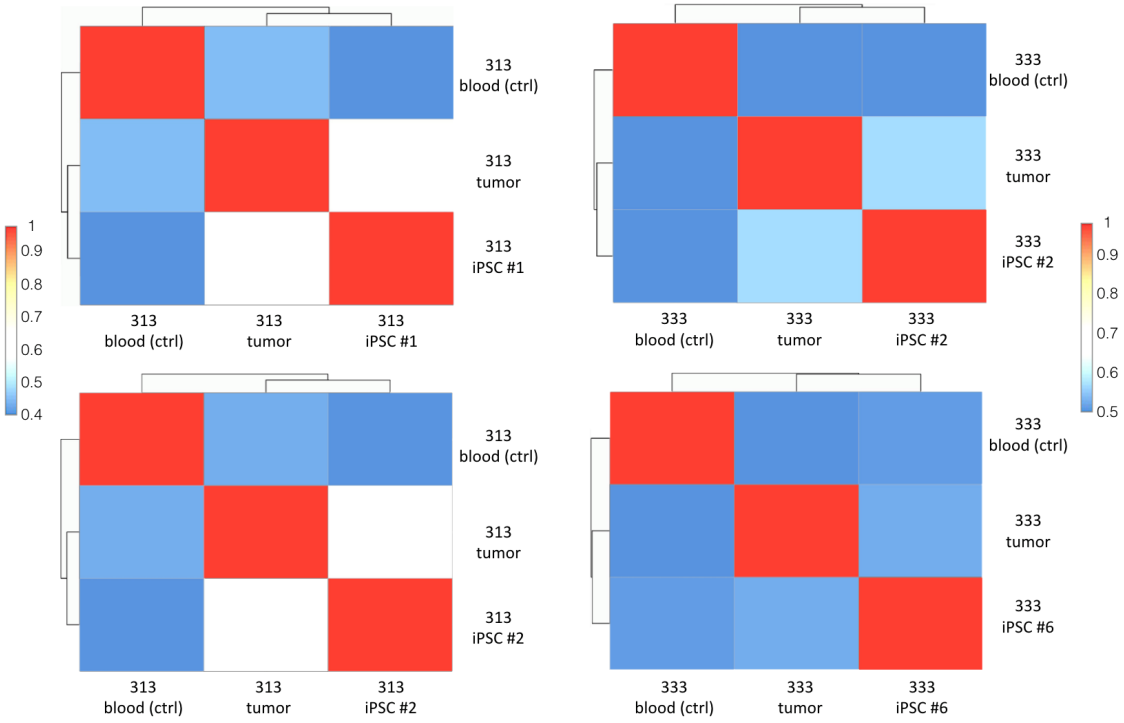

Figure 15. iPSC mutations frequency is more correlated with the one in tumors.

Correlation analysis of the SNVs frequency in tumors, iPSC and blood. Mutation frequency threshold: 0.4.

By filtering out mutations that were present in tumors and iPSC but also in the blood samples (germline mutations), we annotated somatic tumor and iPSC SNV. We additionally used the MuTect pipeline (Cibulskis et al. 2013) in order to derive high confidence mutations, i.e. those identified by both platforms pipelines. We defined the overlap between mutations called with each tool, highlighting a consistent difference in mutation calling for the two algorithms. Therefore, we decided to use only the common mutations to verify the reprogramming of tumor cells.

We intersected the somatic genetic signature of the parental tumors with the one derived from matched iPSC and showed that at least a fraction of the

mutations found in the tumors were retained in iPSC, suggesting the reprogramming of a tumoral subclone rather than a normal tumor-associated cell (Figure 16).

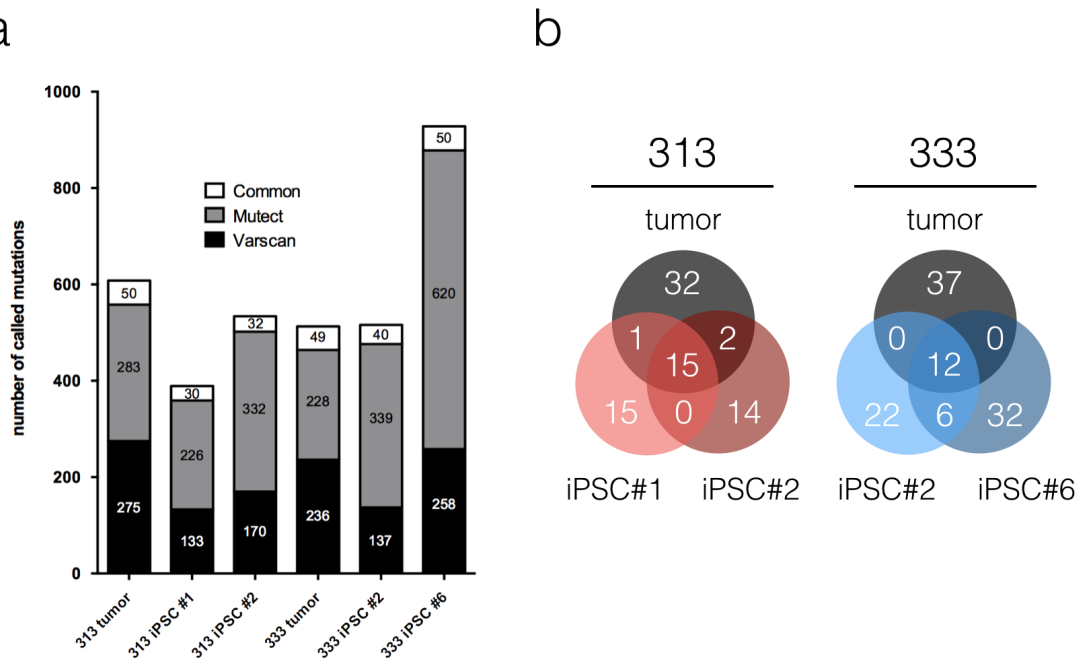

Figure 16. iPSC share a fraction of the parental somatic mutations.

Somatic mutation analysis of parental tumors and their derived iPSC. A. Histogram showing the number of mutations called for each sample by VarScan2 (black bars), MuTect (grey bars), and common between the two algorithms. B. Overlap analysis of high confidence mutations present in tumors and derived iPSC.

7.3.3 Copy number variations (CNVs) analysis

As an additional proof that tumor cells were the ones that underwent the reprogramming process, we decided to use the Control-FREEC tool (Boeva et al. 2011) on sequencing data, to analyze the content of copy number variations (CNVs) in tumors and derived iPSC. Sequencing data coming from peripheral bloog gDNA were used as reference to assess gain or losses of genomic content

(Figure 17) in the other samples.

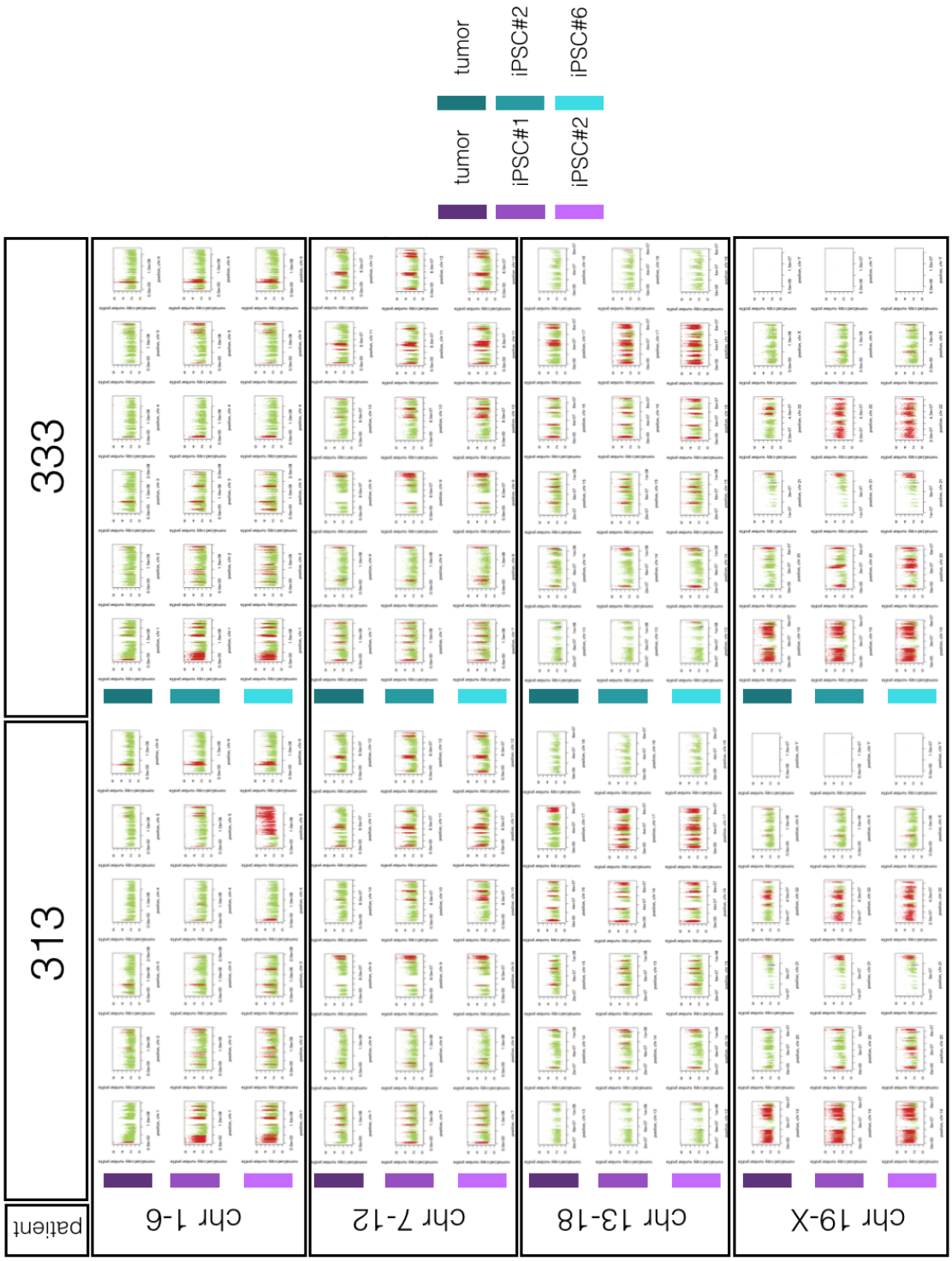

Figure 17. Schematic representation of CNVs in tumors and tumor-derived iPSC. Sequencing reads from blood-derived genomic DNA were used as a reference. Red signal: amplifications of genomic regions; Blue signal: deletions of genomic regions; Green signal: non-altered regions.

The vast majority of CNVs are represented by amplification of genomic areas, with a minor proportion of DNA loss in both tumor samples. By comparing tumors and their iPSC we found that roughly 50% of the CNVs detected in 313 and slightly more than 70% of the CNVs detected in 333 are present in their respective iPSC with 100% identity between the parental tumor and its derived iPSCs. Moreover, more than half of shared CNVs are common between the two iPSC clones derived from the same tumor, suggesting that they are shared between different tumor subclones. An estimation of the copy number of CNVs reveals that: i) increase in signal is an indication that clonal iPSC are derived from tumor subclones; ii) constant intensity of signal indicates that those CNVs are common to the vast majority of cells in the tumor (Figure 18).

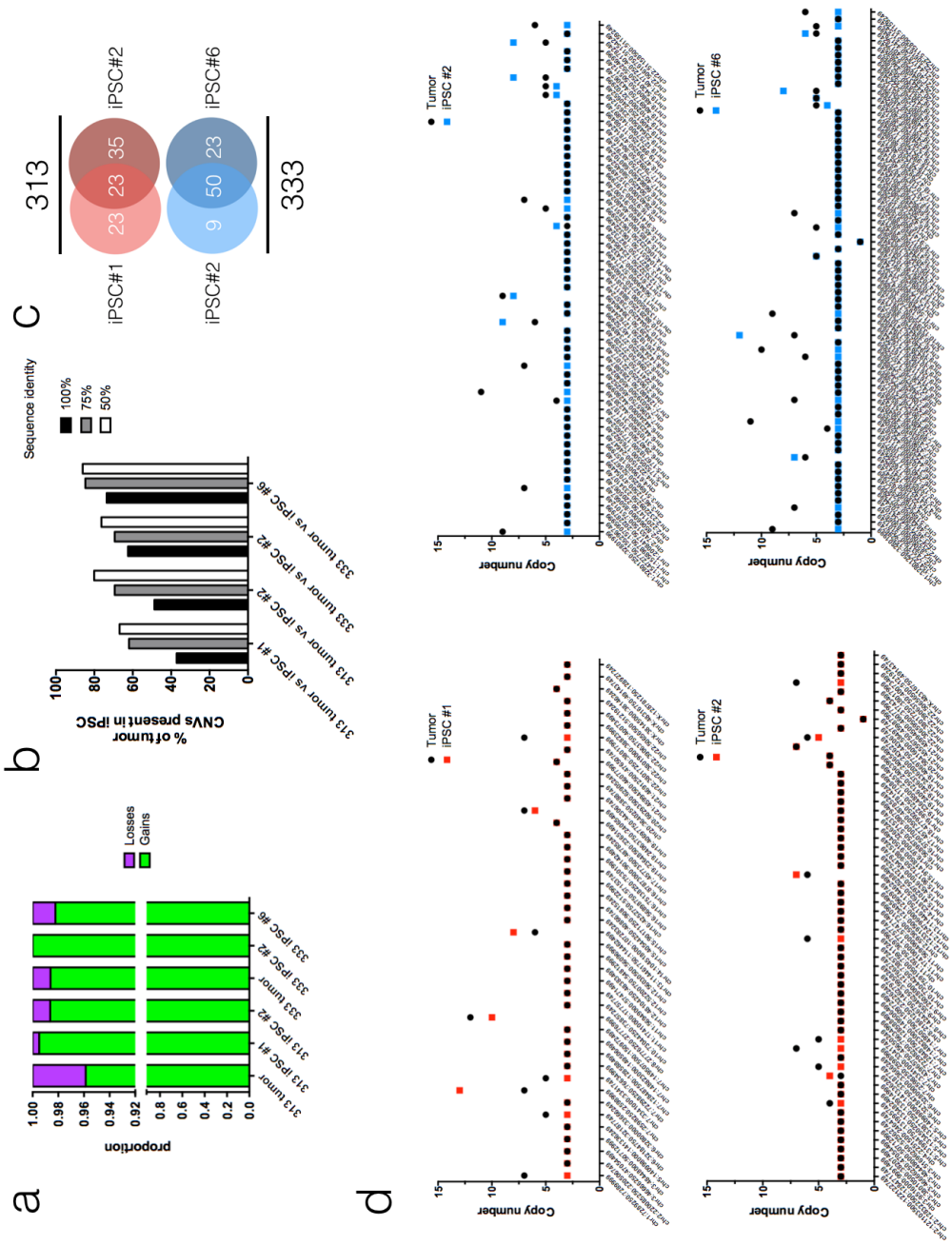

Figure 18. CNVs present in tumors are retained in iPSC.

A. Proportion of gains (green bars) and losses (purple bars) of genetic sequences in the considered samples. B. Percentage of CNVs retained from tumors into iPSC with different matching tolerance. C. Overlap analysis of retained CNVs in iPSC that are common between the two considered clones for each tumor. D. Predicted copy number of CNVs shared by iPSC and their respective tumors in the samples considered. Left panels: 313 background; Right panels: 333 background.

Taken together, these data indicate, both at the SNVs and CNVs level, that our iPSC were derived from tumor cells and not from normal contaminating cells in culture.

7.4. Directed differentiation of pluripotent cells into mesodermal MIXL1+ cells

In order to study the effect of the epigenetic resetting driven by somatic cell reprogramming on tumor cells, I sought to set up a protocol for the differentiation of iPSC into mesodermal progenitors, common to both fimbria (FI) and ovarian surface epithelium (OSE), the two putative cells of origin of ovarian cancer. For the set up I made use of a MIXL1-GFP hESC line, kindly provided by Andrew Elefanty's laboratory. Being the MIXL1 gene expressed during primitive mesoderm differentiation, by this approach it is possible to isolate early mesodermal progenitor cells to be used for the differentiation *in vivo* of pluripotent cells towards the female reproductive tract epithelium (Ye et al. 2011). The protocol relies on the generation of embryoid bodies from ESC and their differentiation towards the mesodermal lineage by administration of BMP4 and ACTIVIN-A. I subjected the MIXL1-GFP hESC line to differentiation and evaluated the expression of GFP by FACS analysis. Up to 47% of cells scored positive for GFP (Figure 19).

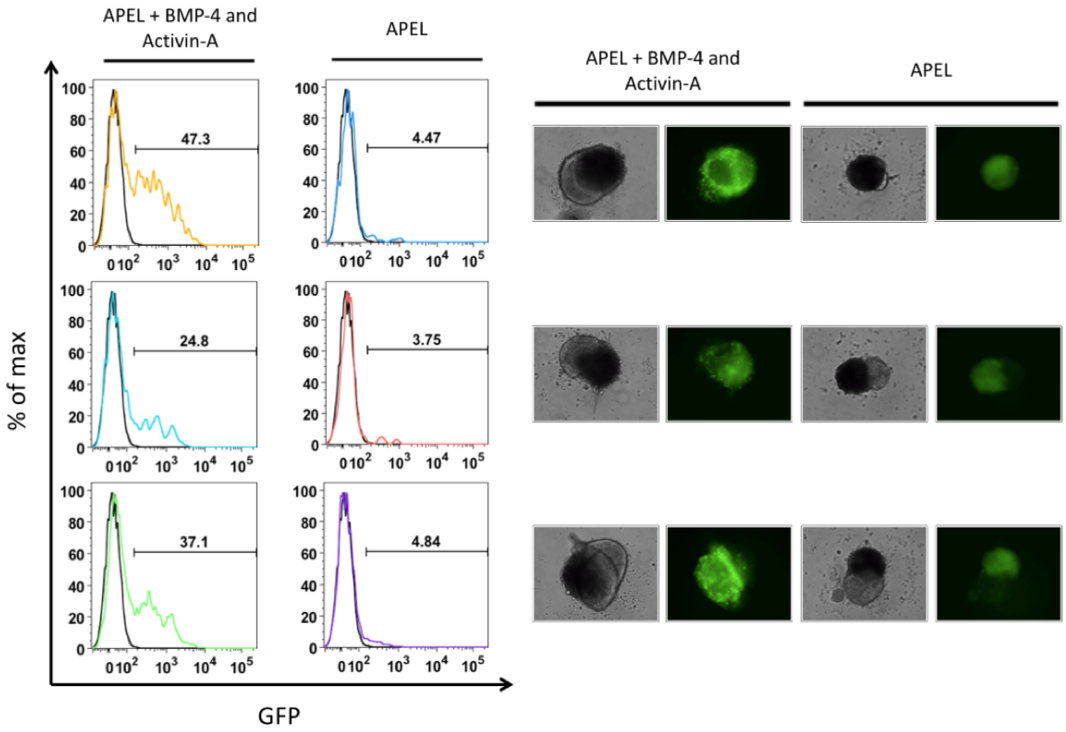

Figure 19. *In vitro* differentiation of the MIXL1-GFP hESC into mesodermal progenitors.

Day 6 post induction embryoid bodies were tested for GFP expression by either FACS analysis (left panels) or fluorescence microscopy (right panels). As negative controls for FACS analysis, EBs derived from iPSC (GFP negative) were used. APEL: Albumin Polyvinylalcohol Essential Lipids medium; BMP-4: Bone Morphogenetic Protein 4.

7.5 CRISPR-Cas9 based gene targeting to track mesodermal progenitors

In order to derive a pure population of mesodermal progenitors that could be used as a starting point for further differentiation into FI and OSE, I resorted to CRISPR-Cas9 technology to target a selection cassette in the MIXL1 locus of tumor-derived iPSC. FI and OSE will then be used to define specific gene expression and epigenetic signatures characterizing each cell type.

I designed a cassette that allows: i) G418 selection of clones carrying the

integration of the targeting construct, either on- or off-site; ii) selection of mesodermal progenitors during differentiation, either by FACS sorting of GFP-positive cells or by puromycin administration when the construct is integrated in frame with the endogenous ATG. FRT sites would allow the excision of the PGK-NeoR by administration of the FLP recombinase (Figure 20).

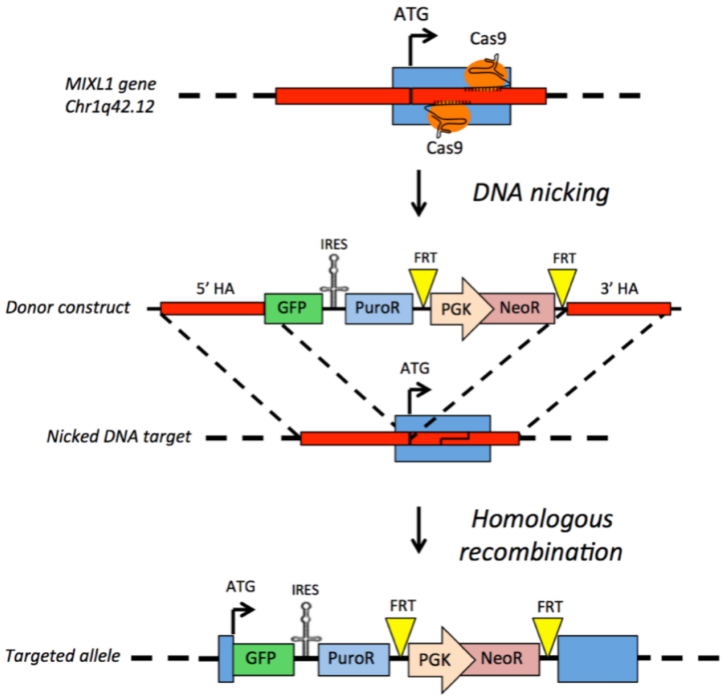

Figure 20. Gene targeting construct and strategy.

HA: homology arm, 500 bp each; GFP: green fluorescent protein; IRES: internal ribosome entry site; PuroR: puromycin resistance gene (puromycin *N*-acetyl-transferase); FRT: FLP recombinase target; PGK: phosphoglycerate kinase promoter; NeoR: neomycin resistance gene.

I first set up the optimal electroporation conditions for the targeting constructs (pCas9-2A-GFP, the guide RNA expressing plasmid, hereafter gRNA, and the donor plasmid) in a normal iPSC line previously generated in our lab (3391B). Since the Cas9-expressing plasmid also expresses GFP, it is possible to assess

the efficiency of transfection by FACS analysis. (Figure 21)

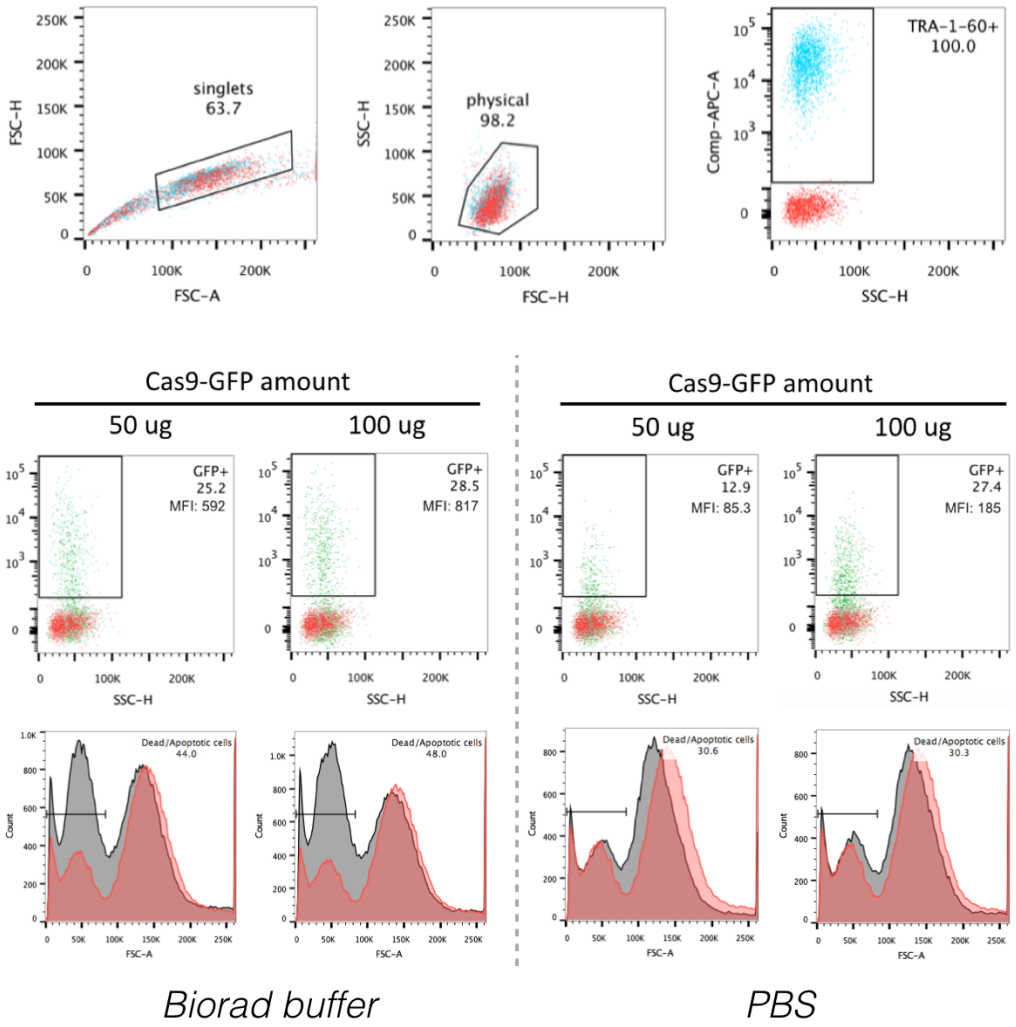

Figure 21. Setup of the electroporation conditions.

FACS analysis performed 48 hours after transfection. Top panels: gating strategy. Cells were selected on the basis of physical parameters for viability and single cell dissociation, and on the basis of TRA-1-60 expression (red – unstained control, blue – stained cells). Middle panels: percentage of GFP+ cells (green) in the considered conditions (red – non transfected cells). MFI: mean fluorescence intensity. Lower panels: histogram gating for dead/apoptotic cells based on the physical parameter FSC (black plot – considered sample, red plot – non transfected cells).

Transfection in Biorad buffer resulted in high toxicity, even if it increased transfection efficiency (higher MFI), therefore I decided to use PBS as preferential buffer for electroporation, and a total of 100 μg of plasmid DNA.

I designed two gRNA to be used in gene targeting experiments, either together in combination with the nickase version of Cas9 (Cas9D10A) or singularly with the wild type Cas9. Cas9D10A allows the delivery on close genomic regions of two single strand breaks that are interpreted by cell's DNA damage response as a double strand break, stimulating DNA repair either by non homologous end joining or homologous recombination. Single strand breaks occurring far apart in the genome in the case of off-target activity do not stimulate homologous recombination, thus reducing potential insertions of the donor construct at unintended sites.

I transfected the appropriate combination of gRNA and Cas9 together with the donor plasmid into 3391B cells (condition A: gRNA1 plus Cas9; condition B: gRNA2 plus Cas9; condition C: gRNA1 and gRNA2 plus Cas9D10A). 48 hours after transfection I added G418 to the culturing medium to select clones that were transfected with the donor plasmid and subsequently integrated the construct. I kept iPSC under G418 selection for three weeks and I picked and expanded 11 independent clones from G418-resistant colonies that have formed. In order to to assess the correct insertion of the targeting cassette, I performed PCR with primers located at the 5' of the construct, outside the homology arms and inside the cassette, I assessed the integration of the cassette

on site. Indeed, some of the G418-resistant clones scored positive in all three transfection conditions (Figure 22a). I also performed a digestion with BamHI and NaeI that cut outside the 5' homology arm and inside the cassette, respectively, showing that the integration was site specific (Figure 22b).

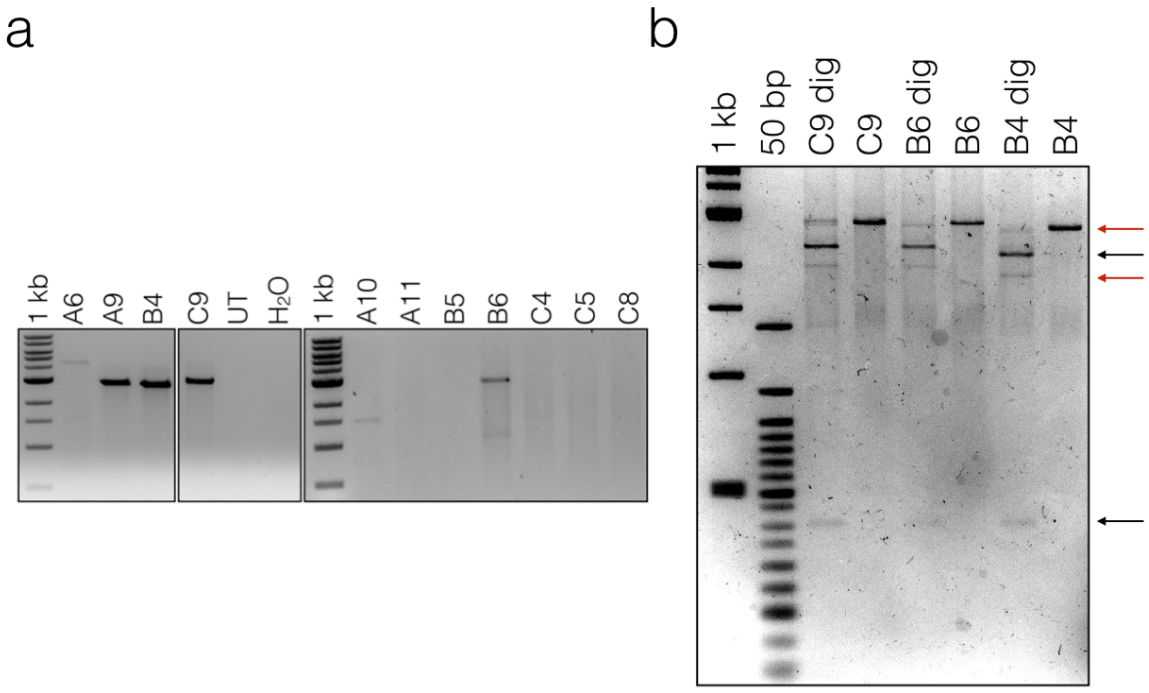

Figure 22. Effective gene targeting at the MIXL1 locus in iPSC.

A. PCR analysis performed on genomic DNA extracted from G418 resistant clones, specific for the 5' end of the integrated cassette (expected size of the amplicon: 2742 bp). B. Digestion of the amplicons belonging to representative clones with BamHI and NaeI restriction enzymes (expected size of the digested bands: 2244 bp, 406 bp and 89 bp, black arrows). Red arrows: undigested DNA and aspecific cutting.

Moreover, as a proof of principle I differentiated to mesodermal progenitors clone B6 showing that GFP+ cells are also Brachyury T+, another defining marker of mesodermal differentiation (Figure 23).

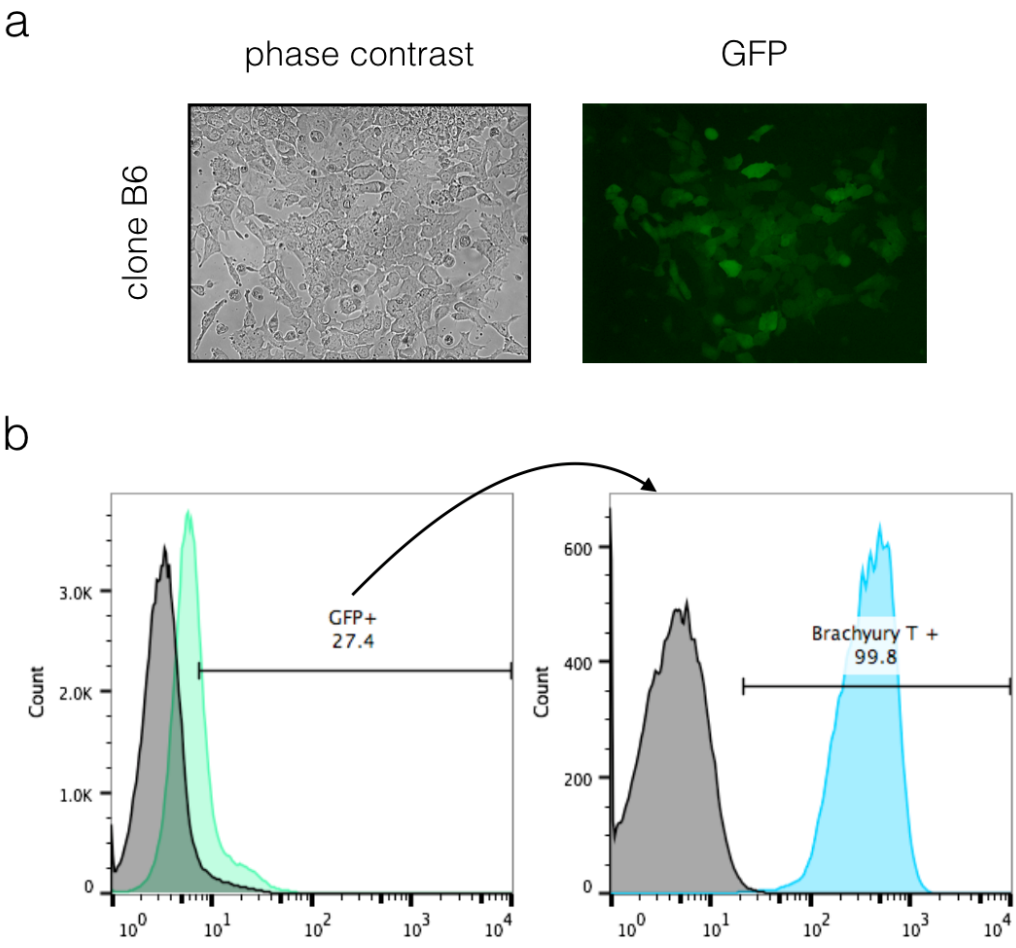

Figure 23. B6 clone shows reactivation of GFP and expression of Brachyury T upon differentiation into mesoderm.

A. Phase contrast and direct fluorescence image of clone B6 at day 3 post-induction of differentiation. B. FACS analysis of the expression of Brachyury T in GFP+ cells at day 3 post-induction of differentiation.

These results indicate that it is possible to track differentiation into mesoderm by gene targeted iPSC in the MIXL1 locus. This approach will be translated also to tumor-derived iPSC as an initial step towards the generation of FRTE *in vivo* and for the analysis of molecular phenotypes *in vitro*.

7.6. Molecular characterization of HGSOC and its cell of origin

From DNA methylation analysis of our iPSC and parental cells, I noticed that the hierarchy between the parental cells was maintained even after reprogramming-induced epigenetic resetting (Figure 12). Indeed, in the parental cell group, 333 tumor was closer to the fimbria compared to 313 tumor. The same holds true in the iPSC group, where 333-derived iPSC are closer to fimbria-derived iPSC compared to 313-derived iPSC. This result suggests that despite genome-wide changes in DNA methylation, epigenetic memory of the parental cell type is still present.

I hypothesized that the same phenomenon could occur during tumor transformation, so that it could be possible to identify DNA methylation signatures reminiscent of those present in the cell of origin in tumor cells. This can be especially relevant for OC, where the identification of the cell of origin is still controversial. Moreover, in-depth analysis of HGSOC epigenetic modifications has been very limited thus far, so there is the need for a better understanding of how epigenome can influence transcription and how this contribute to tumor pathogenesis.

Hence, I proceeded with a more detailed molecular characterization of tumor cells from HGSOC (EOC), tumor cells isolated from HGSOC patients with ascites (AS), fimbrial epithelium cells (FI) and ovarian surface epithelium cells (OSE).

7.6.1. DNA methylation

I subjected to genome wide DNA methylation analysis four fimbria, four OSE, four EOC and five AS independent samples, all derived from different patients. Aim of the analysis was to determine if DNA methylation was a suitable parameter to partition these four categories.

Indeed, unsupervised clustering based on the top 1000 variable sites across all samples showed that these four categories have distinct DNA methylation patterns (Figure 24).

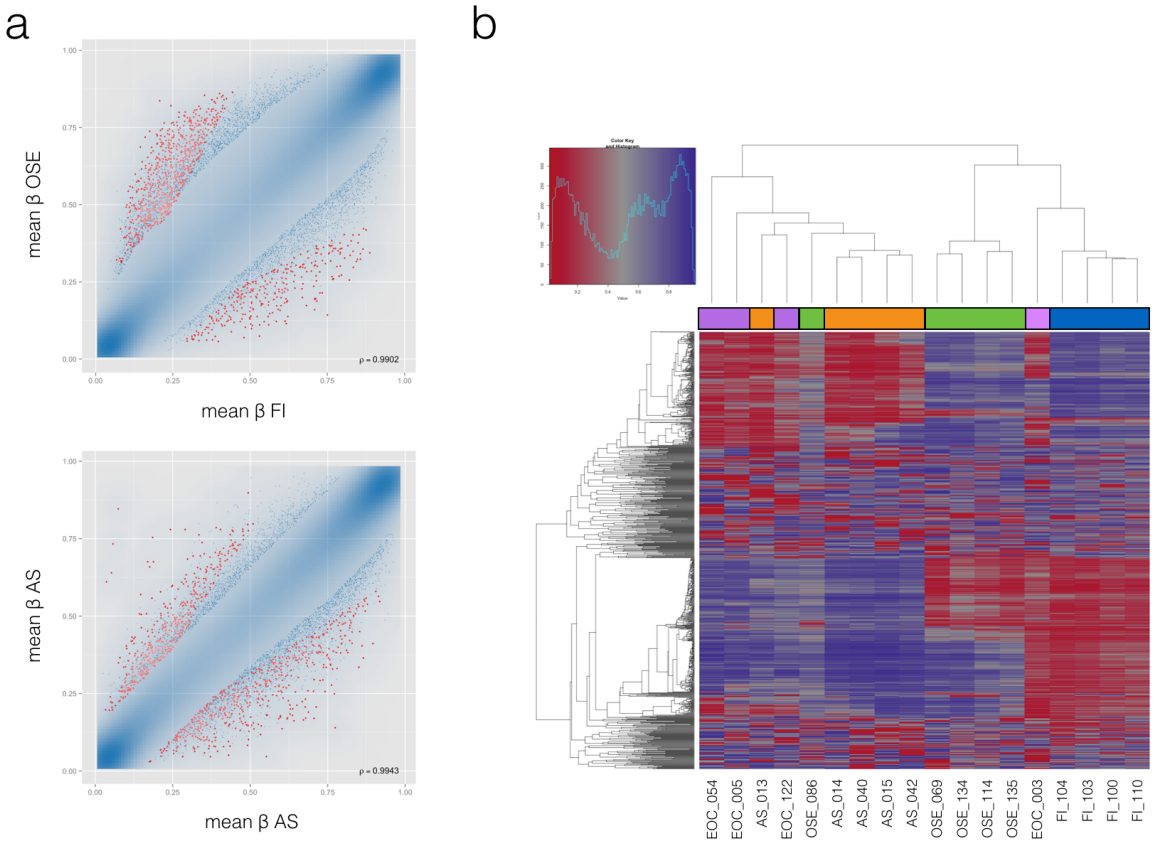

Figure 24. DNA Methylation is able to distinguish gynecological tumor and normal samples.

A. Spearman correlation dotplot between the mean methylation values of FI and OSE (top panel) and AS and EOC (bottom panel). The transparency corresponds to the dots density.

The 1% of the points in the sparsest populated plot regions are drawn explicitly. Red points represent differentially methylated sites in the comparison, defined as the top 1000 ranking sites. Spearman ρ correlation value is shown. B. Unsupervised hierarchical clustering of the top 1000 variable sites across all samples. Colored bars on top define groups: FI (blue), OSE (green), EOC (purple), AS (orange).

Differential methylation analysis between FI and OSE samples uncovered a methylation signature specific for each category, which I then used to query EOC and AS methylomes in order to stratify the samples according to the putative cell of origin.

Interestingly, unsupervised clustering based on the group-specific signatures showed that only EOC_003 clustered with FI samples, while all the other tumor samples showed closer similarity to OSE (Figure 25).

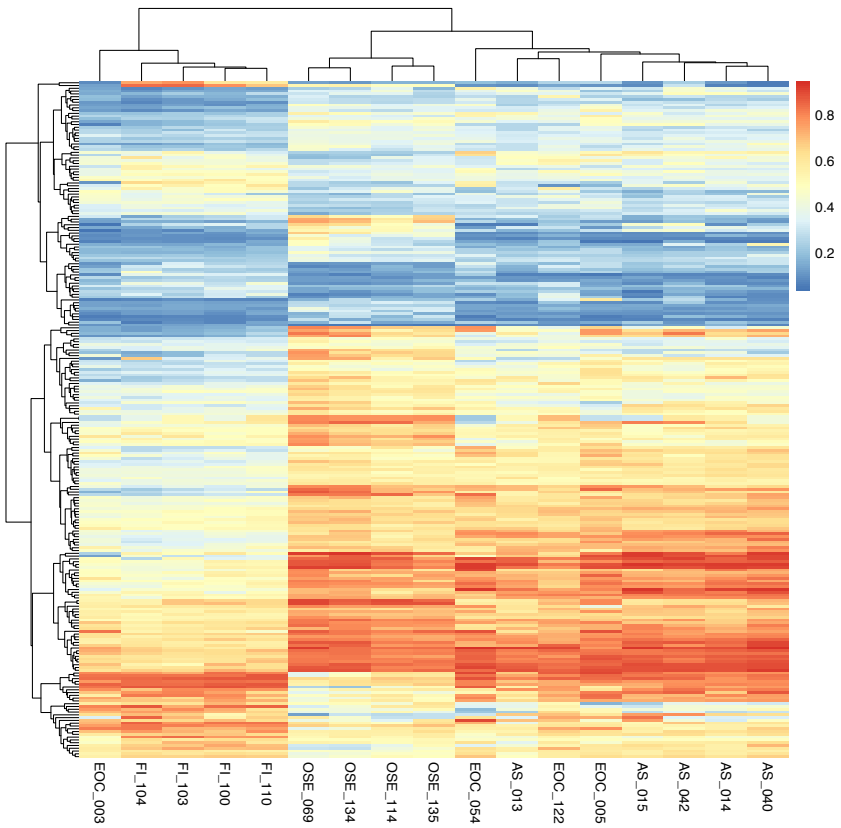

Figure 25. Differentially methylated genes in FI and OSE are able to stratify HGSOC samples.

Unsupervised clustering based on top ranked differentially methylated promoters in FI and OSE.

Even if more samples would be needed to strengthen this point, so far we can say that specific DNA methylation signatures of fimbrial epithelial cells and ovarian surface epithelium cells are partially retained in tumor samples and therefore could be used as a tool to stratify HGSOC samples according to the putative cell of origin of this tumor.

7.6.2. Chromatin Immunoprecipitation

In the past, cancer was considered to be caused mainly by genetic alterations. Nowadays, the view is changed and epigenetic modifications are known to influence gene expression, contributing to tumor development. Nevertheless, the precise contribution of histone modifications to HGSOC pathogenesis is still not well defined. Thus, we performed H3K27me3 and H3K4me3 promoters profiling and active and closed enhancers characterization by using different combinations of H3K4me1, H3K27ac and H3K27me3 modifications. In this initial phase we analyzed 3 HGSOC samples (EOC) and 2 tumor samples from ascites (AS), that allowed simultaneous analysis of multiple marks in cell at the same passage thanks to the abundance of cells derived from these samples.

7.6.2.1 Promoters

We focused our attention on transcription start sites (TSS), deriving localization maps for H3K27me3 and H3K4me3 in a range of 1kb up- and down-stream the TSS. As expected, we were able to classify TSS in four classes according to histone mark occupancy, namely those marked by i) H3K4me3 only, ii) H3K27me3 only, iii) both marks (bivalent TSS) or iv) neither marked TSS (Figure 26).

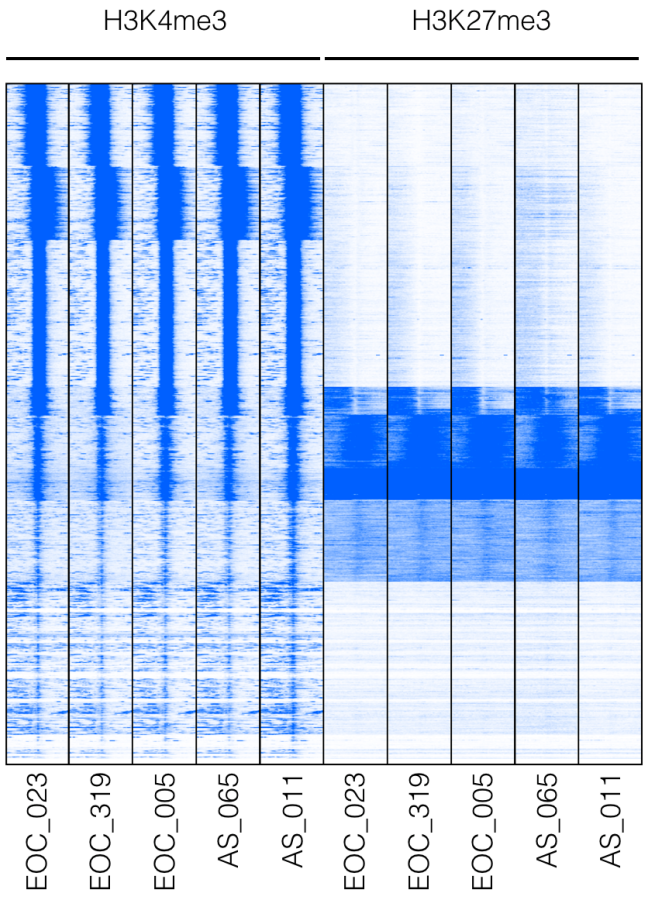

Figure 26. HGSOC samples show homogenous TSS marking across samples.

Heatmap showing the distribution of H3K27me3 and H3K4me3 marks on all TSS in EOC and AS.

This analysis did not highlight any major difference in the overall distribution of the marks between the two groups of samples (EOC and AS). We then analyzed more in details TSS belonging to categories i) and ii) as previously defined. Interestingly, the overlap of H3K4me3^{+} TSS across different samples of the same kind (EOC or AS) is greater than 80% (86% on average for EOC and 89% on average for AS samples). On the contrary, the overlap of H3K27me3^{+} TSS across different samples of the same kind is about 16% on average for both EOC and AS samples.

We then used the results of the overlaps representing the common TSS in EOC and in AS (and therefore overcoming the patient-specific distribution of histone marks) to compare the two different stages of the tumor (EOC vs AS) (Figure 27a). The vast majority of H3K4me3^{+} TSS are shared between EOC and AS, while less than 30% of H3K27me3^{+} TSS bear the mark at both stages (Figure 27b), suggesting that while the H3K4me3 is less variable at different tumor stages, H3K27me3 has more stage-specific localizations.

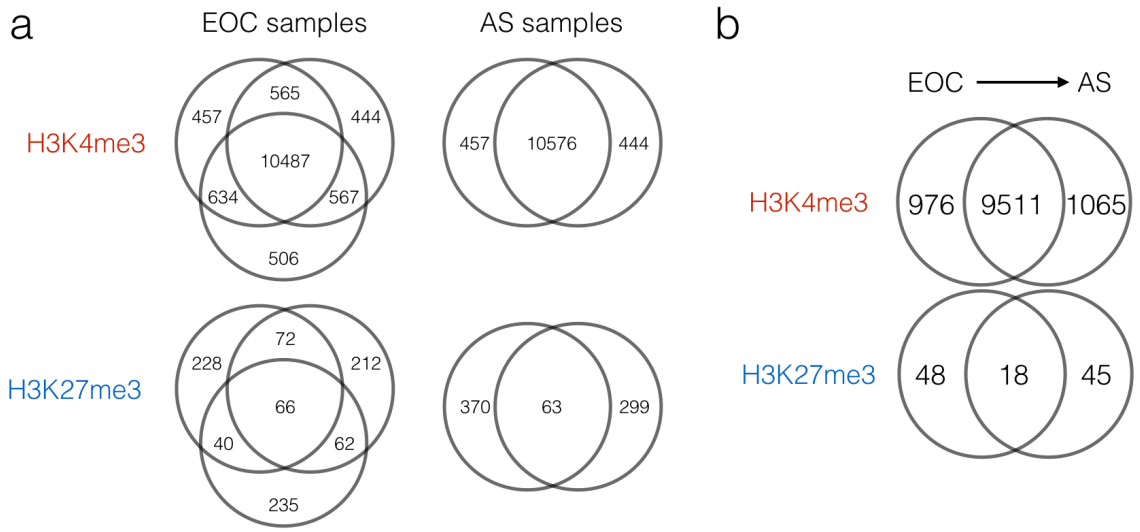

Figure 27. Differential occupancy of H3K27me3 and H3K4me3 histone marks at TSS in EOC and AS.

A. Group-based overlap analysis of H3K4me3- and H3K27me3-marked TSS in EOC and AS samples; B. Identification of exclusive and common H3K27me3- and H3K4me3-marked TSS in EOC and AS.

H3K27me3^{+} TSS were too few to perform an enrichment test, and therefore we performed ingenuity pathway analysis only on H3K4me3^{+} TSS in EOC and AS. For EOC, the analysis uncovered an enrichment for genes involved in the inhibition of matrix metalloproteases, Wnt/beta-Catenin and cAMP signaling. For AS, instead, the best ranking pathways are associated to mediators of intra-cellular response (Figure 28).

Figure 28. Ingenuity pathway analysis performed on exclusive H3K4me3-marked TSS in EOC and AS.

We reanalyzed with our computational pipeline the dataset produced by Chapman-Rothe and colleagues (Chapman-Rothe et al. 2013) on a single HGSOC and verified that the gene promoters they selected and validated as belonging to H3K27me3 only (EN1 and ZIC4), H3K4me3 only (FBXO33 and IKBIP) and bivalent (ALX1 and COCH) groups were still marked with the same histone modifications in the re-analyzed dataset. Next, we sought to verify whether their findings could still hold true in our dataset.

Interestingly, we found that while their H3K4me3 marked promoters fell into the same group also in our EOC samples, H3K27me3 marked promoters switched invariably to the H3K4me3^{+} category in our dataset and bivalent promoters were either confirmed or found in the H3K4me3^{+} class (Table 9). Moreover, the overlap of TSS belonging to each category in our classification and in the one presented by Chapman-Rothe and coworkers is overall very low, with the greatest overlap in the H3K4me3 only category (around 55%) (Figure

29).

Gene promoter	HGSOC (Chapman-Rothe)	EOC_319	EOC_05	EOC_23
EN1	H3K27me3	H3K4me3	H3K4me3	H3K4me3
ZIC4	H3K27me3	H3K4me3	H3K4me3	H3K4me3
FBXO33	H3K4me3	H3K4me3	H3K4me3	H3K4me3
IKBIP	H3K4me3	H3K4me3	H3K4me3	Not present
ALX1	Bivalent	Bivalent	H3K4me3	H3K4me3
COCH	Bivalent	H3K4me3	H3K4me3	Bivalent

Table 9. Categorization of the selected genes in the published sample and in our dataset of EOC.

Green: coherent categorization with our dataset; Red: Incoherent categorization with our dataset.

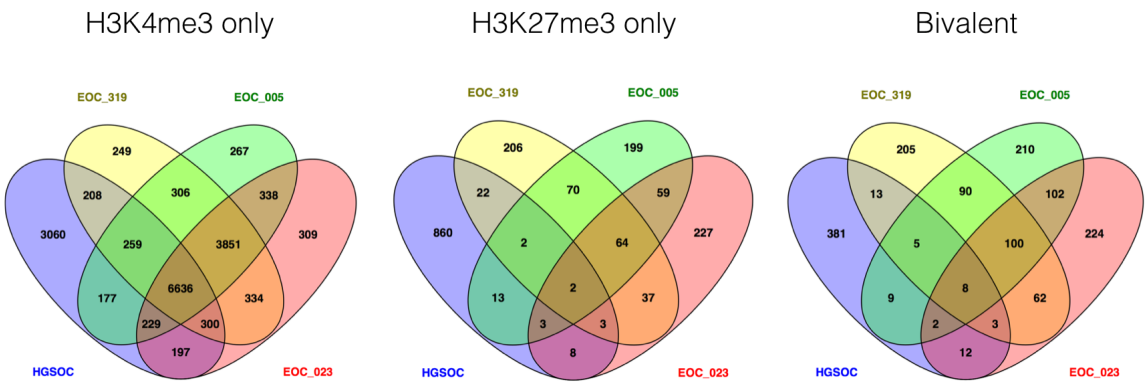

Figure 29. HGSOC published dataset shares a minority of peaks with our samples.

Venn diagrams showing the overlap of marked TSS in the considered groups between our samples and the published HGSOC.

These results suggest that the number of samples used greatly influences the definition of specific marks at promoter regions in HGSOC.

7.6.2.2. Enhancers

In a preliminary analysis, we sought to identify putative enhancers specific for the EOC and for the AS samples. We defined enhancers taking into account a region spanning 1 to 5kb upstream the TSS, and classified them as active, when we found co-localization of H3K4me1 and H3K27ac, or as closed/poised when H3K4me1 co-localized with H3K27me3 (Heintz et al. 2015).

By comparing active or closed enhancers in EOC and AS, we saw that the overlap was greater for active enhancers, while closed/poised enhancers were less conserved between the two groups (overlap percentages: 51% in EOC and 65% in AS for active enhancers and 10% in EOC and 36% in AS for poised enhancers). As shown by the Venn diagrams, at each tumor stage we found a set of specific active and poised/closed enhancers (Figure 30).

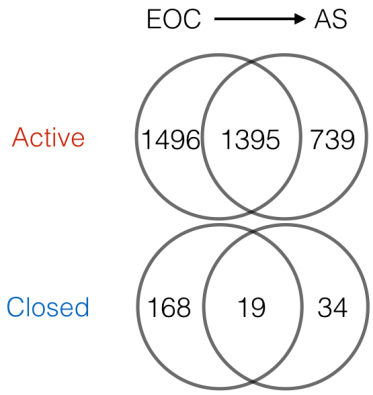

Figure 30. Distribution of putative active and closed/poised enhancers in EOC and AS.

Venn diagram showing common and exclusive active and closed/poised enhancers in EOC and AS.

Both in the case of promoters and enhancers, it will be interesting to compare the ChIP-seq data with transcriptomic data. If the active or “repressed” status suggested by promoters and enhancer is reflected by gene expression, this might represent an interesting molecular finding to shed new light on epigenetic contribution to ovarian cancer pathogenesis.

8. DISCUSSION

Ovarian Cancer is a leading cause of cancer-related death, with high incidence and mortality due to the failure of surgery and chemotherapy to eradicate the disease. Currently available cellular and animal models fail to fully recapitulate both tumor genetic heterogeneity and histopathology. Also, the uncertainty regarding the actual cell of origin of this class of tumors has hampered the identification of relevant targets for therapy. Additionally, to date very little is known about the epigenetic vs. genetic contribution to this disease, thus pointing to a strong need for innovative research that might contribute to the elucidation of fundamental mechanisms driving OC pathogenesis.

The work presented in this thesis made use of two complementary approaches based on somatic cell reprogramming to pluripotency to address the following issues: i) the functional dissection of the reciprocal contribution of epigenetic and genetic contribution to ovarian cancer pathogenesis through reprogramming of tumor cells and differentiation into the female reproductive tract epithelium, and ii) a fine characterization of HGSOC molecular signatures that can be used to identify the cell of origin of this tumor type.

For the first part, my central hypothesis was that reprogramming to iPSC would allow the erasure of tumor-associated epigenetic marks without affecting the underlying mutated genome. Then, the in vivo differentiation of OC-derived iPSC to female reproductive tract epithelium (FRTE) would unfold the epigenetic mechanisms that evolved during the parental tumor progression. In

this context, the analysis of differences at the epigenomic and transcriptomic level compared to the parental tumor would allow the description of relevant pathways for the disease progression.

Thus far, it is still unknown whether all cancer types can be reprogrammed to the pluripotent state, since some genetic lesion might hamper the reprogramming process per se or the stability of pluripotency. Moreover, a careful choice of the most appropriate reprogramming platform is needed in order to: i) minimize the risk for insertional mutagenesis, i.e. insertion of the reprogramming vector in a gene body, causing the knockout of the gene, or the up-regulation of nearby genes' expression, due to the interaction with splicing sites and transcription factor binding sites that are present in some of the promoters used (Lombardo et al. 2011); ii) avoid residual expression of the reprogramming transgenes either at the iPSC stage or during differentiation. Indeed, the reprogramming cocktail include oncogenes as c-MYC and KLF4, whose transient expression has also been demonstrated to drive tumorigenesis per se (Ohnishi et al. 2014).

Keeping in mind these caveats, I made use of the STEMCCA LV (Somers et al. 2010), a single LoxP-flanked excisable vector expressing the four Yamanaka reprogramming factors, that allows the reprogramming of fibroblasts to iPSC from a single integrated copy. This would minimize the risk of unwanted inter-/intra-chromosomal recombination upon Cre-recombinase administration during the excision of the reprogramming vector, that is done in

order to avoid expression of the transgenes during differentiation.

However, this approach was inefficient to drive reprogramming in most of the available samples, with few exceptions that were reprogrammed despite with very low efficiency. This could be attributed either to an intrinsic inefficient reprogramming of tumor epithelial cells from a single-copy vector or to the presence of mutations that could hamper the reprogramming process.

Thus, I took advantage of recent successful reports describing the use of either mRNA or miRNAs related to pluripotency to generate iPSC (Warren et al. 2010, Anokye-Danso et al. 2011) and I applied an integration-free approach based on the daily transfection of a combination of both miRNA and mRNA, that our lab had previously described to generate vector-free iPSC from fibroblasts in less than 15 days (Adamo, Atashpaz, Germain et al. 2015).

With this method, I was able to obtain iPSC from two samples (one undifferentiated ovarian carcinoma and one low grade serous ovarian cancer) that in the first experiment were refractory to viral-based reprogramming (Table 8 and Figure 9). These iPSC lines could be expanded for more than 30 passages, maintaining an undifferentiated phenotype and the expression of pluripotency markers. At the DNA methylation level, they closely resemble human ESC lines genome-wide, with consistent de-methylation of NANOG and OCT4 promoters.

The two samples were reprogrammed at different efficiencies, with a surprisingly higher efficiency in the low grade tumor with respect to the

undifferentiated tumor. It has been shown that reprogramming is more amenable in cells that are less differentiated, such as in the case of adult and fetal neural stem cells (Kim et al. 2008, Kim et al. 2009). Moreover, more aggressive HGSOC express pluripotency markers NANOG and OCT4 (Siu et al. 2013, Peng et al. 2010). Finally, surface markers CD133 and CD44 have been associated with putative cancer stem cells in HGSOC (Kryczek et al. 2012, Zhang et al. 2008), suggesting that there are subpopulations in higher grade OC characterized by a more undifferentiated state and therefore more prone to reprogramming. Despite this, I found that most of the aforementioned genes were not expressed, with the exception of CD44 that had a comparable expression in both samples. The lack of expression of stem-associated markers suggests that the difference in reprogramming efficiency cannot be explained by the presence or absence of stem-like cells in the tumors.

I performed whole exome sequencing analysis on tumor samples and their derived iPSC to rule out the possibility that the cell culture established from low grade tumor contained normal mesenchymal tumor-associated cells. In this scenario, if tumor mutations are detrimental to the reprogramming process, normal cells could have been reprogrammed instead of tumor cells. Anyway, as showed both at the SNV and CNV level, iPSC share a good fraction of the parental tumor genetic aberrations. This evidence proves that iPSC are indeed derived from tumor cells and not from contaminant cells.

Moreover, if early-stage mutations, i.e. present in all cells of the tumor, were

incompatible with the pluripotent state, we would have failed in obtaining any iPSC. The reprogrammed clones therefore carry early stage aberrations that occurred early in the parental tumor.

The differentiation potential of OC-derived iPSC was assessed by teratoma formation assay *in vivo*. OC-derived iPSC were able to differentiate into all three germ layers, proving the pluripotency of these cells. Interestingly, one clone gave rise to a teratocarcinoma recapitulating the features of a medulloepithelioma (Ulbright et al. 2010), a primitive neuroectodermal tumor (PNET). The medulloepithelioma unexpectedly expressed PAX8, one of the defining markers of serous ovarian cancer (Laury et al. 2011), that was already expressed in the parental tumor but that has never been associated with PNET.

Several explantions can justify this finding. One could be that there were mutations in the parental tumors found with very low frequency, possibly due to the heterogeneity of the tumor bulk, that we were not able to detect with the depth of sequencing we used. A portion of these mutations might be exclusively represented in one clone of iPSCs and be causal to specific features of each single iPSC clone (e.g., the aberrant expression of PAX8 in the tumor and in the teratocarcinoma). More in general, this would support the idea that mutations we found to be exclusive to iPSC could be instead present in the tumor and not being detected, arguing against de novo acquisition of mutations during reprogramming.

Another hypothesis could be that epigenetic resetting occurred during the

reprogramming process was able to suppress the tumor phenotype in some clones, while tumor transformation upon differentiation with, as an example, persistence of PAX8 expression could be the result of epigenetic memory of the parental tumor in other clones.

Combining transcriptomic data and a more detailed analysis of DNA methylation in the parental tumor and iPSC, could help in shedding light on this result.

CNV analyses on tumors and iPSC also provided a snapshot of the tumor genetic landscape. Indeed, since iPSC are generated from a single cell in culture, they act as a "magnifying lens" on the genetic composition of subclones in the heterogeneous tumor bulk. An in depth analysis of mutated and gained/lost genes on a greater number of iPSC clones per tumor would allow the genetic reconstruction of the subclones present in the parental tumor and to deconvolute tumor genetic heterogeneity. Moreover, it will also help in tracking the evolution of the tumor: since iPSC are clonal, a consistent increase in frequency of genetic alterations from tumor to iPSC would indicate that these mutations are less represented in the parental tumor. Unchanged frequencies instead would indicate that the mutations are present in the vast majority of tumor cells, pointing to early events in tumor pathogenesis. Also, the clonality of iPSC would allow to assign defined haplotypes to subclones present in the tumor bulk, allowing the reconstruction of the genetic evolution of the considered tumor.

Next, I sought to set up a streamlined platform for the generation and isolation of mesodermal progenitors from iPSC. This would allow to study *in vivo* the progression of differentiating tumor-derived iPSC, and to study the effect of epigenetic resetting on the cancer genome, by comparing the parental tumor and the differentiating cells both at the histopathological and molecular level. Our lab has previously shown that disease-relevant transcriptional phenotypes are already detectable at the iPSC stage and are amplified during differentiation (Adamo, Atashpaz, Germain et al. 2015). Since tumor-derived iPSC carry tumor mutations, it is plausible that during differentiation any effect at the molecular level might be present already in mesodermal progenitors, a common ancestor of both fimbria and ovarian surface epithelium. Moreover through this intermediate step, it would be possible to terminally differentiate iPSC *in vivo* by generating female reproductive tract epithelium (Ye et al. 2011).

I resorted to CRISPR/Cas9 technology to integrate a GFP-puro cassette in the MIXL1 locus, in frame with the endogenous ATG, and managed to isolate iPSC clones carrying proper integration that allows selection of a pure population of mesodermal progenitor cells during differentiation (Figure 21-22).

These cells will be used for the identification of stage-type specific epigenetic and transcriptional signatures by ChIPseq, DNA methylation and RNAseq analyses.

To this regard, this same approach could be implemented for the characterization of HGSOC cell of origin, that thus far is still controversial.

Mesodermal progenitors derived from normal iPSC from female donor could be used to define fimbria and OSE specific epigenetic and transcriptional signatures. One can expect that, despite being the result of genetic and epigenetic aberrations, HGSOCs would still retain epigenetic features that are characteristic of either Fimbria or OSE. The preliminary analysis based on DNA methylation of samples belonging to fimbria, OSE, solid HGSOC and ascitic cells showed that it is possible to distinguish all the four categories accordng to their DNA methylation profiles (Figure 24). Particularly, I could show that HGSOC and ascites more closely correlate with OSE, with the exception of one sample that instead showed very high correlation with fimbria promoters methylation pattern. As far as we can evaluate from this limited set of samples, these data confirm the initial hypothesis that different HGSOC samples can derive from both cell types, and DNA methylation could represent a good marker for the identification of the cell of origin. In order to strengthen this result, more samples will be subjected to the same analysis.

Distribution of well characterized histone modifications has been examined in a cohort of HGSOC and ascitic samples. As a first level of analysis we selected histone modifications marking active or repressed promoter regions (H3K4me3 and H3K27me3, respectively) (Byrd and Shearn 2003, Kirmizis et al 2004) and active or closed/poised enhancers (H3K4me1 in combination with either H3K27ac or H3K27me3, respectively) (Heinz et al. 2015). TSS were classified in four groups according to histone mark occupancy, namely those marked by i) H3K4me3 only, ii) H3K27me3 only, iii) both marks (bivalent TSS) or iv)

neither marked TSS.

We focused on the first two classes and showed that the vast majority of H3K4me3-marked TSS is conserved across tumor samples, while we could score higher inter-sample variability in the case of H3K27me3 TSS. Moreover, most H3K4me3-marked TSS were shared in the transition from EOC to AS, while H3K27me3-marked TSS were more specific to each category. This finding could be attributed to a higher conservation of activation marks, with lower consistency of silenced genes in HGSOC and in AS. Another possibility could be that the different nature of H3K27me3 deposition at TSS, characterized by less sharp peaks and a broader distribution, results in more difficult peak calling by current statistical methods. Indeed, we found that the number of peaks identified for H3K27me3 was greatly lower with respect to H3K4me3. Moreover, depending on the distribution of the peak (i.e. centered on the TSS, spread on the promoter with a narrow depletion on the TSS, spread on the whole gene body), it has been proposed that this mark can either suppress or favour transcription (Young et al. 2011), adding another level of complexity to the understanding of its functional role. In addition to this, we are currently profiling the transcriptome of these cells. The correlation between the distribution of epigentic marks (both histone modifications and DNA methylation profiles) and gene expression on a large number of samples could consequently clarify how chromatin modifications affect tumor pathogenesis, overcoming the inter-patient variability.

By comparing our data with the only dataset present in the literature (Chapman-Rothe et al 2013), we also argued that analyses performed on single or few tumor samples could be of limited value for the dissection of HGSOC pathogenesis. Increasing the number of samples would be mandatory to identify reliable epigenetic features of HGSOC.

The difficulty in peak calling for H3K27me3 could also be the reason why the number of closed/poised putative enhancers we identified was considerably lower with respect to active ones. To address this issue, we will perform NOMe-seq analysis (Kelly et al. 2012) to evaluate nucleosome occupancy at these sites, since more compacted or open chromatin at enhancers has been associated with activity or poisedness (Heinz et al. 2015).

Normal samples and mesodermal progenitors from normal iPSC will be included in the study, to finely characterize specific epigenetic features of fimbria and OSE that are retained in HGSOC samples, addressing the problem of the cell of origin of the tumor. This would allow to refine the identification of relevant dysregulated pathways in HGSOC according to the tumor-specific relevant cell of origin, hopefully leading to the discovery of clinically relevant targets for a better care of OC patients.

9. BIBLIOGRAPHY

Adamo A, Atashpaz S, Germain PL, Zanella M, D'Agostino G, Albertin V, et al. (2015). 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. Nat Genet 47: 132–41.

Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad M, Sharma R, et al. (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J. Pathol. 221: 49–56.

Anokye-Danso F, Trivedi CM, Juhr D, Gupta M, Cui Z, Tian Y, et al. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell. 8: 376–88.

Arisan S, Buyuktuncer ED, Palavan-Unsal N, Caşkurlu T, Cakir OO, and Ergenekon E. (2005). Increased expression of EZH2, a polycomb group protein, in bladder carcinoma. Urol. Int. 75: 252–7.

Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, and Bock C. (2014). Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods 11: 1138–40.

Audeh MW, Carmichael J, Penson RT, Friedlander M, Powell B, Bell-McGuinn KM, et al. (2010). Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. Lancet 376: 245–51.

Auersperg N. (2013). Ovarian surface epithelium as a source of ovarian cancers: unwarranted speculation or evidence-based hypothesis? Gynecol. Oncol. 130: 246–51.

Baldwin RL, Nemeth E, Tran H, Shvartsman H, Cass I, Narod S, et al. (2000). BRCA1 promoter region hypermethylation in ovarian carcinoma: a population-based study. Cancer Res. 60: 5329–33

Barton CA, Clark SJ, Hacker NF, and O'Brien PM. (2008). Epigenetic markers of ovarian cancer. Adv. Exp. Med. Biol. 622: 35–51.

Baylin SB, and Ohm JE. (2006). Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat. Rev. Cancer 6: 107–16

Baylin SB, and Jones PA. (2011). A decade of exploring the cancer epigenome - biological and translational implications. Nat. Rev. Cancer 11: 726–34.

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125: 315–26.

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, et al. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol. Cell. Biol. 21: 289–97

Blelloch RH, Hochedlinger K, Yamada Y, Brennan C, Kim M, Mintz B, et al. (2004). Nuclear cloning of embryonal carcinoma cells. Proc. Natl. Acad. Sci. U.S.A. 101: 13985–90.

Boeva V, Zinovyev A, Bleakley K, Vert J-PP, Janoueix-Lerosey I, Delattre O, et al. (2011). Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. Bioinformatics 27: 268–9.

Bowen NJ, Walker L, and Matyunina LV. (2009). Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. BMC medical genomics. 2:71

Bowtell DD. The genesis and evolution of high-grade serous ovarian cancer.

Nat Rev Cancer. 10: 803–8.

Bracken AP, Dietrich N, Pasini D, Hansen KH, and Helin K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. 20: 1123–36.

Bracken AP, Kleine-Kohlbrecher D, Dietrich N, Pasini D, Gargiulo G, Beekman C, et al. (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes Dev. 21: 525–30.

Branco MR, Ficz G, and Reik W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat. Rev. Genet. 13: 7–13.

Briggs R, and King TJ. (1952). Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. Proc. Natl. Acad. Sci. U.S.A. 38: 455–63.

Bryant H, Schultz N, Thomas H, Parker K, Flower D, Lopez E, et al. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nat. 434: 913–7.

Byrd KN, and Shearn A. (2003). ASH1, a Drosophila trithorax group protein, is required for methylation of lysine 4 residues on histone H3. Proc. Natl. Acad.

Sci. U.S.A. 100: 11535–40.

Carette JE, Pruszak J, Varadarajan M, Blomen VA, Gokhale S, Camargo FD, et al. (2010). Generation of iPSCs from cultured human malignant cells. Blood. 115: 4039–42

Carey BW, Markoulaki S, Hanna J, Saha K, Gao Q, Mitalipova M, et al. (2009). Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc. Natl. Acad. Sci. U.S.A. 106: 157–62.

Chapman-Rothe N, Curry E, Zeller C, Liber D, Stronach E, Gabra H, et al. (2013). Chromatin H3K27me3/H3K4me3 histone marks define gene sets in high-grade serous ovarian cancer that distinguish malignant, tumour-sustaining and chemo-resistant ovarian tumour cells. Oncogene. 32: 4586–92

Cheng X, Ying L, Lu L, Galvão AM, Mills JA, Lin HC, et al. (2012). Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. Cell Stem Cell 10: 371–84.

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. 31: 213–9.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science. 339: 819–23.

Cowan CA, Atienza J, Melton DA, and Eggan K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science 309: 1369–73.

Davis RL, Weintraub H, and Lassar AB. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51: 987–1000.

Domcke S, Sinha R, Levine DA, Sander C, and Schultz N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nat Commun 4: 2126.

Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, et al. (2001). Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. J. Gen. Virol. 82: 1013–25.

Eggan K, Baldwin K, Tackett M, Osborne J, Gogos J, Chess A, et al. (2004). Mice cloned from olfactory sensory neurons. Nature 428: 44–9.

Eminli S, Utikal J, Arnold K, Jaenisch R, and Hochedlinger K. (2008).

Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. Stem Cells 26: 2467–74.

Esteller M. (2008). Epigenetics in cancer. N. Engl. J. Med. 358: 1148–59.

Evans MJ, and Kaufman MH. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature 292: 154–6.

Flesken-Nikitin A, Hwang CI, Cheng CY, and Michurina TV. (2013). Ovarian surface epithelium at the junction area contains a cancer-prone stem cell niche. Nature. 495:241-5.

Geisler JP, Hatterman-Zogg MA, Rathe JA, Buller RE. (2002). Frequency of BRCA1 dysfunction in ovarian cancer. J. Natl. Cancer Inst. 94:61-7.

Hylander BL, Punt N, Tang H, Hillman J, Vaughan M, Bshara W, et al. (2013). Origin of the vasculature supporting growth of primary patient tumor xenografts. J Transl Med 11: 110.

Gurdon JB, Elsdale TR, and Fischberg M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature 182: 64–5.

Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, et al.

(2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell 133: 250–64.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38: 576–89.

Heinz S, Romanoski CE, Benner C, and Glass CK. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. 16: 144–54.

Hochedlinger K, and Jaenisch R. (2002). Monoclonal mice generated by nuclear transfer from mature B and T donor cells. Nature 415: 1035–8.

Hochedlinger K, Blelloch R, Brennan C, Yamada Y, Kim M, Chin L, et al. (2004). Reprogramming of a melanoma genome by nuclear transplantation. Genes Dev. 18: 1875–85.

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nat. Biotechnol. 27: 851–7.

Hockemeyer D, Wang H, Kiani S, Lai CS, Gao Q, Cassady JP, et al. (2011).

Genetic engineering of human pluripotent cells using TALE nucleases. Nat. Biotechnol. 29: 731–4.

Honda H, Pazin MJ, Ji H, Wernyj RP, and Morin PJ. (2006). Crucial roles of Sp1 and epigenetic modifications in the regulation of the CLDN4 promoter in ovarian cancer cells. J. Biol. Chem. 281: 21433–44.

Honda H, Pazin MJ, D'Souza T, Ji H, and Morin PJ. (2007). Regulation of the CLDN3 gene in ovarian cancer cells. Cancer Biol. Ther. 6: 1733–42.

Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, et al. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature 460: 1132–5.

Hou Z, Zhang Y, Propson NE, Howden SE, Chu L-FF, Sontheimer EJ, et al. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. U.S.A. 110: 15644–9.

Hu S, Yu L, Li Z, Shen Y, Wang J, Cai J, et al. (2010). Overexpression of EZH2 contributes to acquired cisplatin resistance in ovarian cancer cells in vitro and in vivo. Cancer Biol. Ther. 10: 788–95.

Jia F, Wilson KD, Sun N, Gupta DM, Huang M, Li Z, et al. (2010). A nonviral minicircle vector for deriving human iPS cells. Nat. Methods 7: 197–9.

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. (2013). Mutational landscape and significance across 12 major cancer types. Nature 502: 333–9.

Karst AM, and Drapkin R. (2010). Ovarian cancer pathogenesis: a model in evolution. J Oncol 2010: 932371.

Katsaros D, Cho W, Singal R, Fracchioli S, Rigault De La Longrais IA, Arisio R, et al. (2004). Methylation of tumor suppressor gene p16 and prognosis of epithelial ovarian cancer. Gynecol. Oncol. 94: 685–92.

Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, and Jones PA. (2012). Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. 22: 2497–506.

Kim JB, Zaehres H, Wu G, Gentile L, Ko K, Sebastiano V, et al. (2008). Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. Nature 454: 646–50.

Kim JB, Greber B, Araúzo-Bravo MJ, Meyer J, Park KI, Zaehres H, et al.

(2009). Direct reprogramming of human neural stem cells by OCT4. Nature 461: 649–3.

Kim J, Coffey DM, Creighton CJ, Yu Z, Hawkins SM, and Matzuk MM. (2012). High-grade serous ovarian cancer arises from fallopian tube in a mouse model. Proc. Natl. Acad. Sci. U.S.A. 109: 3921–6.

Kim J, Hoffman JP, Alpaugh RK, Rhim AD, Reichert M, Stanger BZ, et al. (2013). An iPSC line from human pancreatic ductal adenocarcinoma undergoes early to invasive stages of pancreatic cancer progression. Cell Rep. 3: 2088–99.

Kim J, Coffey DM, Ma L, and Matzuk MM. (2015). The ovary is an alternative site of origin for high-grade serous ovarian cancer in mice. Endocrinology 156: 1975–81.

Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, et al. (2004). Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev. 18: 1592–605.

Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. (2003). EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc. Natl. Acad. Sci. U.S.A. 100:

11606–11

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22: 568–76.

Kouzarides T. (2007). Chromatin modifications and their function. Cell 128: 693–705.

Kryczek I, Liu S, Roh M, Vatan L, Szeliga W, Wei S, et al. (2012). Expression of aldehyde dehydrogenase and CD133 defines ovarian cancer stem cells. Int. J. Cancer 130: 29–39.

Kuhn E, Kurman RJ, Vang R, Sehdev AS, Han G, Soslow R, et al. (2012). TP53 mutations in serous tubal intraepithelial carcinoma and concurrent pelvic high-grade serous carcinoma--evidence supporting the clonal relationship of the two lesions. J. Pathol. 226: 421–6.

Kumano K, Arai S, Hosoi M, Taoka K, Takayama N, Otsu M, et al. (2012). Generation of induced pluripotent stem cells from primary chronic myelogenous leukemia patient samples. Blood 119: 6234–42.

Kuperwasser C, Hurlbut GD, Kittrell FS, Dickinson ES, Laucirica R, Medina

D, et al. (2000). Development of spontaneous mammary tumors in BALB/c p53 heterozygous mice. A model for Li-Fraumeni syndrome. Am. J. Pathol. 157: 2151–9.

Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, and Graf T. (2006). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. *Immunity* **25**: 731–44.

Langmead B, Trapnell C, Pop M, and Salzberg SL. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Laugesen A, and Helin K. (2014). Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell 14: 735–51.

Laury AR, Perets R, Piao H, Krane JF, Barletta JA, French C, et al. (2011). A comprehensive analysis of PAX8 expression in human epithelial tumors. Am. J. Surg. Pathol. 35: 816–26.

Lewis EB. (1978). A gene complex controlling segmentation in Drosophila. Nature 276: 565–70.

Lewis PF, and Emerman M. (1994). Passage through mitosis is required for

oncoretroviruses but not for the human immunodeficiency virus. J. Virol. 68: 510–6.

Li H, and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–9.

Li H, Collado M, Villasante A, Strati K, Ortega S, Cañamero M, et al. (2009). The Ink4/Arf locus is a barrier for iPS cell reprogramming. Nature 460: 1136–9

Liu H, Zhu F, Yong J, Zhang P, Hou P, Li H, et al. (2008). Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. Cell Stem Cell 3: 587–90.

Lombardo A, Cesana D, Genovese P, Di Stefano B, Provasi E, Colombo DF, et al. (2011). Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat. Methods 8: 861–9.

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, et al. (2013). CAS9 transcriptional activators for target specificity screening and

paired nickases for cooperative genome engineering. Nat. Biotechnol. 31: 833–8

Martin GR. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl. Acad. Sci. U.S.A. 78: 7634–8.

Mayshar Y, Ben-David U, Lavon N, Biancotti J-CC, Yakir B, Clark AT, et al. (2010). Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. Cell Stem Cell 7: 521–31.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553–60.

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. (2011). A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29: 143–8.

Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, et al. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. 2: E234.

Miyoshi N, Ishii H, Nagai K, Hoshino H, Mimori K, Tanaka F, et al. (2010).

Defined factors induce reprogramming of gastrointestinal cancer cells. Proc. Natl. Acad. Sci. U.S.A. 107: 40–5.

Miyoshi N, Ishii H, Nagano H, Haraguchi N, Dewi DL, Kano Y, et al. (2011). Reprogramming of mouse and human cells to pluripotency using mature microRNAs. Cell Stem Cell. 8: 633–8

Münz C, Bickham KL, Subklewe M, Tsang ML, Chahroudi A, Kurilla MG, et al. (2000). Human CD4(+) T lymphocytes consistently respond to the latent Epstein-Barr virus nuclear antigen EBNA1. J. Exp. Med. 191: 1649–60.

Ng A, and Barker N. (2015). Ovary and fimbrial stem cells: biology, niche and cancer origins. Nat. Rev. Mol. Cell Biol. 16:625-38.

Ohnishi K, Semi K, Yamamoto T, Shimizu M, Tanaka A, Mitsunaga K, et al. (2014). Premature Termination of Reprogramming In Vivo Leads to Cancer Development through Altered Epigenetic Regulation. Cell 156: 663–677.

Okita K, Ichisaka T, and Yamanaka S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature 448: 313–7.

Okita K, Nakagawa M, Hyenjong H, Ichisaka T, and Yamanaka S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors.

Science 322: 949–53.

Orkin SH, and Hochedlinger K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell 145: 835–50.

Parrott JA, Nilsson E, Mosher R, and Magrane G. (2001). Stromal-epithelial interactions in the progression of ovarian cancer: influence and source of tumor stromal cells. Mol. Cell. Endocrinol. 175:29-39.

Pan G, Tian S, Nie J, Yang C, Ruotti V, Wei H, et al. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell Stem Cell 1: 299–312.

Papp B, and Plath K. (2013). Epigenetics of reprogramming to induced pluripotency. Cell 152: 1324–43.

Park I-HH, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, et al. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature 451: 141–6.

Pasi CE, Dereli-Oz A, Negrini S, Friedli M, Fragola G, Lombardo A, et al. (2011). Genomic instability in induced stem cells. Cell Death Diff.18: 745–53.

Peng S, Maihle NJ, and Huang Y. (2010). Pluripotency factors Lin28 and Oct4

identify a sub-population of stem cell-like cells in ovarian cancer. Oncogene 29: 2153–9

Perets R, Wyant GA, Muto KW, Bijron JG, Poole BB, Chin KT, et al. (0). Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca;Tp53;Pten models. Cancer Cell 24: 751–65.

Prat J, Belhadj H, Berek J, Bermudez A, Bhatla N, Cain J, et al. (2015). Abridged republication of FIGO's staging classification for cancer of the ovary, fallopian tube, and peritoneum. Eur. J. Gynaecol. Oncol. 36: 367–9.

Qu X, Liu T, Song K, Li X, and Ge D. (2012). Induced pluripotent stem cells generated from human adipose-derived stem cells using a non-viral polycistronic plasmid in feeder-free conditions. PLoS ONE 7: e48161.

Raaphorst FM, Meijer CJ, Fieret E, Blokzijl T, Mommers E, Buerger H, et al. (2003). Poorly differentiated breast carcinoma is associated with increased expression of the human polycomb group EZH2 gene. Neoplasia 5: 481–8.

Ramos-Mejia V, Muñoz-Lopez M, Garcia-Perez JL, and Menendez P. (2010). iPSC lines that do not silence the expression of the ectopic reprogramming factors may display enhanced propensity to genomic instability. Cell Res. 20: 1092–5.

Rao Z-YY, Cai M-YY, Yang G-FF, He L-RR, Mai S-JJ, Hua W-FF, et al. (2010). EZH2 supports ovarian carcinoma cell invasion and/or metastasis via regulation of TGF-beta1 and is a predictor of outcome in ovarian carcinoma patients. Carcinogenesis 31: 1576–83.

Reddington J, Sproul D, and Meehan R. (2014). DNA methylation reprogramming in cancer: Does it act by re-configuring the binding landscape of Polycomb repressive complexes? BioEssays 36.

Roland IH, Yang W, Yang D, Daly MB, Ozols RF, Hamilton TC, et al. (2003). Loss of surface and cyst epithelial basement membranes and preneoplastic morphologic changes in prophylactic oophorectomies. Cancer 98: 2607–2623.

Sampaziotis F, Cardoso de Brito M, Madrigal P, Bertero A, Saeb-Parsy K, Soares FA, et al. (2015). Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. Nat. Biotechnol. 33: 845–52.

Scott C, Becker M, Haluska P, and Samimi G. (2013). Patient-Derived Xenograft Models to Improve Targeted Therapy in Epithelial Ovarian Cancer Treatment. Front. Oncol. 3:295.

Singer G, Oldt R 3rd, Cohen Y, Wang BG, Sidransky D, Kurman RJ, et al. (2003). Mutations in BRAF and KRAS characterize the development of low-grade ovarian serous carcinoma. J. Natl. Cancer Inst. 19:484-6.

Siu MK, Wong ES, Kong DS, Chan HY, Jiang L, Wong OG, et al. (2013). Stem cell transcription factor NANOG controls cell migration and invasion via dysregulation of E-cadherin and FoxJ1 and contributes to adverse clinical outcome in ovarian cancers. Oncogene 32: 3500–9.

Smith J, Grizot S, Arnould S, Duclert A, Epinat J-CC, Chames P, et al. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res. 34: e149.

Somers A, Jean J-CC, Sommer CA, Omari A, Ford CC, Mills JA, et al. (2010). Generation of transgene-free lung disease-specific human induced pluripotent stem cells using a single excisable lentiviral stem cell cassette. Stem Cells 28: 1728–40.

Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, and Mostoslavsky G. (0). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells 27: 543–9.

Stadtfeld M, Nagaya M, Utikal J, Weir G, and Hochedlinger K. (2008). Induced

pluripotent stem cells generated without viral integration. Science 322: 945–9.

Steffen PA, and Ringrose L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat. Rev. Mol. Cell Biol. 15: 340–56.

Stricker SH, Feber A, Engstrom PG, Caren H, Kurian KM, Takashima Y, et al. (2013). Widespread resetting of DNA methylation in glioblastoma-initiating cells suppresses malignant cellular behavior in a lineage-dependent manner. Genes Dev. 27: 654–69.

Szotek PP, Chang HL, Brennand K, Fujino A, Pieretti-Vanmarcke R, Celso LC, et al. (2008). Normal ovarian surface epithelial label-retaining cells exhibit stem/progenitor cell characteristics. Proc Natl Acad Sci U S A 105: 12469–73.

Taapken SM, Nisler BS, Newton MA, Sampsell-Barron TL, Leonhard KA, McIntire EM, et al. (2011). Karotypic abnormalities in human induced pluripotent stem cells and embryonic stem cells. Nat. Biotechnol. 29: 313–4.

Takahashi, K & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–76

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131: 861–72.

Tada M, Takahama Y, Abe K, Nakatsuji N, and Tada T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr. Biol. 11: 1553–8.

Teif VB, Beshnova DA, Vainshtein Y, Marth C, Mallm J-PP, Höfer T, et al. (2014). Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. Genome Res. 24: 1285–95.

Vang R, Shih IeM, Kurman RJ. (2009). Ovarian low-grade and high-grade serous carcinoma: pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems. Adv. Anat. Pathol. 16:267-82.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. (1998). Embryonic stem cell lines derived from human blastocysts. Science 282: 1145–7.

Ulbright TM, Hattab EM, Zhang S, Ehrlich Y, Foster RS, Einhorn LH, et al. (2010). Primitive neuroectodermal tumors in patients with testicular germ cell tumors usually resemble pediatric-type central nervous system embryonal neoplasms and lack chromosome 22 rearrangements. Mod. Pathol. 23: 972–80.

Utikal J, Maherali N, Kulalert W, and Hochedlinger K. (2009). Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. J. Cell. Sci. 122: 3502–10.

Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, et al. (2002). The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 419: 624–9.

Vaughan S, Coward JI, Bast JR, Berchuck A, Berek JS, Brenton JD, et al. Rethinking ovarian cancer: recommendations for improving outcomes. Nat Rev Cancer. 11: 719–25.

Warren L, Manos PD, Ahfeldt T, Loh YH, Li H, Lau F, et al. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell. 7: 618–30

Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hämäläinen R, et al. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature 458: 766–70.

Wei SH, Chen C-MM, Strathdee G, Harnsomburana J, Shyu C-RR, Rahmatpanah F, et al. (2002). Methylation microarray analysis of late-stage

ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. Clin. Cancer Res. 8: 2246–52

Wei SH, Balch C, Paik HH, Kim Y-SS, Baldwin RL, Liyanarachchi S, et al. (2006). Prognostic DNA methylation biomarkers in ovarian cancer. Clin. Cancer Res. 12: 2788–94.

Widschwendter M, Jiang G, Woods C, Müller HM, Fiegl H, Goebel G, et al. (2004). DNA hypomethylation and ovarian cancer biology. Cancer Res. 64: 4472–80.

Wilmut I, Schnieke AE, McWhir J, Kind AJ, and Campbell KH. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature 385: 810–3.

Woloszynska-Read A, James SR, Link PA, Yu J, Odunsi K, and Karpf AR. (2007). DNA methylation-dependent regulation of BORIS/CTCFL expression in ovarian cancer. Cancer Immun. 7: 21.

Wong K-KK, Tsang YT, Deavers MT, Mok SC, Zu Z, Sun C, et al. (2010). BRAF mutation is rare in advanced-stage low-grade ovarian serous carcinomas. Am. J. Pathol. 177: 1611–7.

Ye L, Mayberry R, Lo CY, Britt KL, Stanley EG, Elefanty AG, et al. (2011).

Generation of human female reproductive tract epithelium from human embryonic stem cells. PLoS One. 6: e21136

Yee JK, and Zaia JA. (2001). Prospects for gene therapy using HIV-based vectors. Somat. Cell Mol. Genet. 26: 159–74.

Yoshioka N, Gros E, Li H-RR, Kumar S, Deacon DC, Maron C, et al. (2013). Efficient generation of human iPSCs by a synthetic self-replicative RNA. Cell Stem Cell 13: 246–54.

Young RH. (2007). From Krukenberg to today: the ever present problems posed by metastatic tumors in the ovary. Part II. Advances in anatomic pathology.

Young MD, Willson TA, Wakefield MJ, Trounson E, Hilton DJ, Blewitt ME, et al. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. Nucleic Acids Res. 39: 7415–27.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318: 1917–20.

Yu J, Hu K, Smuga-Otto K, Tian S, Stewart R, Slukvin II, et al. (2009). Human

induced pluripotent stem cells free of vector and transgene sequences. Science 324: 797–801.

Xie H, Ye M, Feng R, and Graf T. (2004). Stepwise reprogramming of B cells into macrophages. Cell 117: 663–76.

Zhang H, Zhang S, Cui J, Zhang A, Shen L, and Yu H. (2008). Expression and promoter methylation status of mismatch repair gene hMLH1 and hMSH2 in epithelial ovarian cancer. Aust N Z J Obstet Gynaecol 48: 505–9.

Zhang S, Balch C, Chan MW, Lai H-CC, Matei D, Schilder JM, et al. (2008). Identification and characterization of ovarian cancer-initiating cells from primary human tumors. Cancer Res. 68: 4311–20.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9: R137.

Zhang R, Calixto CP, Tzioutziou NA, James AB, Simpson CG, Guo W, et al. (2015). AtRTD - a comprehensive reference transcript dataset resource for accurate quantification of transcript-specific expression in Arabidopsis thaliana. New Phytol. 208: 96–101.

Zhao XD, Han X, Chew JL, Liu J, Chiu KP, Choo A, et al. (2007). Whole-

genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. Cell Stem Cell 1: 286–98.

Zhao Y, Yin X, Qin H, Zhu F, Liu H, Yang W, et al. (2008). Two supporting factors greatly improve the efficiency o human iPSC generation. Cell Stem Cell 3: 475–9.

Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C, et al. (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet. 7: e1002389.

10. ACKNOWLEDGEMENTS

I would like to thank my boss, Prof. Giuseppe Testa, for giving me the opportunity to work on this exciting project. I am really glad he gave me so much space to develop my own ideas and scientific reasoning.

I would like to thank my added co-supervisor Prof. Pier Paolo Di Fiore for sustaining my research and for insightful lab meetings on the subject.

I would like to thank our collaborators Ugo Cavallaro, Michela Lupia and the whole UC group for helping me out to take my first steps in the Ovarian Cancer field and for punctual support throughout the project development.

I would like to thank my internal advisor Dr. Stefano Casola, for giving me punctual insights on the project, and my external advisor Prof. Azim Surani for helpful comments on my work during these four years.

I would like to thank Vivek Das and Pasquale Laise for having performed the Whole exome sequencing, RNAseq and ChIPseq bioinformatic analyses.

I would like to thank Dr. Gilles Gasparoni and Prof. Joern Walter for the extreme availability, the insightful discussions in Saarbruecken, and for having performed the processing of samples for DNA.

I would like to thank Prof. Andrew Elefanty for providing the MIXL1-GFP hESC lines and for sharing differentiation protocols.

I would like to thank my friends and colleagues Giulia Barbagiovanni and Federico Boem for the continuous support, the exciting scientific conversations and for their friendship.

A special thank goes to my colleague Elena Signaroldi, that really supported me during these four years and especially in these last days of thesis writing.

Thanks also to the entire Molecular Medicine programme laboratory, and especially Dario and Rose Mary, for their friendship and support.

Most of all I would like to thank the entire lab of stem cell epigenetics (past members included), because they made these four years together really special.

A huge thank you is dedicated to my parents: to my mother that would have been so cheerful and proud for this achievement, and to my father without whom I would not be here writing these lines.

The final thank goes to Laura, who supports me every day and who has made the last year so special.