UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Using micropatterning of hiPSCs to induce PGC formation

Francisco Vieira Bretes

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Susana M. Chuva de Sousa Lopes, PhD
Gabriela Rodrigues, PhD

2019

ACKNOWLEDGEMENTS

A big thank you to my mother, grandparents and brother for supporting my studies, diet and being a positive influence in general. As well as my friends, for being the outside world, hope and being also a lifeline (you know who you are). A big thank you to everyone in the lab for being such nice people, hardworking and supportive people in general.
In particular, I thank Jasin for being a friend, teaching me about cell culture, PCR and being so patient too. I thank Sander for being a friend as well, talking to me about bioinformatics, showing me the basics on how to start up with transcriptomics data with Jasin and the experience he carries.
I thank Esther for being a massive help with the experiments and maintaining the cells, they wouldn´t be possible without you, and as well, being a friend. I thank Erica for the efforts on the work with the Cytoos, testing them in her project, before they could be used in mine and, as with everyone, for the niceness. I thank Christiaan Arendzen from the iPS Facility for maintaining the iPSCs and bringing them to Chuva de Sousa Lopes Lab. I thank Palmira too for being an example, including all the PhD students, Xueying, Julieta, Silwia. Not forgetting, Maaike and Ioannis.
I thank Professor Gabriela Rodrigues for being my supervisor in Portugal and being a great teacher. I also thank Professors Solveig, Octávio, Gabriel, Vitor, Bernard and Élio or heir teachings during the first year of the master.
Most of all, lastly but the reason for me being able to work in this project, Dr Susana, for her willingness for talking to me initially, accepting me in her group and being my supervisor.

SUMÁRIO

O estudo *in vitro* do desenvolvimento dos gâmetas é essencial para compreender os aspetos determinantes da fertilidade em mamíferos e em particular em humanos. Nesse âmbito inclui-se nesta tese a investigação das células precursoras deste tipo celular, as Células Germinais Primordiais (PGCs).

De acordo com estudos anteriores em ratinho, as células precursoras das PGCs (pPGCs) podem ser identificadas durante o período de gastrulação. Restrições éticas e legislativas relativas à utilização de embriões humanos em fase de gastrulação se insere impedem o estudo destas últimas com recurso a cultura de embriões humanos. Como estas restrições excluem a possibilidade de efetuar análises no período em que se estima abranger a gastrulação humana, não é possível atualmente estudar estes acontecimentos com base em observações feitas com embriões humanos.

Por este motivo recorreu-se a um método de cultura celular usando diferenciação de células humanas pluripotentes induzidas (hiPSCs) e células estaminais embrionárias humanas (hESCs). Este método permite replicar até um certo ponto a gastrulação *in vivo*. O método 2D de cultura de células desenvolvido por Warmflash et al. (2014) foi escolhido como base. Este método apoia-se na micropadronização de lamelas (com fibronectina) para obter diferenciação celular, com estimulação da via da proteína morfogenética do osso (bone morphogenetic protein4, BMP4). Tentou-se reproduzir o aparecimento dos folhetos germinativos e a sua organização em domínios distintos mutuamente repressivos, devido a moléculas secretadas por cada folheto germinativo que asseguram que células características de um não se formem nos restantes folhetos. O método de Warmflash et al. (2014) recapitula, deste modo, a gastrulação *in vitro*. No entanto, tal organização foi impossível de obter devido a cobertura sub-ótima dos micropadrões. Este grau de cobertura terá resultado de incompatibilidades entre células e a proteína constituinte dos micropadrões, a fibronectina (*FN1*). Por este motivo, um outro método baseado na cultura de células estaminais embrionárias humanas (hESCs) em lamelas revestidas com matrigel foi usado, também almejando a geração de PGCs *in vitro*. Este método produziu resultados positivos, visto que células positivas para três marcadores-chave de PGCs foram detetadas.

Estas células foram detetadas ao aplicar imunofluorescência (usando um painel de anticorpos) refinado por meio de análise bioinformática, a partir de marcadores extraídos da literatura. Estes marcadores permitiram evidenciar, por meio de imagiologia às lamelas, a presença de potenciais PGCs.

Com este método foi possível induzir potenciais PGCs, abrindo-se um novo meio para investigar os mecanismos e vias de sinalização para estudar estas células sem ter de recorrer ao uso de animais ou embriões humanos em cultura.

Palavras-chave

PGC Micropadronização hiPSC Gastrulação in vitro hESCs

ABSTRACT

Understanding PGC development is key to uncover new strategies to assist the reproduction of humans and non-human organisms. However, the study of human early gametogenesis falls under the same restrictions associated with studying human early development. This means culturing human embryos past two weeks (14 days), the estimated stage at which gastrulation should begin is out of reach under the law. This motivated the current efforts for investigating methods that replicate gastrulation *in vitro* as closely as possible. A 2D protocol developed initially by Warmflash et al. (2014) potentially fulfils this necessity according to tests performed by the authors. This allows the use of micropatterning for differentiating colonies of human induced pluripotent stem cells (hiPSCs) with supplementation of bone morphogenetic protein4(BMP4). The authors showed that colonies develop the three germ layers (ectoderm, mesoderm, endoderm and extraembryonic ectoderm). The promise that this 2D platform provides the means to recreate the cellular interactions of gastrulation prompted for the question driving this study: Can primordial germ cells (PGCs) be induced from hiPSCs by differentiating them in micropatterns?

To answer this question, a search for genes, that could be used as markers, was first conducted. Thus, several genes were selected from established marker genes and others whose validation has been less extensive in the literature. This was followed by a bioinformatics analysis was performed on a data set containing both germ and somatic cells from Li et al. Cell (2017). This data set was analysed in parallel with one other containing primed and naïve stem cells from Messmer et al. Cell (2019). In this manner, the exploration of these datasets resulted in a basic analysis, which was the source of proposed antibody combinations to detect PGCs.

After obtaining the appropriate marker combinations for immunostaining, the 2D method was tested for its suitability to induce PGCs. However, the seeding of these micropatterns with hiPSCs or human embryonic stem cells (hESCs) revealed to be technically challenging for this study. Continuous experimental difficulties with achieving uniform attachment (coverage) of hiPSCs to the micropatterns motivated searching for PGCs *in vitro* through a different method, while also relying on the previously identified marker combinations. The new method used seeding of hESCs on Matrigel coated glass coverslips in conjunction with BMP4 stimulation, as applied on the micropatterns. This protocol was then used to look for indications of possible human PGCs in culture through immunostaining of the fixed coverslips, for selected marker combinations. After imaging it was possible to observe cells with PGC morphology and co-expression of key PGC markers. Due to the resemblance of the observed cells to PGCs, it is likely that PGCs may have been generated using the proposed method of stimulation. The same that is used in micropatterns by Warmflash et al. (2014). Taken together, this thesis describes a potential method to study PGC specification without animal experimentation or resorting to human embryo culture.

KEYWORDS

PGC Micropatterning hiPSC Gastrulation In vitro hESCs

Index

List of tables and figures

ABREVIATIONS LIST

APC/C	Anaphase-promoting complex or Cyclosome
AVE	Anterior visceral endoderm
AKT1	*AKT* serine/threonine kinase 1
TF	Transcription factor
BMP	Bone Morphogenetic Protein
CCNA1	Cyclin A1
CCNA2	Cyclin A2
CCND1	Cyclin D1
CCNE1	Cyclin E1
CCNE2	Cyclin E2
CTNNB1	Catenin beta 1
CER1	Cerberus1
CDC20	cell division cycle 20
CDH1	E-Cadherin
CDH2	N-Cadherin
CDH5	Cadherin 5
CDH11	Cadherin 11
CDK	Cyclin-dependent kinase
CXCL12	Chemokine (C-X-C motif) ligand 12
CXCR4	C-X-C chemokine receptor type 4
DP	Double Positive/double positive
DPPA3	Developmental pluripotency associated 3, Stella
DKK1	Dickkopf WNT signaling pathway inhibitor 1
EMI1	early mitotic inhibitor 1

ECM	Extra-cellular matrix
EOMES	Eomesodermin
hESC	Human Embryonic Stem Cell
hiPSC	Human induced Pluripotent Stem Cell
ExE	Extraembryonic Ectoderm
FERMT2	Fermitin family member 2
FN1	Fibronectin
FNDC5	Fibronectin type III domain containing 5
h	Human
ITGA1	Integrin subunit alpha 1
ITGB1	Integrin subunit beta 1
JUP	Junction plakoglobin
KIT	KIT proto-oncogene, receptor tyrosine kinase
KITL	KIT ligand or stem cell factor
PCNA	Proliferating cell nuclear antigen
p	Primed
PECAM1	Platelet and endothelial cell adhesion molecule 1
PDPN	Podoplanin
m/hPGC	Mouse/human Primordial germ cell
P	P-value
POU5F1	POU class 5 homeobox 1
PVE	Posterior visceral endoderm
PRDM1	PR/SET domain 1
PSC	Pluripotent stem cell
MCM2	Minichromosome maintenance complex component 2
MASTL	Microtubule associated serine/threonine kinase like
NANOG	*NANOG* homeobox
LAMA2	Laminin subunit alpha 2
LEFTY1	Left-right determination factor 1

LAMA4	Laminin subunit alpha 4
LAMB1	Laminin subunit beta 1
LAMB2	laminin subunit beta 2
LAMC1	Laminin subunit gamma 1
LAMC2	Laminin subunit gamma 2
SOX2	SRY-box transcription factor 2
SOX17	SRY-box transcription factor 17
scRNA-seq	single cell Ribonucleic Acid sequencing
SSC	spermatogonial stem cell
TET1	TET methylcytosine dioxygenase 1
TFAP2C	Transcription factor AP-2 gamma
TPM	Transcripts per million
TUBA1C	Tubulin alpha 1c
TUBB3	Tubulin beta 3 class III
TSIX	*TSIX* transcript, *XIST* antisense RNA
UBE3A	Ubiquitin-conjugating enzyme E2A
WNT	*WNT* family member
XIST	X inactive specific transcript

1. Introduction

1.1 – Gametogenesis and the relevance of studying germ cell development

Reproduction is known as the means by which species propagate and maintain themselves. In the case of sexually reproducing species, this is dependent on the ability to produce gametes. These are originated from precursors, named Primordial Germ Cells (PGCs). PGCs are formed as a result of a process called specification, during gastrulation, following the blastocyst stage, by a variety of signals produced by surrounding cells. The combination of these signals and the spatial arrangement of the cells producing them determines where and how many PGCs are specified[1]. Specification is ultimately a consequence of how embryo morphologies in each animal group, or Phylum, change from the blastocyst stage onwards and more precisely, species wise [1,2].

The process of gametogenesis is not error proof or exempt from the influence of internal or external(environmental) factors, such as toxins. Among these errors, aneuploidies, DNA replication related mutations, chromosomal aberrations, epigenetic changes and mitochondrial heteroplasmy are found. Such susceptibility of genomes to alterations is at the root of fertility problems or embryonic lethality in humans. The causes of infertility are variable and affect both sexes approximately equally, but have been mostly studied in male[3–5].

Various techniques have been developed to bypass these obstacles and achieve pregnancy. These are regarded as part of an array commonly referred to as *in vitro* fertilization procedures (IVFP). IVFP require an intense hormonal stimulation in the case of females. In order to obtain multiple oocytes, from which the best quality ones are selected. The referred hormonal treatment results in very intense discomfort and stress, aside from being greatly inefficient, the effects of intense stimulation are unknown in the resulting next generation across their lifespan[6,7]. Another drawback lies in the non-existence of precise enough screening methods capable of detecting non-well characterized mutations, such as de novo mutations on germ cells[8].

Additionally, extraction of human gametes through these methods and their manipulation with the goal of producing embryos for research purposes is forbidden by regulations and ethical boundaries in most countries [9].

The use of human induced Pluripotent Stem Cells (hiPSCs) can circumvent these limitations, as they can be generated from a specific donor by conversion to a pluripotent state[10]. The main procedures trigger pluripotency through viral integration in the genome or transfection of core pluripotency genes. Once the donor's somatic cells are converted to pluripotent hiPSCs, they can be screened by nuclear and mitochondrial genome sequencing. Thus, healthy PGCs could be generated from donors carrying mutations or epigenetic alterations. Additionally, when enough knowledge from model animal studies is gathered, the creation of platforms to replicate early development closely, gains reproducibility. As such, the possibility to study the early stages for better assistance in the reproduction of human and non-human animals will be enhanced.

1.2 – Comparing model organisms: Putting the pieces together

As mentioned above, the requirements for specification of PGCs can vary among different species. In addition, this differs between mammalian and non-mammalian species. Even within mammals differences are found, for example, between rodents and other species, such as, rabbits, humans and various bovines. This means that, the signalling required to instruct cells, leading to the organization of a developing embryo, are not the same in every organism [1,11,12].

These differences stem from as early as the fusion of both sex gametes into a zygote, for example, the origin of the centrosome. This distinctive characteristic is variable inside mammals already around the time gametes meet, with the centrosome being provided by sperm cells in most cases and by oocytes in rodents[13]. Another differential aspect is the zygote's genome methylation between rodents (mice, rats) and humans or rabbits[14].

Additionally, differences continue to be observed as embryos close in on the occurrence of gastrulation events, such as, implantation timing, an event that happens earlier in the mouse (E4.5). [15]. In humans, the blastocyst grows for longer before implantation occurs(E6-12). These differences have not been connected to the regulation and time of onset of gastrulation[15,16]. Considering species-specific diversity during the early stages of mammalian development, investigating human development indirectly using a model can prove to be a challenge. Resorting to evolutionary closer organisms is practically possible, even with minuscule, species-specific differences between primates. Although, these are attributable to variability in study outputs[2,17,18].

Recent studies have demonstrated that the modes and timing of PGC specification to be largely conserved between humans, monkeys and pigs. These observations lead to the understanding that specification arises due to the events described as gastrulation in mammals [17,19]. Additionally, studies on mice, also one of the closest species, support this[20–22]. The precursors of mPGCs are specified around embryonic day(E) 6.25 as a cluster of 6 cells, that in the following 24 hours increase to about 45 cells. At E7.25, that cluster of 45 cells at the base of the allantois, located in the posterior part of the embryo, becomes lineage restricted. These are the only cells that enter the germline, which later give rise to the gametes with the genetic information transmitted to the next generation[22–24].

In humans, it is less clear at what stage of gastrulation hPGCs are specified, what are the molecular pathways involved and how many cells are specified initially[25–27]. Nevertheless, the similarities between the species mentioned priorly make those reasonable model organisms to consider when modelling the formation of functional germ (haploid) cells in humans[28]. Considering the common conclusions of previous studies, the following sections will further elaborate on the applications and aspects of stem cells, namely hiPSCs, while focusing on this period.

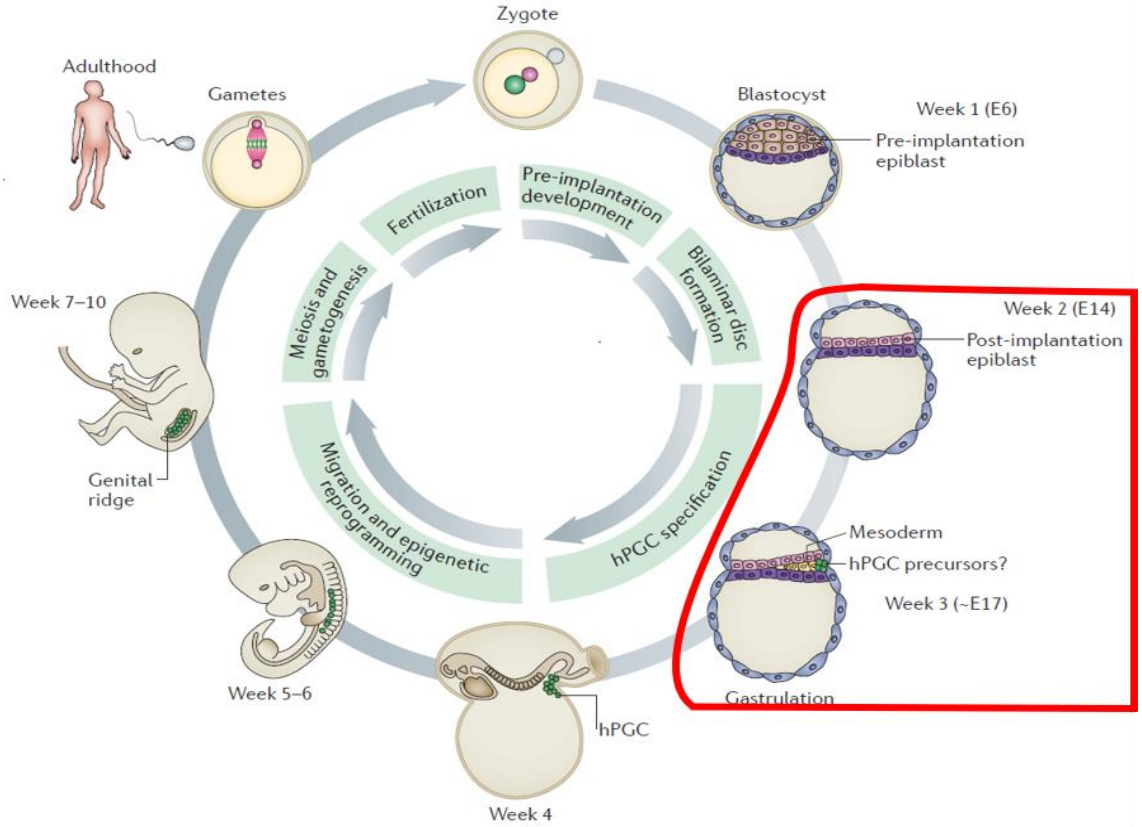


Figure 1.2. Germ cells throughout the stages of Human development, male perspective. Fertilization – The zygote becomes the blastocyst, which is composed off pluripotent pre-implantation epiblast cells. Epiblast cells give rise to all embryo proper lineages, including the germ line, that are the main objective in this study. Post-implantation – In the uterine wall, the blastocyst develops a bilaminar embryonic disc and at a later stage undergoes gastrulation. At this point, ectoderm, mesoderm and endoderm germ layers are formed. Human PGCs (hPGCs) are likely specified approximately at the stage off gastrulation. This means, posteriorly to the onset of gastrulation (roughly, embryonic day 17(E17)). However, the precise timing of hPGC specification in *vivo* is still unclear and thought to occur during 2- 3 weeks post-fertilization(wpf). At around 4 wpf – hPGCs localize near the yolk sac wall close to the allantois, and later migrate through the hindgut to the developing genital ridges. Migratory hPGCs undergo genome-wide epigenetic reprogramming, including global DNA demethylation, imprint erasure. During *in-utero* development and including adulthood, gonadal germ cells require meiosis and gametogenesis to differentiate into sperm and eggs. Towards the end of their journey, cells remethylate the genome reinstating sex-specific epigenetic marks. Thus, making the generation of a totipotent zygote upon fertilization, possible. Adapted from [27].

1.3 – Gastrulation: Specification of PGCs *in vivo*

Both in mouse as in human, the stimulus described to be the critical inducer responsible for PGC specification is BMP4. This morphogen is known to be produced in the extra-embryonic ectoderm (ExE) and in the amnion of the two species, respectively[27,29].

Mouse gastrulation has been described to start at E6.5, when the primitive streak is formed. In this species, ExE producing BMP4 leads to the formation of a cup shaped cylinder while in humans a bilaminar disc is formed[27,30]. At the same time, BMP2 and *WNT*3 are produced in the posterior visceral endoderm (PVE). BMP4, along with BMP2 and *WNT*3(through β-catenin) then induce expression of the mesodermal transcription factor (TF) T(Brachyury), whose expression is maintained from there on[31]. This, together with BMP4 triggers PR domain zinc finger protein 1(*PRDM1*) positive (+) and *PRDM1*4+ cells to differentiate in the proximal-posterior most region of the epiblast[32,33]. Thereafter, *PRDM1* binds directly to the genome to repress somatic and mesodermal genes[34]. These *PRDM1*+ cells are restricted to the posterior visceral endoderm (PVE), due to the inhibitory effect of BMP antagonists such as cerberus1(*CER1*), left-right determination factor (LEFTY1) and *WNT* antagonists. *PRDM1* along with *TFAP2C*, it´s downstream target cooperates with *PRDM1*4 to activate the transcriptional circuitry of PGCs[32,35–37]. BMP8B produced in the ExE prevents the anterior visceral endoderm (AVE) from reaching the posterior epiblast. Contributing to an environment permissive for the specification of PGCs and where BMP signalling is effective towards PGC fate, thus, ensuring PGCs are specified in the adequate location. In the absence of a fully operational AVE, *PRDM1*+ cells, are induced also in the anterior epiblast. It is not clear through which exact signalling molecules of which pathways *PRDM1* is induced in the mouse. *SOX2* is thought to contribute to this phenotype, and to PGC specification in mice or rodents in general. It is known however, that this gene is essential to mPGC differentiation and proliferation *in vitro*, an effect arising from the experimental impairment of these events [1,27,38,39].

In human, although much less is known about the specification of hPGCs, significant progress has been made[1]. It is estimated to take place inside the 2-3 weeks post-fertilization(wpf) interval, a period corresponding to Carnegie stage 7[40]. Studies in hPGCs and hPGC-Like Cells(hPGCLCs) have been able to assemble a pathway that explains how hPGCS are specified. The authors of these studies were able to do so by culturing the cell types mentioned above *in vitro* in various modes. Irie et al. Cell (2015) obtained hPGCLCs through culturing of embryoid bodies/embryoids(EBs), Sasaki et al. (2015) and Yokobayashi et al. (2017) used floating aggregates preceded by incipient mesoderm-like cell (iMeLCs) induction[34,41–43]. From these studies, it was concluded that *WNT* signalling induces expression of *EOMES*odermin (*EOMES*) on cells on the posterior side of the epiblast. Following this event, *SOX17* starts being expressed cell autonomously in cells in which *EOMES* has been induced and, before that, T. Ultimately, *SOX17* is responsible for the expression of *PRDM1*, which in turn, represses mesodermal and endodermal somatic fates and enables the germ cell fate[34,42].This signalling however, cannot be studied *in vivo*, as ethical constrains prevent any possible observations that could be made with human embryos during gastrulation[9]. BMP4 and BMP8B are also expressed in humans, being that, BMP4 signalling is restricted spatially, as in the mouse. The areas responsible for the formation of this domain are suspected to be distinct from the ones in the mouse. In the case of humans, signalling is

thought to occur similarly to rabbits[1]. The *WNT*3 signals are thought to originate from the posterior epiblast and hypoblast, whereas, BMP4 and BMP8B are expected to be expressed form an area surrounding the epiblast[27]. This interplay is thought to cause the same effect as in the mouse, that is, BMP signalling has the highest levels of effect in the posterior epiblast[27]. The exact *in vivo* signalling responsible for PGC specification is still unknown in humans. Knowing this, methods that allow the creation of a window into the stage of gastrulation are fundamental to understand the signalling at the root of specification. Similar to the narrow time window in which PGCs are specified, shortly afterwards, PGCs start migrating. At this stage a portion of the signalling is thought to remain active since specification[27,44].

1.4 – Migration

Having been specified, PGCs initiate a collective movement from the posterior region of the epiblast to the gonadal primordium (sexually undifferentiated gonad). The migration of PGCs has been studied in mouse and human, having been divided into three stages([45], 1989). These are named: Separation Phase(early), the transition of the cells from gut epithelium to mesentery ECM; Migratory Phase(mid), the span cells take to reach the crest of the bipotential and lastly; the Colonization Phase, when cells move into the gonadal primordium(late)[46–48]. Immediately after mPGCs are specified at embryonic(E) day 7.25, DNA demethylation and imprint erasure have not been initiated yet. Not long after, PGCs initiate a collective migration, using the epithelium of the hindgut and dorsal mesentery to move towards the gonadal ridges at E8.0-8.5. At this point in time, mPGCs undergo reacquisition of pluripotency and genome-wide epigenetic reprograming. This reprogramming is accompanied by proliferation of PGCs, which, in the human homolog, occurs from approximately 4wpf[44]. This epigenetic reprograming is extensive as all imprints are erased along with gene expression repressive marks(methylation), with the first occurring around colonization[49]. In mouse, as in human, a recent study on hPGCs by Eguizabal et al. (2016) noted that H3K27me3(repressive) increased from 6 weeks onwards, while H3K9me2(activating) decreased in females, contrary to their observations in male cells for the latter histone modification[50]. On the other hand, Gomes Fernandes et al. (2018) were able to characterize migrating hPGCs at 4.5 weeks, corresponding to Carnegie stage 12-13. This latter study tested a comprehensive panel of markers showing that H3K27me3 is found in a perinuclear location/coating the nuclear envelope[51].

Around this age, 5-methylcytosine(5mC) and Ten-Eleven Translocation methylcytosine dioxygenase enzymes (TET enzymes), 5-hydroxymethylcytosine(5hmC) are detectable. DNA methylation maintenance is concomitantly suppressed through repression of novo DNA methyltransferases (*DNMT3A*, DNMT3B), and UHRF1 in hPGCs, in addition to passive demethylation with each cell division due to downregulation of *DNMT3A* and DNMT3B[44,52–54]. During the demethylation process, the genome becomes susceptible to disruption of chromatin reconfiguration due to random repetitive transposable element (TPE) reintegration. TPEs are known to be repressed by DNA methylation, but they can also be degraded by small non-coding RNAs (sncRNA). P-element induced wimpy testis-like (PIWIL) are the sncRNAs in question and they are the main silencers of TPEs (Yang and Wang, 2016a). The silencing of TPEs is the result of PIWIL molecules complexing with piRNA silencing complexes(piRISC) of the Argonaut family. An association that causes degradation by cleavage of target upon

complexing[55–57]. Three paralogues are found in mice, known as *PIWIL1*, 2 and 4, while humans contain an additional one coded in the genome, *PIWIL3*[58]. Along with these markers, a study using other markers found in the literature, such as *TFAP2C*, *POU5F1*, *PDPN*, *ALPL* and *PRDM1* also proved to be reliable. Additionally, one other study had previously observed that expression of some of those genes seems to be maintained at noticeable levels while PGCs migrate[44]. Migration specific markers integrating signalling pathways previously documented were validated as well, such as, integrinα6(*ITGA6*), P-element-induced wimpy testis Like 4(*PIWIL4*) and KIT. As well as, *CD38*, a surface marker previously used to sort hPGCLCs[34,51].

It has been suggested that extracellular matrix (ECM) molecules take part in aiding and directing PGCs in their migration. Namely, fibronectin and laminin[59,60]. These proteins have been detected in the migratory path of PGCs, on the surface of somatic cells that PGCs migrate on and on cellular processes of PGCs[59,61,62]. In the mouse, migration occurs along a discontinuous basement membrane underlying the coelomic epithelium[63], which is coated by type IV collagen, strongly adhesive laminin[64,65], proteoglycans[66], and fibronectin[59]. Other proteins such as, proteoglicans, laminin and tenascin are also present throughout the migratory route. The α1 chain of laminin is found in all segments of the migration route and in all stages. This chain/subunit is also found to be the most adhesive one for mPGCs[67]. This points to a role of ECM in directing PGCs towards the genital ridges and to the bipotential gonad[68].

mPGCs also express cadherins (CDHs)[69], platelet endothelial cell adhesion molecule-1 or PECAM-1[70], α6, α5, α3 and β1 Integrins (ITGs)[71]. Embryos show expression of integrin subunit α3, α6, αv and β1 during mPGC migration. However, the absence these subunits cause no serious defects in the migratory phenotype. Integrin receptor subunits αv and β3 regulate the binding to ECM molecules found to be in the route of mPGCs, such as fibronectin and laminin[72,73]. Mouse embryos lacking the β1 integrin subunit arrest early in development, shortly after implantation[74,75]. Although the expression of these genes has been previously linked to PGCs, very few have been validated. One such gene is *ITGA6*, for having been used successfully by few studies using human fetal material or human cells[34,51]. Taking these studies together, it is possible to notice that there is a lack of information concerning the expression of most of these genes in human.

Species Embryonic Comparison Timeline

Carnegie	Stage	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Human	Days	1	2-3	4-5	5-6	7-12	13-15	15-17	17-19	20	22	24	28	30	33	36	40	42	44	48	52	54	55	58
Mouse	Days	1	2	3	E4.5	E5.0	E6.0	E7.0	E8.0	E9.0	E9.5	E10	E10.5	E11	E11.5	E12	E12.5	E13	E13.5	E14	E14.5	E15	E15.5	E16
Rat	Days	1	3.5	4-5	5	6	7.5	8.5	9	10.5	11	11.5	12	12.5	13	13.5	14	14.5	15	15.5	16	16.5	17	17.5

Figure 1.4. Correspondence between the developmental stages three mammal species. Correspondence of embryonic days enables a comparison on when events take place, such as, implantation. Additionally, this table intends to give the reader a bridge to understand why certain ages of model organisms are used as a proxy to comprehend those events in human. Adapted from[76].

1.5 – In vitro generation using hESCs and hiPSCs

Protocols aiming for the differentiation of PGCs *in vitro* occurred only in recent years with more success by using embryonic stem cells(ESCs) from mouse cultured in medium with two inhibitors(2i) through TGFβ (BMP) pathway stimulation, among others[19,34,43]. Mouse ESCs(mESCs) are known to be heterogeneous, even when cultured with two inhibitors (2i) and human ESC´s are also documented as somewhat heterogeneous[77–79]. These cells can fluctuate between states of naive or primed pluripotency, giving rise to different populations within the culture. Additionally, hESCs are a lot more heterogeneous than mESCs[80]. However, the 2i condition is still very effective in keeping hESCS homogeneous[81,82].

Ultimately, the efforts behind these protocols have managed to generate PGC-like cells, also known as PGCLCs. This designation arises from transcriptomic data comparisons to cells extracted from fetal gonads, human PGCs (hPGCs), as well as, mPGCs[44,83]. Most protocols mentioned to this point in this section are based on the stimulation of embryos or cells in culture using BMP4 and BMP8b. This was first described by Hayashi et al. (2011), one of the first studies to achieve PGCLC differentiation[84].

Most of the *in vitro* grown PSCs, hiPSCs or hESCs, are alkaline phosphatase (*ALPL*)+ colonies and share the expression of pluripotent regulatory genes, *POU5F1*, *SOX2*, and *NANOG*. Stem cells in vitro are generally assigned to two states according to regulatory signalling pathways and morphology. The described states of pluripotency are designated "naïve" and "primed", respectively. The naïve cell type requires leukaemia inhibitor factor (LIF) to maintain itself and forms a compact dome-shaped colony. hiPSCs, as opposed to naïve, are not dependent on LIF, since they are not responsive to it[85]. They are, however, sometimes dependent on bFGF and form larger flattened colonies. Primed cells incur in a significant extent of apoptosis when passaged as single cells. Mouse ESCs and EpiSCs in particular represent post-implantation epiblast blastocysts; both express *POU5F1* but are driven by a different enhancer[1,86]. In the case of human PSCs, pluripotency genes, such as, *POU5F1*, *SOX2*, and *NANOG*, are expressed commonly across all states(primed-naïve). These genes are also expressed by *in vivo* PGCs, as such, they do not constitute good PGC markers by themselves. Other pluripotency markers are also shared *by in vitro* cultured pluripotent cells, as is the case for *SOX2* in mPSCs, hPSCs and mPGCs. This gene is characteristic of primed hESCs, and thus can be considered to differentiate both cell types in vitro[1,85,87]. The glycosphingolipids SSEA-1 and SSEA-4 are also regarded as pluripotency markers although they are not required in order to maintain it. Although, they are not expressed in *in vivo* hPGCs[51,88].

Different studies have shown that hESCs and hiPSCs express a panel of markers in common with those of PGCs. The markers included in this common panel are composed by SSEA-4, *POU5F1*, *NANOG*, *STELLAR* (*stella*-related) , *ALPL*, and *PRDM1*, *NANOS1*, *NANOS3*, *DAZ*, *DAZL* in some ESC lines, and *C-KIT*, but not *SSEA1*, *CXCR4*, and VASA (*DDX4*) or synaptonemal complex protein 1 and 3 (*SYCP1* and *SYCP3*) markers of pre- and meiotic germ cells. ESCs and iPSCs also express markers exclusive to these types of cells. Among them, SSEA-3, tumor rejection antigen 1–60(TRA-1-60), TRA-1-81, and *SOX2* can set

hPSCs apart from PGCs, since they are not expressed *in vitro* or *in Vivo* [1,89,90]. Setting PGCLCs apart from undifferentiated or hESCs/iPSCs in the process of differentiating by sorting with antibody combinations for SSEA-1, SSEA-4 and *C-KIT*, and *CXCR4* have been used[91–93]. Using mESCs/iPSCs, PGCs can be isolated using intrinsic cell-surface markers, integrin-β3(ITGB3 gene) and SSEA1[84]. However, it remains unclear if these cell surface markers can be used to sort hESCs/hiPSCs. In addition, although cell sorting based on these cell-surface markers is useful and convenient, the expression of these markers is not exclusive to germline cells.

According to Irie et al. (2015) hPGCLCs share similarities with hPGCs in terms of gene expression, concerning genes such as, *PRDM1, NANOS3, ITGB3* and *KIT*. Other pluripotency related genes were also common in both cell types, for example, OCT4 (*POU5F1*), *NANOG* and *PRDM14*[34]. In a study by Kerr et al. (2008), germ cells were analysed for genes expressed specifically after colonization. These were shown to express stem cell markers like stage-specific embryonic antigen (SSEA)-1, SSEA-4, EMA-1, and *ALPL*. A minority of those (<1%) expressed *POU5F1*, *CKIT*, and *NANOG*[94]. On the other hand, *SOX2* is not detectable in human gonadal germ cells[38].

1.6 – Existing 2D and 3D methods

In order to study PGC development and their specification, a method that replicates gastrulation *in vitro* as closely as possible is necessary. This is due to an increasing need to develop models *in vitro* to study the processes that occur early during development, at the time of gastrulation. Using human pluripotent stem cells to mimic this developmental period *in vitro* would create an important information source. Recently, two groups have used the method of "embryoid bodies"(3D) to study early PGC formation[34,95]. However, this method consists of making a compact ball of cells that can eventually cavitate. This method is suitable to study mouse early development and recently the inclusion of several types of stem cells (embryonic stem cells, trophectoderm stem cells and trophoblast stem cells) and has led to the formation of the so-called "artificial embryos"[96,97].

In contrast to mouse embryos that are "cup-shaped", human embryos are "flat-shaped". Hence, optimizing models to mimic early development with a flat geometrical configuration would be beneficial to faithfully understand the molecular mechanisms taking place. The 2D method developed by Warmflash and colleagues provides an interesting tool to explore to understand human PGC formation[98–101]. This model describes the differentiation of cultured human pluripotent stem cells that are confined to micropatterned areas. These micropatterned areas consist of extracellular matrix proteins, such as laminin and fibronectin. Careful modulation with *WNT*, BMP and TGFβ signalling resulted in concentric radial domains of cells expressing ectoderm, mesoderm, endoderm and extraembryonic ectoderm markers. The micropatterned culture wells are used to create a platform for cells to sense each other mechanically and indirectly, using short range diffused signals (paracrine). Through this, colonies differentiated and mounted an organized response to a stimulus similar to a gastrulating embryo[102]. The addition of BMP4 to the medium initiates waves of differentiation on the periphery of colonies, which move towards the centre as the stimulus continues to be present. These colonies end up developing a primitive streak like structure after 52h, making this a good method to study events during mammalian gastrulation. Definitive endoderm marker *SOX17*,

which is also critical for PGC specification, is found to be expressed in the micropatterned colonies, after starting from PSCs[42,98,103]. The expression of *SOX17* motivated to the formulation of the hypothesis which follows. This study proposes that cells expressing both *SOX17* and pluripotency marker *POU5F1* could be differentiated using the chips. Because germ layers can already be distinguished at 48h, micropatterned chips will be seeded with cells to be cultured in a mode based on the micropattern based method[98,99].

The 2D protocol developed initially by Warmflash et al. (2014) fits these requirements according to tests performed by the authors. These authors showed that differentiating micropatterned hESCS colonies leads the to develop all three germ layers concentrically (ectoderm, mesoderm, endoderm and extraembryonic ectoderm)[98]. For this study hiPSCs and hESCs will be used. To this moment, no protocols using hiPSCs have been attempted with the goal to generate patient specific PGCs and, if successful their use can contribute to future protocols.

1.7 – OBJECTIVES

The objectives to be achieved during this project are:

1. Determine which PGC markers from the literature are supported by a bioinformatics analysis of existing Datasets containing germ cells;
2. Determine which genes of adhesion receptors and ligands are expressed by PGCs at 4-5wpf;
3. Study which epigenetic factors, are likely to contribute to PGC phenotype;
4. Determine if other novel markers characterize the germ cell fate.

Some of the information on PGC gene expression comes from human embryos analysed at the single cell level. Such analysis has been performed on embryos as young as 4wpf, which corresponds to an age when PGCs are migrating[104]. Hence, the possibility to investigate such data will be highly informative and requires the analysis of data describing those ages or earlier.

The analysis of a dataset constructed by Li et al. Cell (2017) and markers for PGCs gathered by reviewing literature on the development of this cell type will be evaluated for reliability. On the other hand, the analysis of a dataset by Messmer et al. Cell (2019) will allow comparison of the cell types in Li et al. Cell (2017) with PSCs, primed or naïve. Genes will be discarded if their expression is not consistent in a specific group of samples (germ or somatic cells) or reinforcing them as candidates to detect PGCs if the opposite is verified. From the first dataset, the earliest ages will be considered and chosen as representative to fundament the choice of markers[87,105].

By resorting to these methods, it is expected that the information obtained can contribute to the present knowledge on the early stages of gametogenesis. It is also expected that by doing so, new directions in medical assistance to reproduction can be uncovered.

2. Materials and Methods

2.1 scRNA-seq dataset analysis

The single cell RNA-sequencing(scRNA-seq) datasets analysed in this study to extract gene expression information were Log2 transformed versions (TPM/10+1). Where TPM (transcripts-per-million) values were obtained as the number of unique molecular identifiers (UMI) of each gene divided by all UMIs of a given cell. The resulting value was then multiplied by 1,000,000.

2.1.1 scRNA-seq dataset analysis: Germ cells vs somatic cells

This dataset and the UMI counts data frame (DF) were obtained from the Li et al. (2017) study Github page: http://github.com/zorrodong/germcell. Raw sequencing data can be downloaded from accession number EO: GSE86146. TPM values were divided by 10 because the UMI number of most of the single cell samples did not reach the order of 1,000,000 transcripts. Which was a criterion for ruling out cells according to the authors, no further explanation was disclosed by these authors[105]. Annotation to complete the information displayed about gene expression in this study was produced from information e*XIST*ing in the dataset.

2.1.2 scRNA-seq dataset analysis: primed cells vs naïve cells

In addition, a scRNA-seq composed of naïve, primed cells and an intermediate population with characteristics of both from Messmer et al. (2019). Raw counts were normalised as in the Li dataset. This dataset was obtained as raw counts through the accession number: ArrayExpress: E-MTAB-6819. The original code that was obtained from the Github page: https://github.com/MarioniLab/NaiveHESC2016 and adapted to this study. Furthermore, this data, with this composition of cells was chosen with the intent of comparing the epression of established PGC markers in PSCs. For the Messmer et al. (2019) data set, annotation was produced from data table 3 and is referred to in the text, including figure legends, as metadata[87].

The open source software R, version 3.5.3, was used to perform all analyses on the datasets. Packages used to wrangle and prepare the data for plotting gene expression in heatmaps included readxl, dplyr, tibble and R base functions. Function chosen to output the expression data was Pheatmap, which received as inputs: gene vectors, to select what expression data of which genes to include; and annotation from a DF generated separately with additional information concerning each cell, when further distinction between cells was necessary.

2.2 hiPSC colony growth analysis

All statistical analysis concerning size of colonies grown in culture plates were performed in excel, including two-way ANOVA. Means and standard deviation were also calculated using the same software.

2.3 Culture of hiPSCs

Daily maintenance – Cells were refreshed with 3ml of TeSR-E8 culture medium and checked with an inverted microscope for any visible contamination and differentiation.

Passaging of cells – New 6-well culture plates were prepared in the previous working day, by adding a cold solution of Laminin-521 and placing them at 4^0C until they are needed. Preparation of the laminin solution was made as follows. First, 950 µl per well of PBS+/+ at 4^0C is added to a tube, followed by 50 µl per well of Laminin-521 and homogenization inside the flowhood. Once homogenized, 1ml of Laminin-521 solution was added to the wells intended for use and the plates were placed at 4^0C (cold room). All steps required to make the Laminin-521 solution were performed using ice cold pipette tips to avoid Laminin-521 polymerization inside and assure the correct concentration for use. On passaging days, the Laminin-521 plates were put in the incubator (37^0C) for at least 2 hours for Laminin-521 to polymerize on the bottom of wells. While this happened, in the flow hood, culture plates were observed for suitable colonies. Afterwards, the culture medium was replaced with Gentle Cell Dissociation solution and left for up to 4 minutes or until small orifices were observed in the colonies. Non-suitable colonies, meaning ones with differentiation were removed with a pipette tip during this period. Any colonies too small were left untouched. To stop the effect of the dissociation solution, this was then replaced with 1ml of culture medium. At this point, suitable colonies were scrapped off using a pipette tip and left shortly in the wells while their destined wells were prepared. Lastly, the excess of non-polymerized Laminin-521 was taken out, replaced with 1ml of medium and 1ml of cells was then seeded after a soft resuspension. This protocol was repeated every 3 days.

2.4 Differentiation of hiPSCs and hESCs

The main method used in this project relied on applying extra cellular matrix (ECM)-coating layer that enables cell attachment only inside a defined and uniquely shaped area that repeats itself. This forms a pattern on the culture surface, in this case, a coverslip meant to lay on a well. As an alternative, ready to use micropatterned chips are commercially available from the CYTOO company. These were chosen instead, coated with circular fibronectin shapes varying in diameter(80-1000µm), to perform the experiments on micropatterns.

Cells were seeded using steps based on the ones developed for hESCs, by Warmflash et al. (2014) and described in detail by Deglincerti et al (2016). Starting with the making of a small clump suspension from cells ideally at or near confluence (Fig. 2.1). Using such an approximate concentration, cells were seeded in the micropatterns. When these conditions were met, cells were differentiated through mTeSR1+5%Penicilin/Streptomicin(mTeSR1) culture medium with BMP4 (50ng/ml). After approximately 72 hours, a span of time considered to be sufficient for differentiation, the cells were fixed with 4% paraformaldehyde (PFA) and immunostained.

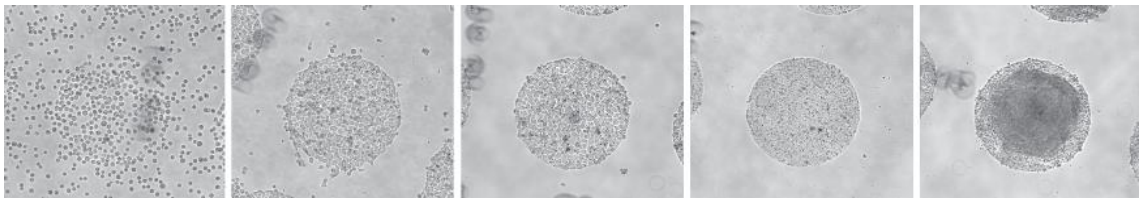


Figure 2.4. Differentiation of hiPSCs and hESCs on micropatterns. Expected progression of coverage of the micropatterns by hESCs and hiPSCs fom the moment of seeding of Cytoo chips. Images correspond to, from left to right: 10 minutes, 2 hours, 2,5hours, 20hours and 68hours after seeding (from Deglincerti et al. (2016)) [99].

2.5 Differentiation of hiPSCs and hESCs: Micropattern seeding and differentiation

Cell lines used:

WIS1 – Primed hESC line obtained from the Hanna lab, Weizman Institute.

Cl002 – Primed hiPSC line obtained from the LUMCs iPS Facility, led by Christian Freund and Harald Mikkers.

Micropatterned chips used:

Cytoo Arena micropatterned chips – Manufactured by and ordered from the Cytoo company.

Maintenance and passaging

The WIS1 cell line was passaged once a week for maintenance, with a ratio of 1:1, 1:10 when passaging to 12-well plates with coverslips and 1:6 when passaging to Cytoo chips. WIS1 cells were kept on Matrigel while on maintenance wells and for differentiation experiments on coverslips.

Cl002(Control 2) cells were kept on 6-well plates coated vitronectin, as maintenance, by the iPS Facility. These cells passaged according to a 1:6 ratio to full new 6-well plates, from 1 6-well well and were then received for further maintenance or experiments.

Part I: Cytoo chips

First, Cytoo chips were kept in 6ml PBS+\+, in a small petri dish, until cells were added. Afterwards, 2ml of mTeSR1 medium from the refreshing cells in the previous day from each well was removed. 1ml PBS+\+ was added for washing and replaced by 1ml Gentle Cell Dissociation Solution. This solution was left on the well for 3 minutes and 1ml of mTeSR1 medium was added to each well, wells were then scrapped using the cell scrapper. After this, hESCs and hiPSCs were resuspended by pipetting up and down. 300µl of cell suspension was added to each Cytoo chip, carefully, in order to create a dome with the shape of the Cytoo. Once the dome was established and starting in the areas not covered by the chips and making sure these areas contact was made with the dome slowly, preferably through the formation of multiple contact points (objective: decrease cell disturbance). mTeSR1 medium was added, totalling 2ml per dish. The 2ml were left in the dish for 5h and then refreshed with mTeSR1 medium + BMP4(50ng/ml).

Part I: WIS1

Cells were first passaged using the same method as the chips. Differences were related to the coating and amounts transferred to each well. WIS1 cells were passaged from matrigel coated 6-well plates to 12-well plates with matrigel coated coverslips, in 100µl of cell suspension to recipient wells with 900µl.

Part II(D3B4): Cytoo chips and coverslip 12-wells

Finaly, the cells were left to differentiate for 3 days in the medium just mentioned. This period of differentiation was named D3B4 and three conditions were derived from its combination or not with 24h of cell growth. Refreshment off differentiation medium was performed daily, at the same time, except when this period encompassed weekends.

Part II(24h+D3B4): Cytoo chips

The 24h growth period was performed prior to differentiation, which was conducted exactly as in the condition described above.

Part II(1wk*FN1*+D3B4): Cytoo chips

Cells spent one week on a fibronectin coating for adaptation, before being passaged to the chips.

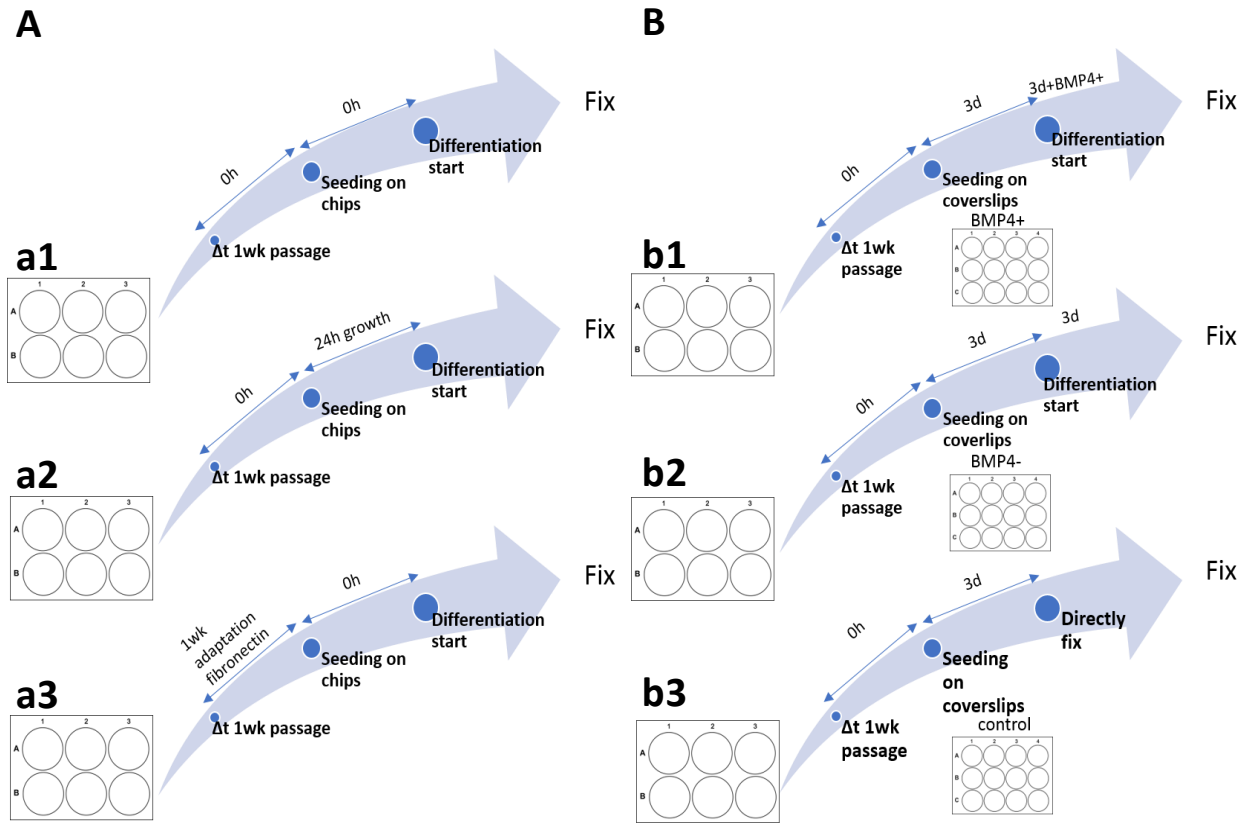


Figure 2.5 Conditions that hiPSCs and hESCs were subjected to, prior to immunofluorescence. Graphical version of the protocol decribed for (A) Cytoo chips, with the three conditions tested: a1, D3B4 condition; a2, 24+D3B4 condition and a3, 1wk*FN1*+D3B4. Using CL002 and hESC WIS1 cells. Graphical version of the protocol decribed for (B) hESC WIS1 cells, with the three conditions tested: b1, 3D+3DB4 condition; b2, 3D+3D condition and 3D condition in b3.

2.6 Immunofluorescence

A panel of antibodies described previously by this lab[51] as well as other labs was used[27,44,51]. Cells were be fixed for 20 minutes at room temperature (RT) in 4% PFA. Followed by permeabilization for 8 minutes at RT, blocked in 1% bovine serum albumin (BSA) in PBST (blocking solution) for 1 hour and incubated with a combination of primary antibodies diluted in blocking solution (BS) overnight(O/N) at 4°C in a humidified chamber. In the next day, the cells were washed with BS for 5 minutes and incubated with a combination of secondary antibodies diluted in BS O/N at 4°C in a humidified chamber. After counterstaining with DAPI in PBST (1/1000), the cells were mounted using prolongold antifade reagent.

Table 2.6.1. Primary antibodies used for immunofluorescence on coverslips seeded with hESCs.

PRIMARY	Host	Catalog number	Manufacturer	Dilution Factor
Goat Anti-*POU5F1*	Goat	sc-8628	Santa Cruz Biotechnology	1:200
Mouse Anti-*PDPN*	Mouse	ab77854	LifeSpan Sciences	1:100
Goat Anti-*SOX17*	Goat	AF1924	R&D Systems	1:200
Mouse Anti-*TFAP2A*	Mouse	sc-12726	Santa Cruz Biotechnology	1:200
Rabbit Anti-*TFAP2C*	Rabbit	sc-8977	Santa Cruz Biotechnology	1:200

Table 2.6.2. Secondary antibodies used for immunofluorescence on coverslips seeded with hESCs.

SECONDARY	Alexa fluorophore	Host	Catalog number	Manufacturer	Dilution Factor
Donkey Anti-Goat	488	Donkey	A11055	Invitrogen	1:500
Donkey Anti-Mouse	594	Donkey	A21203	Life Technologies	1:500
Donkey Anti-Goat	488	Donkey	A11055	Invitrogen	1:500
Donkey Anti-Mouse	594	Donkey	A21203	Life Technologies	1:500
Donkey Anti-Rabbit	647	Rabbit	A31573	Invitrogen	1:500

2.7 Imaging

Image acquisition was performed using a Leica DM6B-Z widefield fluorescence microscope with ColourProc, a LUMC-made software (LUMC, Leiden, The Netherlands) coupled to a Coolsnap Myo CCD camera. For image analysis, all images were saved in TIFF file format and processed using Fiji (version 2.0.0-rc43/1.5k), a software made by Schindelin et al. (2012)[106]. Every cell stained in all the used channels, that gave of signal from the predicted structures and contains a DAPI+ nucleus was considered a potential PGC.

2.8 Statistics

Colony diameters were tested for differences between passages using a two-way ANOVA in excel. $P < 0.05$ was considered significant.

3. Results

3.1 – Expression comparison of genes commonly associated PGC with PGC development scRNA-seq analysis

In order to ascertain which genes, among the ones described in previous studies can be used reliably to detect hPGCs *in vitro* (see Specification and in vitro generation, section 1), gene expression data of PGCs was plotted. From the genes described by studies in section 1, different lists of genes relating to the processes linked to stages of events defining the biology of PGCs (ex: specification and migration) were compiled. Given the scope of the Li et al. (2017) dataset, the earlier ages whose expression data was collected were chosen. These ages comprised the period between 4 and 5wpf, being that 4wpf corresponds to males (only) and 5wpf to females only. These served as proxy to sex differences, previously described commonly expressed characteristics in the 4-5wpf developmental span. These ages are also the ones closest to the probable interval for hPGC specification, that were collected from whole embryos and were analysed at the single cell level. This makes this dataset the Li et al. (2017) dataset the only one in e*XIST*ence to contain data pertaining the *in vivo* and to approximate so much on specification events. In addition to this, based on a prior comprehensive analysis of a 4.5wpf embryo by this lab, a preliminary gene set panel was arranged. The intent of these was to provide a starting point to evaluate how many remain afterwards and can be considered from a standpoint, as candidates do detect PGCs.

To further narrow down these gene combinations from the literature, gene expression was analysed using a Log2 normalized dataset from Li et al. (2017). Moreover, these combinations were refined using the Messmer et al. (2019) dataset using the same normalization method[105,107]. The expression of these genes is compared in the plots.

On all plots shown the values coded by the colour scale are in log2 normalized transcripts per million (from here onwards, TPM). These values were divided by ten to account for low abundance transcripts and referred to by the same unit, TPM.

3.1.1 – Most established PGC markers are not as consistent as previous studies claim

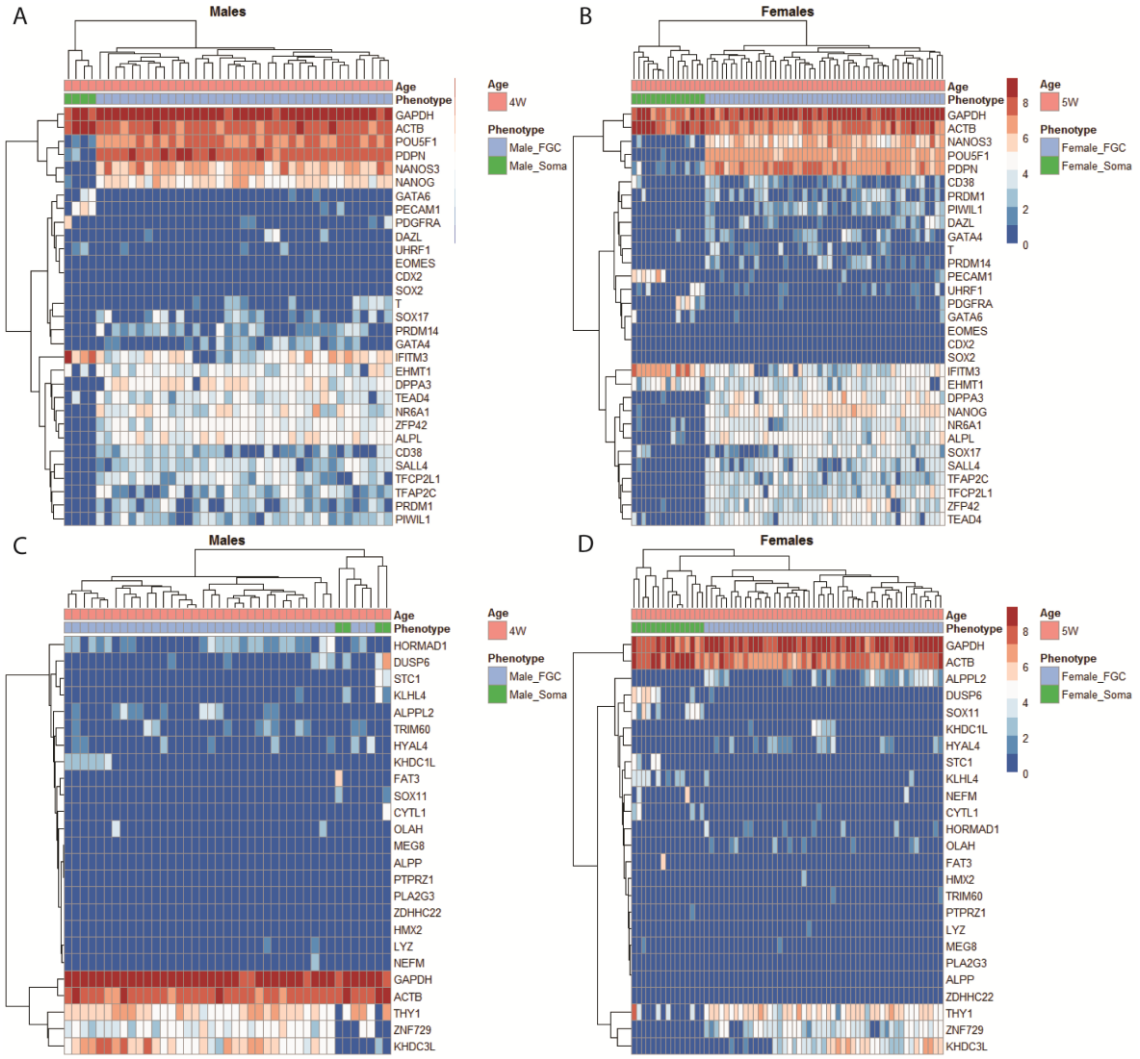


Figure 3.1.1. Heatmap plots of genes expected to identify PGCs, early development and of naïve plus primed pluripotency state genes in germ and somatic cells. (A, B) Log2 normalized transcripts per million (TPM◄TPM/10+1) of PGC gene set at 4 and 5wpf, respectively. (C, D) Genes characteristically expressed in naïve and primed PSCs compared to the initial gene set at 4 and 5wpf, respectively. Cell type labels were made according to metadata constructed from the dataset. Scale refers to relative expression values with blue as lowest expression and red and highest expression. Euclidean clustering was performed according to gene and to sample.

Upon comparing the expression of genes linked to a pluripotent phenotype in combination with genes linked with PGC specification, two cell types are distinguishable. A separate clustering branch is assembled specifically connecting somatic cells, arising from similar gene expression. The same is verified throughout all germ cells, as they are also connected by a large branch, that clusters them together. This branch, however, contains far more sub clusters due to expression variability between germ cells. The genes are also organized into clusters, due to similar expression patterns throughout the cells and differences between groups of genes that are differently expressed. This is exemplified by *NANOG*, *POU5F1*, *NANOS3* and *PDPN*, which show a large difference between somatic and germ cells (Fig 3.1.1A, BB), being more strongly expressed in the latter. Genes like *SOX2*, *GATA6* showed little to no expression on either analysed age. The first, *SOX2*, is completely absent in

all cells, while the latter gene shows a more elevated expression in 50% of somatic cells at 4wpf and noticeably lower values in two germ cells.

PRDM1(Fig 3.1.1A) and *SOX17* (Fig 3.1.1A, B) were among the genes here shown to have a greater variability in expression levels at these ages. Germ cells within the lower left cluster have, on the other hand, more homogeneous expression levels. Only divided by weaker expressing genes, on the lower sub-branch in that branch. Although these genes were mostly or only expressed in germ cells. *PDPN* and *POU5F1* show the highest levels of expression and are also the most homogeneous. Their expression is not entirely restricted to germ cells, as denoted by the lighter shades of blue from soma cells at 4 and 5wpf, although expression is much lower in these. In addition, they are clustered together by the dendrogram on the left.

When the clustering is combined with sample annotation a signature for germ cells becomes more apparent. This is based on the grouping of cells with similar gene expression profiles (top of plots) the grouping genes with similar expression profiles (left of plots) Of note, this expression of *POU5F1* and *PDPN* in soma is not at all close to the values in germ cells (Fig 3.1.1A, B). Lastly, T can be noticed as a highly heterogeneous and somewhat weakly expressed gene in some positive cells. This type of pattern is also present in the expression of late PGC marker *DAZL*. This is reflected by other marker genes, namely, *UHRF1* and *PDGFRA*. Due to the heterogeneity shown by other genes described to be markers, further analysis is requiredThis is in part suggested by shared expression of genes here pointed as PGC specific in previous reports on PSC markers[27].

Several protocols so far have obtained PGCs while having started from pluripotent stem cells (PSCs), which is a shared objective with this study. The ability to distinguish PGCs from hESCs or hiPSCs is a requirement, as the hiPSCs will be used for testing PGC detection after 72h of differentiation. From the naïve and primed genes plotted alongside the literature selected panel of PGC markers (Fig. 3.1.1A, B), most are not consistently expressed in hPGCs (Fig. 3.1.1C, D)[107]. *HORMAD1*(naïve), *KHDC3L*(naïve), *DUSP6*(primed) and *ZNF729*(primed) are exceptions at 4wpf in this aspect, albeit, the former is less homogeneous. However, only *ZNF729* and *KHDC3L* maintain specificity in the 5wpf (female) cells (Fig. 3.1.1D), while *HORMAD1* has lost it. Although *KLHL4*(primed), *STC1*(primed), and *DUSP6*(primed) were also described as stem cell markers by the latter mentioned study. These genes show higher levels in somatic cells, that are not visible on the few germ cells which show expression. Another primed gene, *CYTL1*, is more ubiquitous in comparison to the other genes that characterize the same state (Fig. 3.1.1C, D). Although this expression of primed genes is noticed mostly on somatic cells, as opposed to the naïve genes. The latter is characteristically expressed a relatively less differentiated cell type.

3.1.2 – Most PGC genes are not restricted to one pluripotency state. Nor are all pluripotency state markers restricted to their respective states.

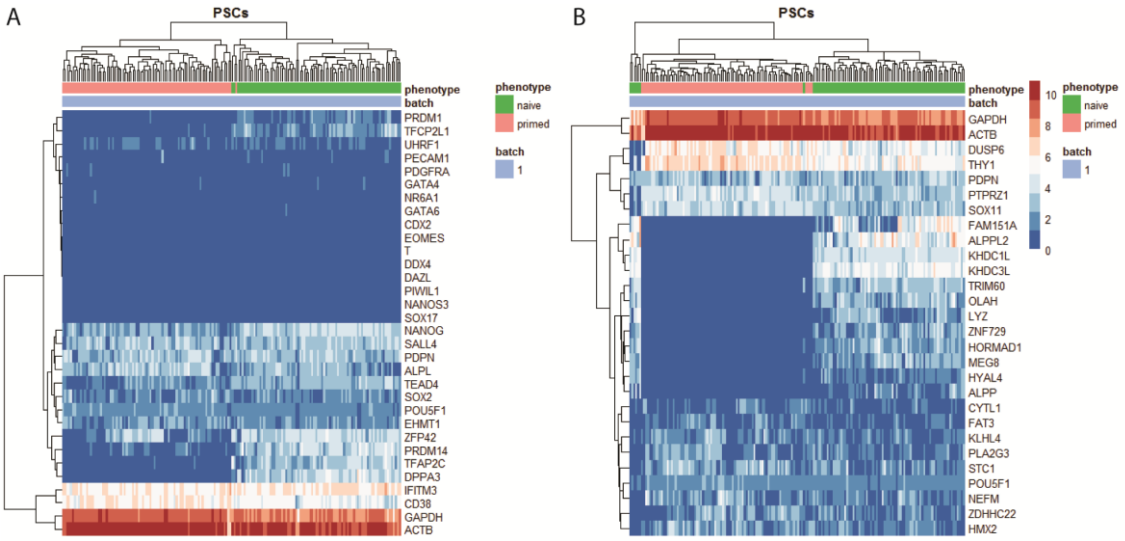


Figure 3.1.2. Heatmap plots of genes encoding proteins established as PGC /early development markers and PSC genes. Log2 normalized TPM (TPM/10+1) of (A) PGC and early development markers, ECM plus adhesion genes (B) Genes connected to the naive and primed states of pluripotency by Messmer et al. (2019). Cell type labels were made according to adjusted metadata from the dataset. Euclidean clustering was performed according to gene and to sample.

As mentioned in the previous paragraph, to achieve distinction between pluripotent stem cells (PSCs, hESCs and hiPSCs) and PGCs, confirmation on whether PGC markers are expressed in the former cells is crucial. Naïve and primed cells from Messmer et al. (2019) using PGC markers and germ layer marker genes together were compared (Fig 3.1.2A and 3.1.2B). Among the genes compared just above, only *TFAP2C*, *DPPA3* show complete or almost complete specificity for naïve cells. Also, in the same gene cluster, *PRDM14* and ZFP42 show still, reduced specificity for naïve cells in comparison with the other two genes. With the expression of the former being more intense in naïve cells, relative to primed. The latter loses distinction power greatly, due to similar expression levels in many of the primed cells. In the cluster above the one containing *DPPA3*, no gene could be seen being expressed only or mostly on one cell type. Instead, these genes were observed to be expressed in almost every cell, while showing no prominence on either type. Less noticeably, *PRDM1* and *TFCP2L1* are expressed in more naïve cells than primed. All other genes did not show any obvious or very slight difference between the two pluripotency states (Fig. 3.1.2A).

On the other hand, genes characterizing the states of pluripotency show a clear distinction between the naïve and primed states. This can be noticed by absence of expression of primed genes in naïve cells and greater expression levels in primed cells (Fig. 3.1.2B). Primed genes, however, are not expressed exclusively or with a large difference to naïve cells. At most, some like *DUSP6* and *THY1*, which cluster together, appear to have relatively higher levels of expression. The cells shown belong to a single batch, since the other batches presented issues concerning the expression of *POU5F1*. Particularly, absence of expression from many cells and inconsistency of expression levels in *POU5F1*+ cells.

Taken together, these results from PSCs (Fig. 3.1.2) show that only two PGC markers (Fig. 3.1.2A) are specific for hPGCs, by themselves (Fig. 3.1.1 and 3.1.2). The same is seen for either pluripotency state, due to many genes not being expressed in one cell type only (Fig. 3.1.2B).

3.1.3 – PGCs multiply while migrating: hPGC proliferation is asynchronous

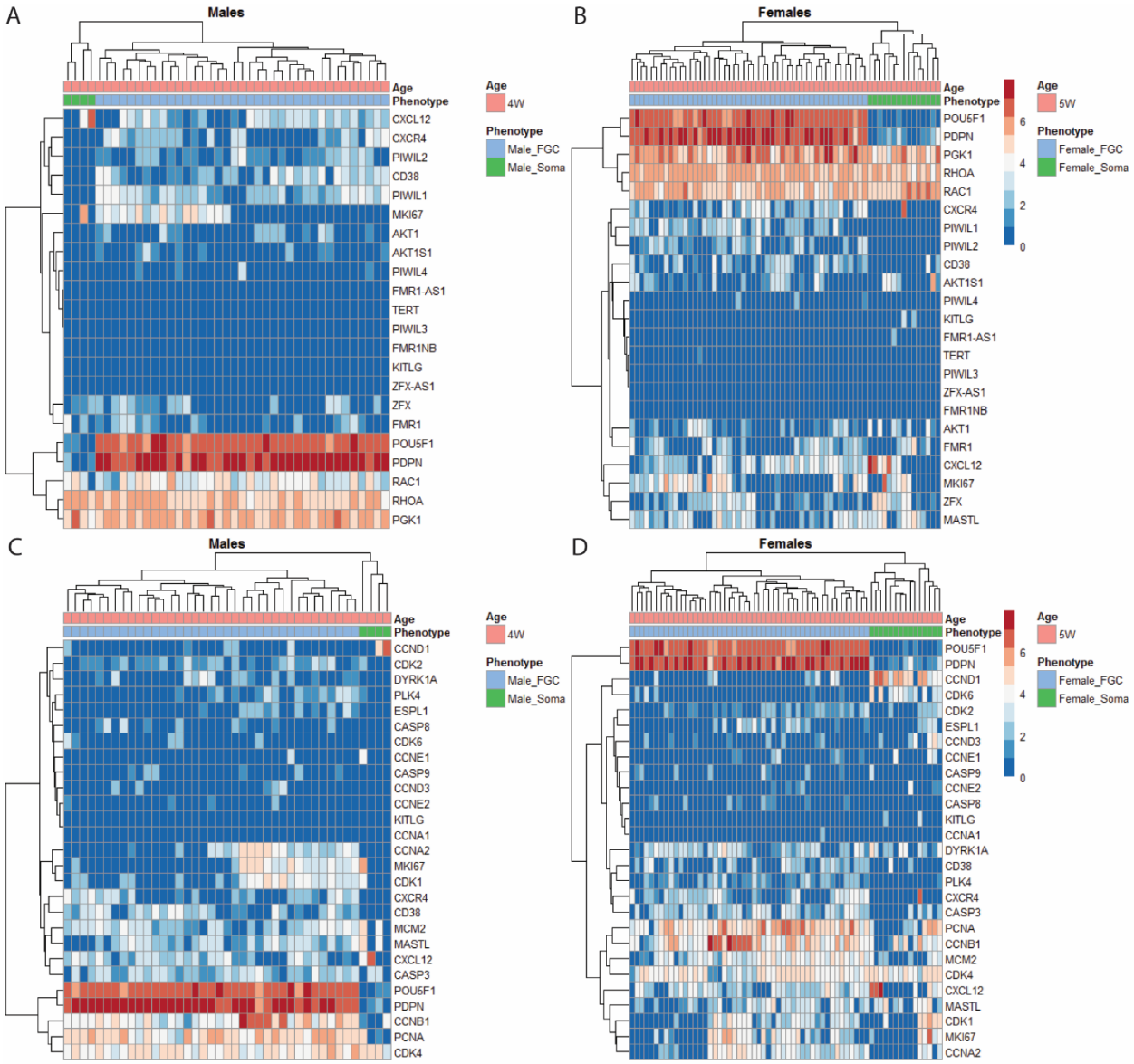


Figure 3.1.3. Heatmap plots of genes encoding proteins playing a role in cell migration and proliferation. Log2 normalized TPM (TPM/10+1) of (A, B) Genes connected to migration at 4 and 5wpf, respectively. (C, D) Cell cycle genes reprogramming associated at 4 and 5wpf, respectively. Cell type labels were made according to metadata constructed from the dataset. Euclidean clustering was performed according and to sample.

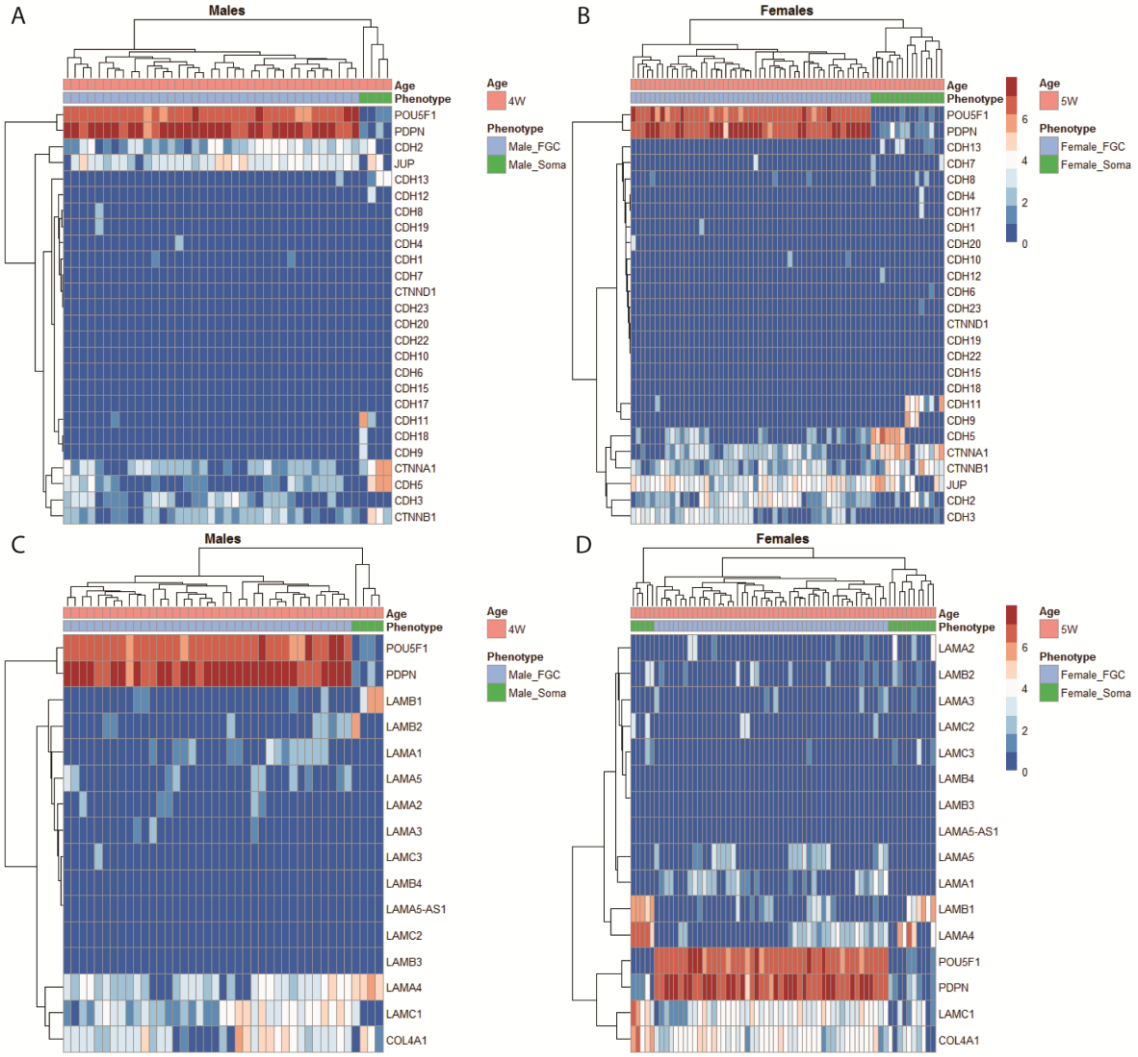
Genes connected to migration (Fig 3.1.3A, B), such as *RHOA*, *RAC1*, *CXCL12*, *CXCR4* and *KITLG* did not show any specificity according to cell type. This is observed although the ages plotted contain hPGC migration activity. Others connected to germ line fate (*CD38*, *AKT*), X chromosome (*FMR1*, *ZFX*, *PGK1*). *TERT* shows almost no expression, that is also characteristic of mPGCs (Fig 3.1.3A, B) [108]. Of the TPE suppressor genes, *PIWIL1* and 2 are

the ones whose expression values can be found inside the germ cluster alone[56,109,110]. *MKI67* seems to be specific to a group of germ cells, clustering them together, undistinguished by other characteristics in this gene's expression. In these cells, expression is also not distinctive or homogeneous. Furthermore, this last gene is accompanied by others that place the *MKI67*+ cells in different phases of the cells cycle (Fig. 3.1.3).

For this reason, these other genes, such as cyclins and cyclin dependent kinases were plotted along with *PDPN*, *POU5F1* and others previously linked to germ cells in the ages in question. All cells *MKI67*+ do not place in G0 or G1, but in any point during the S-M interval of the cycle[111]. Namely, expression of cyclin genes for E and A variants, accompanied by cyclin dependent kinase (*CDK*) 1 and 2 points to cells entering or going through S phase. However, E variant cyclins are weakly expressed in a low number of the germ cell cluster (Fig. 3.1.3B) and are only expressed in some somatic cells at 5wpf (Fig. 3.1.3D). Cells expressing cyclin D variants instead, usually along with *CDK4* and *CDK6*, are entering G1 and have none or low *MKI67*[111]. In this case *CDK6* only shows lower expression levels in some germ cluster cells at 4wpf and mostly very evident to high levels of expression on somatic cells at 5wpf. As almost no germ cells are expressing this gene, they cannot be said to be in G1. The relatively high expression of *CDK1* in cells coinciding with S phase marker expression including *PCNA* + *MCM2* + *PLK4* at both ages suggests these cells are exiting that phase. One cell can be noticed at 5wpf (Fig. 3.1.3D), as it is negative for *MKI67*, *PCNA*, *MCM2* and for Caspase 3(CASP3). Of note, as well, is the close clustering of *PCNA* and CCNB1 at both ages, which becomes closer at 5wpf. Some of the germ cells express *ESPL1*(separase), a gene known to be involved in anaphase, show lower expression at 4wpf and relatively higher expression at 5wpf[112]. Cells entering M phase can also be located by the coincident expression of Cyclins A and B, plus expression of *CDK1*. Among the *CDK*s, *CDK4* is the only gene showing expression ubiquitously. In turn, Cyclin E variants are almost not expressed in most cells, independently of the cluster they are placed in. Cyclin A1 variant does not show any noticeable expression apart from one cell at 5wpf. When combined with the expression of specific cyclins, *CDK* genes and genes like *PCNA*, *ESPL1* indicate that the cell cycle progression in hPGCs is not synchronous. Among germ cells, there is also a group which expresses CCNB1 more intensely, a protein linked to M phase entry and complexing with *CDK1* to favour the former. When looking at *CDK1* expression patterns in somatic cells, it can be said these are in two different phases of the cycle. These phases may correspond to M phase entry or G2 when cells are *CDK1*+ or *CDK1*+ and *CDK2*+, respectively.

Taking the expression of stage specific combinations of proteins above together, it is possible to see that not all germ cells were proliferating at the time point they were analysed.

From the plotting of PGC characteristic genes, four have sufficient support according to heatmaps and previous studies. These are *POU5F1*, *PDPN*, *TFAP2C* and *SOX17*, with the first two being highly expressed, noticeably more so in PGCs. While also showing expression specific to germ cells. Most of the remaining genes are more heterogeneous and/or weakly expressed. Nonetheless, the above plots (Fig. 3.1.1A and B) contain enough information to adjust the starting combinations. This is due to genes that are expressed specifically in PGCs but not in primed PSCs (Fig. 3.1.1 and 3.1.2A).



3.1.4 – Most laminins and caderins are not expressed by germ cells

Figure 3.1.4. Heatmap plots of genes encoding the Cadherin, Laminin and Integrin adhesion protein families. Log2 normalized TPM (TPM/10+1) of (A, B) Cadherins at 4 and 5wpf, respectively, (C, D) Laminins at 4 and 5 wpf, respectively. Cell type labels were made according to metadata constructed from the dataset. Euclidean clustering was performed according to gene and to sample.

Adhesion proteins are known to be involved in PGC movement (Fig 3.1.4 and 3.1.5), as such, their expression levels were compared. As well as, ECM molecules that make the migratory movement possible by being interactors (Fig 3.1.4 and 3.1.5). Owing to the described PGC dependence on these molecules for movement until the cells reach the gonadal primordium[68].

At 5wpf (Fig 3.1.4B, females), laminin B1 shows distinctly higher somatic levels of expression, particularly in left side somatic cluster, albeit, only on four cells. Cells expressing this gene show higher and more homogeneous levels. At the terminal gene clustering branches, *LAMB1* clusters directly with *LAMA4* at 5wpf (Fig. 3.1.4C), as opposed to 4wpf (Fig. 3.1.4D).

Where it clusters with a greater number of genes expressed similarly throughout the samples. On the other hand, the remaining somatic cell cluster (Fig 3.1.4B), presents itself as more heterogeneous and occurring only on some cells. The cluster of soma cells on the last-mentioned cluster also show lower levels, contrasting with another somatic cluster. Genes like *LAMC1*, *LAMA4* and *COLA4* are too heterogeneous and ubiquitous at this age.

Most cadherin genes do not show elevated or low expression (Fig 3.1.4E, F), consistently on either cell type identifiable by the two main clusters of samples(cells). Absence of expression of *CDH1*(E-Cadherin) is verified throughout all samples, apart from one germ cell. Unlike *CDH2*(N-Cadherin), which shows undistinctive expression (Fig 3.1.4A, B). Also, *CDH11* is only expressed in roughly half of somatic cells at 5wpf, apart from an individual germ cell, making it specific for somatic cells.

From these results (Fig. 3.1.4) it is shown that adhesion molecules such as, laminins and cadherins are not good markers for PGCs.

3.1.5 – Integrins are mostly not specific to any cell type, while tubulins show more specificity

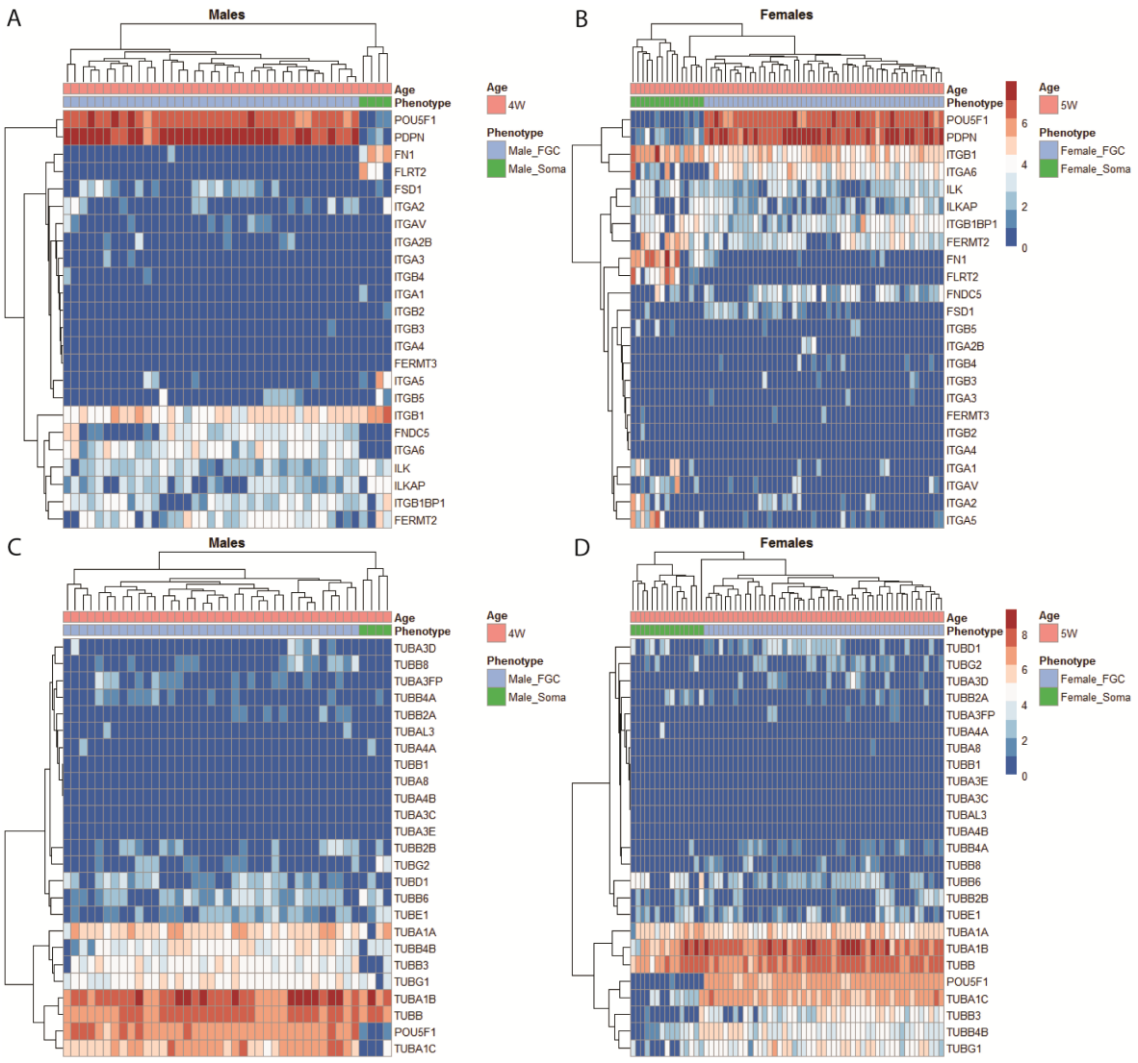


Figure 3.1.5. Heatmap plots of genes encoding integrin proteins expressed on germ or somatic cells and tubulin expression. Log2 normalized TPM (TPM/10+1) of (A, B) Integrins at 4 and 5wpf, respectively. (C, D) Tubulins

expressed in germ and somatic cells, at 4 and 5wpf post-fertilization, respectively. Cell type labels were made according to adjusted metadata from the dataset. Euclidean clustering was performed according to gene and to sample.

The integrin cell surface adhesion family was also analysed for potentially distinctive expression levels in germ cells. Germ cells do not express integrins, among these are the α3, α5, αv, β1 and β3 subunits (Fig. 3.1.5 A and 3.1.5B). All of which are proteins known to be expressed on mPGCs surface, as is *ITGA6*, with distinctly germ specific expression. This gene is the most distinctive between germ and somatic cells, since it is absent from the latter. At 5 weeks, more heterogeneity is observed in the expression of this gene and is not exclusive to cluster 1 or 2 of somatic cells. The β1 integrin subunit is one of the integrins is most critical in the early stages, here shown to be expressed ubiquitously[74,75,103]. The same occurs with *FERMT2*, a gene involved in assisting in integrin-fibronectin connection[113,114]. On the other hand, *FLRT2* shows expression entirely on the somatic cell cluster.

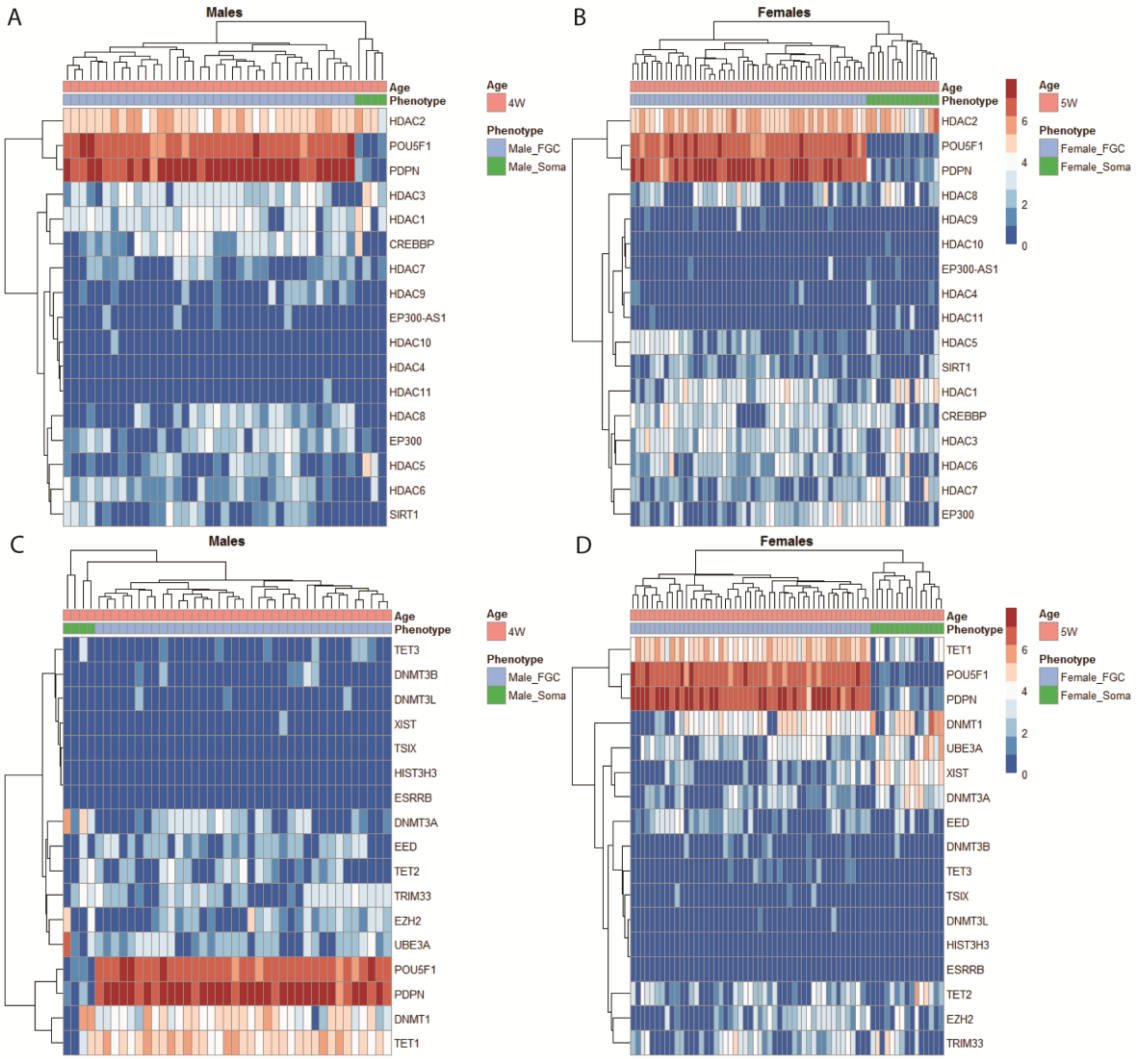
There were no genes, other than *FNDC5* at 4wpf and *ITGA6* (Fig 3.1.5A and 3.1.5B) showing any cell type distinctive expression. Only *ITGA6* is supported by heatmap plot comparison of expression regarding germ cells/somatic during the period plotted. This gene also clusters with other PGC specific described markers at 4wpf and with the most distinct, as well as, highly expressed genes at 5wpf.

Gomes Fernandes et al. (2018) previously reported that tubulin beta III(*TUBB3*) could be detected PGCs on paraffin sections of a 4,5wpf embryo. For this reason, the tubulin superfamily was plotted in the Li et al. (2017). Most tubulin superfamily genes have low and inconsistent expression across the cells, particularly in the upper and middle gene clusters (Fig. 3.1.5C and 3.1.5D). Unlike the genes referred above, the remaining two gene clusters show higher and more consistent expression. These lower clusters show variability in the specificity for cell phenotype between the ages shown. Particularly, *TUBA1C* is comparably more expressed in somatic cells at 5wpf, while *TUBG1* is less expressed. In this case *TUBB3* is an exception since it remains germ cell specific.

Most integrins are not determinant for the specified germ cell/PGC phenotype, due to lack of specificity in these ages. This is verified both ages, except in the case of *ITGA6*. The expression landscape is similar in the tubulin family. However, tubulins have more exceptions that can be used as markers if the antibodies are available and if expression is weak in PSCs.

The objective of these plots was to find genes that can complement combinations of pluripotency genes and different PGC specific genes. Among these genes, *LAMB1*, *FN1*, *PECAM1* and *CDH5* cluster together and show the highest relative expression in the somatic cluster. *LAMC1* seems to be the more specific gene in the germ cell cluster. Of these genes, *LAMC1* and *FN1* are the best supported by the heatmap plots (Fig.3.1.4C and 3.1.4D). Among the remaining genes, *TUBB3* and *ITGA6* are good candidates for markers according to their expression across samples (Fig. 3.1.5).

Extensive epigenetic reprogramming occurs while PGCs migrate, including in the 4-5wks period. Genes related to epigenetic control of gene expression were also plotted and are shown below in the remaining plots.



3.1.6 – Chromatin state and histone modifiers are in general not reliable candidates

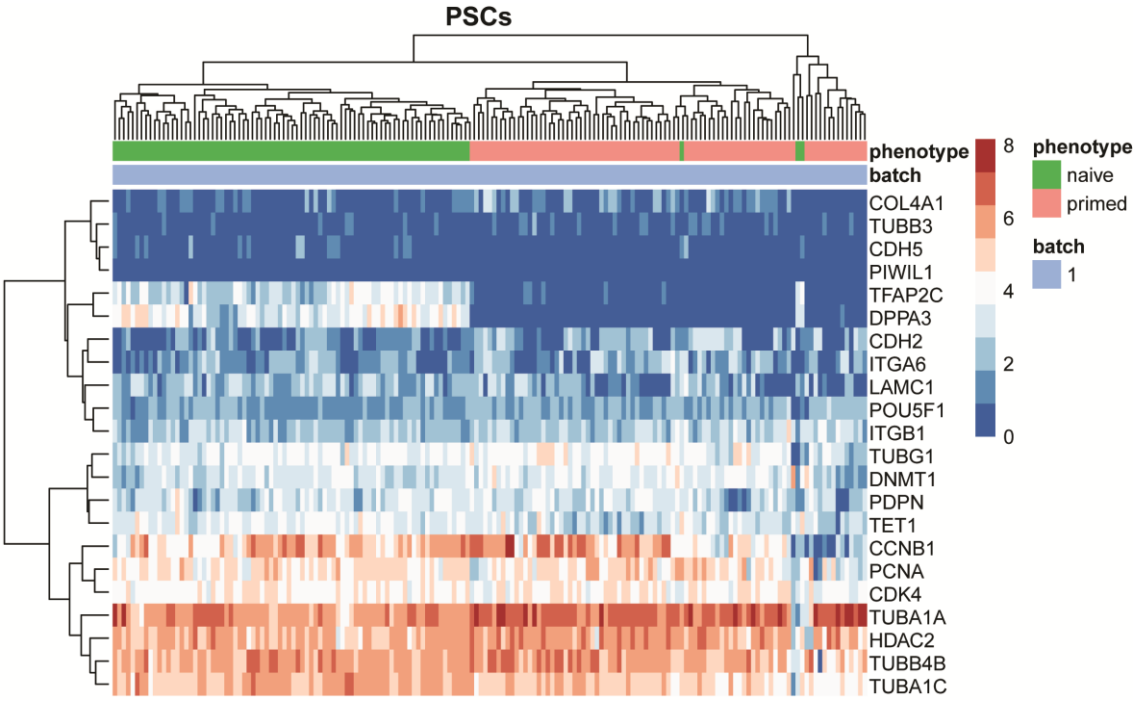
Figure 3.1.6. Heatmap plots of histone deacetylase and epigenetic regulator genes. Log2 normalized TPM (TPM/10+1) of (A, B) Genome modifiers and euchromatin contributors, at 4 and 5wpf, respectively. (C, D) Epigenetic regulators expression, at 4 and 5wpf, respectively. Cell type labels were made according to metadata constructed from the dataset. Euclidean clustering was performed according to gene and to sample.

Having been known to play a role in PGC differentiation and development, through reprogramming of gene expression and overall demethylation of the genome. Genes coding for factors responsible for histone post translational modifications (PTM) and other epigenetic regulators like TET enzymes were plotted (Tang et al., 2015, Tang et al., 2016). Showing a poorly distinctive expression according to cell type at 4 weeks by themselves (Fig. 3.1.6C and 3.1.6D). However, at 5 weeks (Fig. 3.1.6C and 3.1.6D), *TET1* clusters together with *POU5F1* and *PDPN* as a distinctive and similarly expressed gene, despite the heterogeneity in soma. *EZH2*, a gene involved in the methylation of histone H3 is expressed indiscriminately (Fig 3.1.6C and 3.1.6D; Tang et al., 2016). DNMTs show very variable expression and other enzymes, such as TETs (1 and 2) aren´t expressed exclusively in germ cells even if expressed mostly in these. *TET3*, on the other hand, shows no expression at either age. DNMTs (Fig.

3.1.6C and 3.1.6D), an enzyme type described to be downregulated during PGC migration, present lower expression in germ cells, except for *DNMT1* at 4wpf[44]. While, at 5wpf *DNMT3A* becomes more expressed in somatic cells (Fig. 3.1.6D), as does *DNMT1*, an enzyme tasked with preventing ICR demethylation[115,116]. In addition, an ICR linked gene was included (*UBE3A*), but yielded no differentiation in terms of expression between cell types and only shows slightly increased expression at 5wpf. To be noted as well, expression of *XIST* is higher in the somatic cluster and is coincident with an increased number of germ cells expressing this lncRNA at more evident levels (Fig. 3.1.6C and 3.1.6D).

Overall (Fig. 3.1.6C and 3.1.6D), *TET1* is the one gene showing greater expression in germ cells and considerably weaker or no expression among soma cells.

Taking the plots shown so far together, the only genes that present themselves consistently as having distinctive expression levels that are characteristic to a cell type. In this case, to germ cells, with the genes in question being *POU5F1* and *PDPN*. However, some other genes, in the plots above, have shown cell type specificity consistently through immunostainings and qPCR[19,42,51,103,117].



3.1.7 – Genes other than established markers can assist in detecting PGCs?

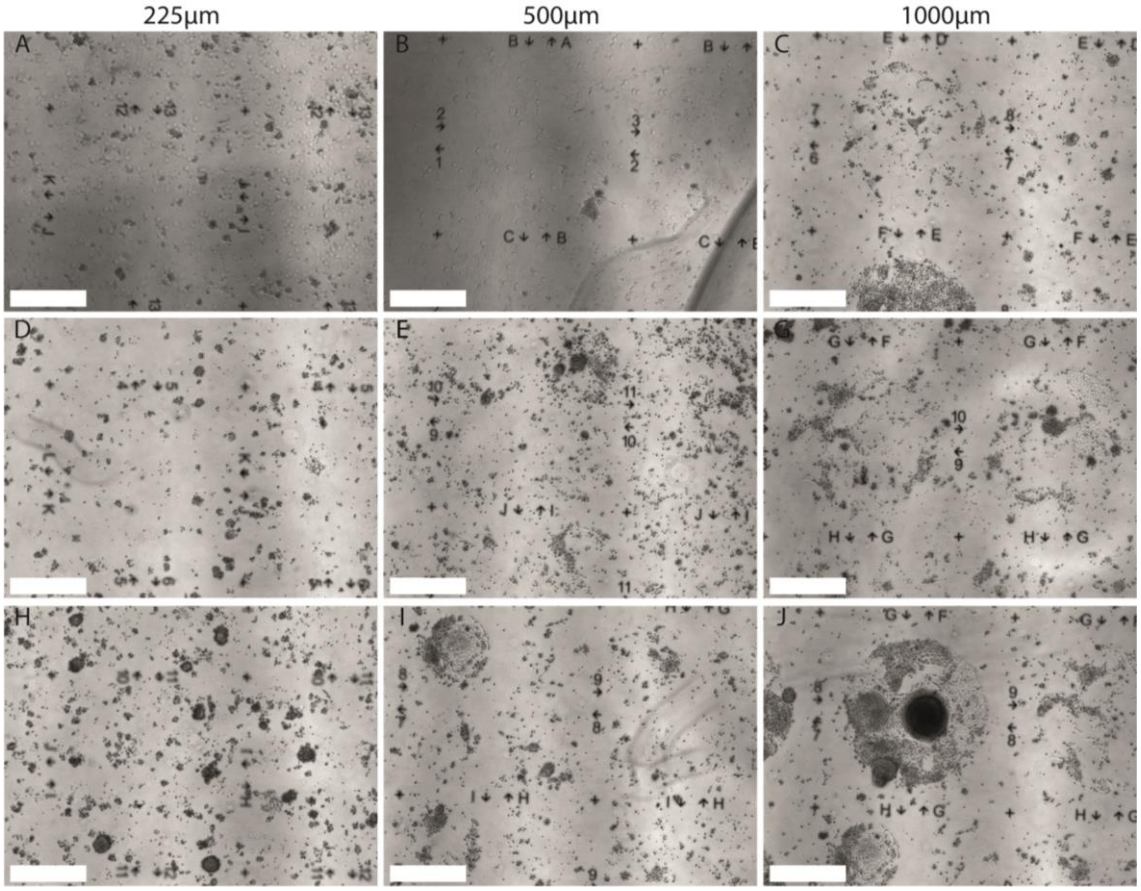
Figure 3.1.7. Heatmap plot of germ and somatic genes with primed and naïve cells. Log2 normalized TPM (TPM/10+1) of (A) Genes selected as markers, for which antibodies are accessible for this study. Cell type labels were made according to metadata constructed from the dataset. Euclidean clustering was performed according to gene and to sample.

For the purpose of testing if any genes previously outlined as germ cell specific, the referred genes were plotted using the PSC DF. Among those genes, *TUBB3* was indeed restricted to germ cells, along with *TUBA1C* (Fig. 3.1.5A and 3.1.5B). However, only *TUBB3* is shown here, weakly expressed in the Messmer dataset, mostly on primed cells. Apart from this gene, *DPPA3* and *TFAP2C* show relatively strong expression, that is seen only on naïve PSCs.

The remainder of the genes compared here do not seem to be specific to a PSC cell type or have a low enough expression throughout the samples (Fig. 3.1.1 to 3.1.7).

Most genes that could be PGC markers, according what is shown by the analysis so far cannot provide an adequate distinction between PSCs and PGCs. This is seen in genes that show expression ubiquitously across all PSCs, but not in genes that generally show almost no expression across all PSCs or only in naïve cells (Fig.3.1.7).

3.2 – Cell culture: Testing micropatterns and supported marker genes



3.2.1 – Reproduction of micropattern seeding with PSCs proved unsuccessful

Figure 3.2.1. Brightfield images of D3B4 condition Cytoos seeded with CL002 cells from different chips taken prior to fixation. All images were acquired at 4x magnification. (A-C) Images on top are representative of lower coverage areas found in this condition. (D-G) Middle row images are representative of median like coverages. (H-J) Bottom row images are representative of similar or higher coverages compared to the middle row. Scale bar = 500µm.

Coverage of D3B4 condition Cytoos seeded with CL002 cells was less in general when cells did not spend 1 week in 6-well plates coated with fibronectin. Considering this, coverage in the D3B4 or 1wk*FN1*+D3B4 condition with hiPSCs (Fig. 3.2.1). The top three radiuses are shown, instead of the 80-1000µm range, for representativity and lack of relevant differences.

WIS1 seeded Cytoos of the same condition show a very different" landscape", as those chips had an appearance much like a deserted area (similar to Fig. 3.2.1B) or less covered. Direct passage and differentiation of primed ESCs and hiPSCs did not reproduce the Warmflash et al. (2014) results.

3.2.2 – Reproduction of micropattern seeding with PSCs also unsuccessful with *FN1* adaptation

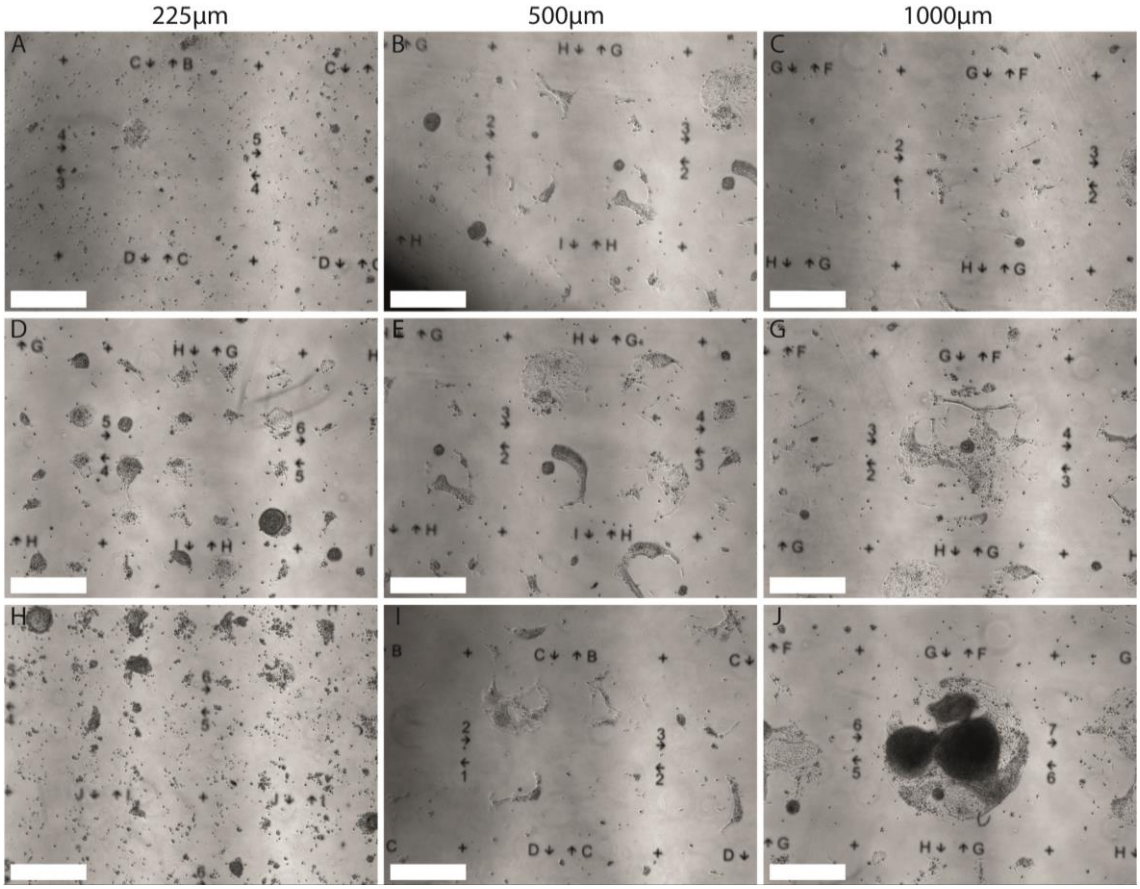


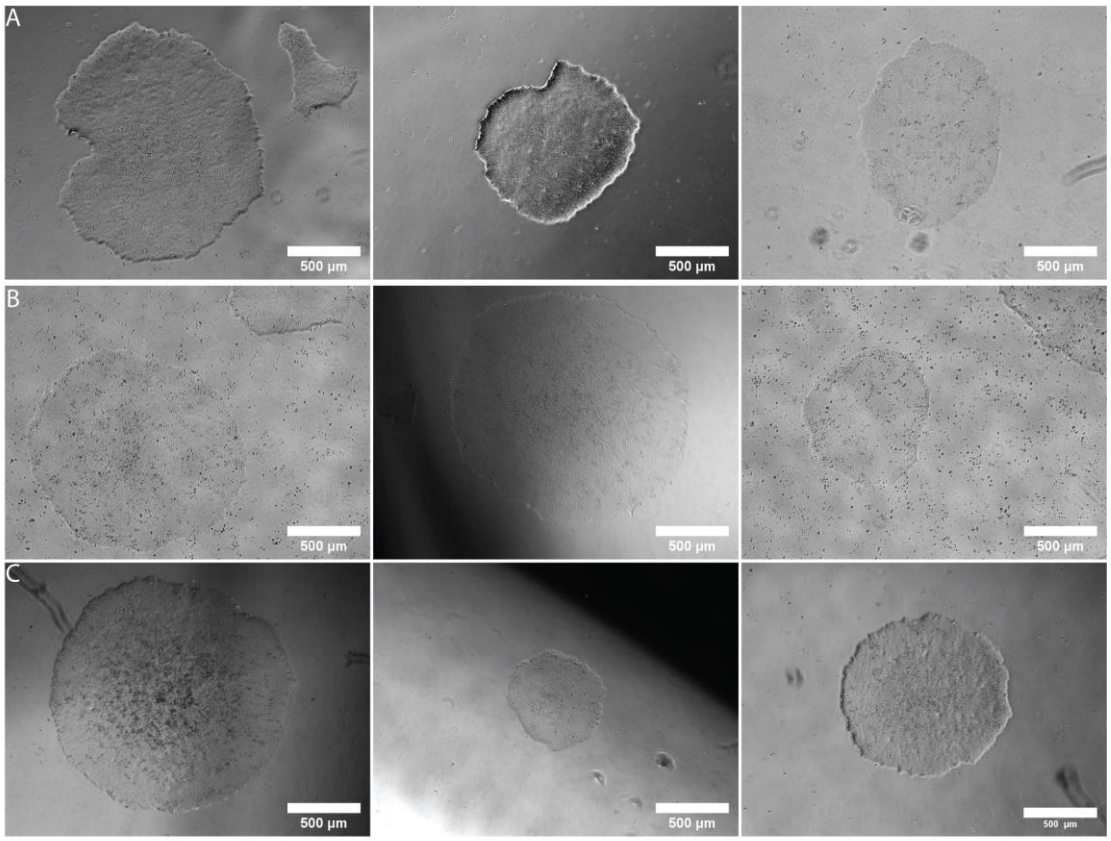
Figure 3.2.2. Brightfield images of 1wk *FN1*+D3B4 condition Cytoos seeded with CL002 cells from different chips taken prior to fixation. Images were taken at 4x magnification. (A-C) Images on top are representative of lower coverage areas found in this condition. (D-G) Middle row images are representative of median like coverages. (H-J) Bottom row images are representative of similar or higher coverages compared to the middle row. Scale bar = 500µm.

Recapitulating Warmflash et al. (2014) was not possible due to incomplete coverage of micropatterns of any size[98]. The seeding of CL002 hiPSCs which have spent 1 week attached on fibronectin (*FN1*) provided better attachment and less cell death. This observation was made due to widespread absence of darker small dot-like shapes on and around the micropatterns (Fig. 3.2.2).

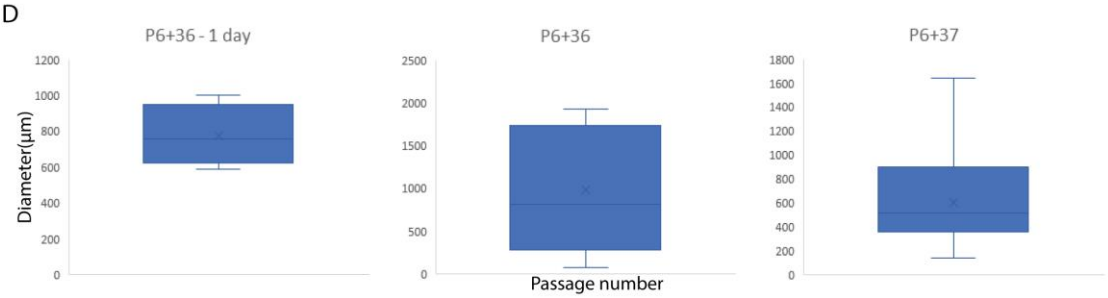
Overall, micropatterns were never fully covered. Only a few circles in each chip are almost completely covered and in some of these cases, clumps can be observed above the shape. Such clumps seem to be protruding from not completely attaching and cells expanding from it. Suboptimal attachment could be observed in all diameters and can be observed in the darker regions of the diameters shown (Fig. 3.1.2.1B, 3.1.2.1E and 3.1.2.1J).

This condition allowed cells a 1-week period of adaptation to their final seeding and differentiation substrate/coating (Fig. 3.2.2). This improved coverage but not enough to reproduce the minimum requirements for analysis, that were achieved in Warmflash et al. (2014)[98].

3.2.3 – Assessing colonies for antibody testing: a proxy for ideal size of differentiation



D



E

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
P6+36 - 1d	4	3108,27	777,0675	29500,84
P6+36	16	15691,62	980,7263	480377,5
P6+37	31	18794,12	606,2619	139385,6

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1486726	2	743363,1	3,109294	0,053732	3,190727
Within Groups	11475732	48	239077,7			
Total	12962458	50				

Figure 3.2.3. hiPSC colonies and radius analysis for protocol optimization. (A-C) Representative hiPSC colonies at day 6 of culture. (D) Boxplots of diameter ranges found inside each passage shown with diameters in micrometers (µm) and passage number of colonies measured in X axis and Y axis respectively. (E) Average radius and variance analysis between each group prior to passaging from the same dates, respectively. Statistics were performed to test for Gaussian distribution fit of diameter values. Scale bar = 500µm.

After unsuccessful experimentation with micropatterned chips, it was necessary to design a new experiment. Considering this, previously obtained data on hiPSC iCTRL12 LUMC 0030 colonies was used to assess colony growth and decide on a growth period before new differentiation attempts. Colony diameter was evaluated on passage day and one day prior, before passaging, to obtain an estimation on when the colonies could be differentiated after seeding. The diameter/growth of colonies over time was determined not to be different after no variation was found between groups using a two-way ANOVA. Thus, regular colony growth contributed to the devising of conditions for culture time until the differentiation of hESCs.

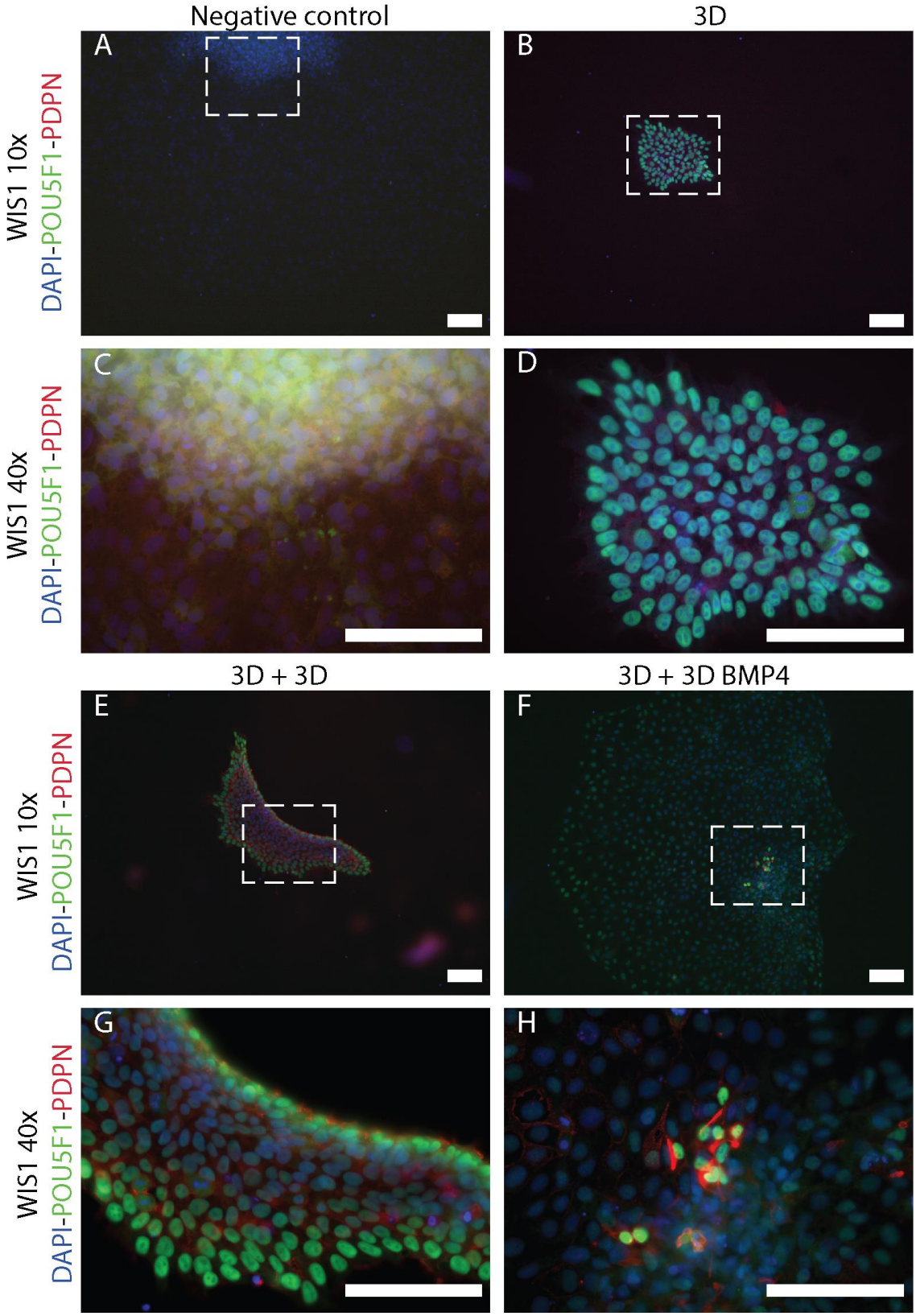
According to the values obtained through hiPSC culture, it can be seen that the growth of colonies is un-altered between passages. In addition, as hESCs were already adapted to their destination ECM (Matrigel), this adds to predictability of colony growth in the various conditions tested. With this design, it was intended that the colonies would differentiate without being restricted by the micropatterns (Fig. 3.2.3). For the purpose of testing antibody combinations, an experiment with colonies of similar sizes was performed using a differentiation method based on the chips. This originated from the decision of using coverslips to seed hESCs on coverslips and to differentiate them for 72h with BMP4 at 50ng/ml.

These conditions were 3D, in which the colonies were only allowed to grow until the start of the differentiation period, non-inclusive. 3D+3D condition colonies were allowed to grow until the end of the differentiation period without the BMP4 stimulus. Lastly, 3D+3DB4 is based on the 3D+3D condition, with the difference that it contains differentiation stimulation with BMP4.

3.3 – Staining hESCs in search of PGCs: Can they be generated with BMP4 supplementation?

Considering this and the observations made above, a set of combinations was attempted. First by allowing hESCs to grow for three days, until colonies reached target range of diameters(500-1000µm) and differentiating them for 72h. This range includes the values of colony diameter in which, as demonstrated by Warmflash et al. (2014) the germ layer patterns become reproducible. In addition, this was the best approximation possible achieved during the course of this project.

By following this protocol, the effect of differentiating the colonies was expected to be discernible, regarding an unrestricted round shape of final intended sizes. Putting together data from Fig. 3.1.1 to Fig. 3.1.7, different combinations were tested: combination 1, mouse anti-*PDPN* + Goat anti-Oct4; combination 2, Goat anti-*SOX17* + Rabbit anti- *TFAP2C*; combination 3 Mouse anti- + Goat anti- Oct4 + Rabbit anti-*TFAP2C*.



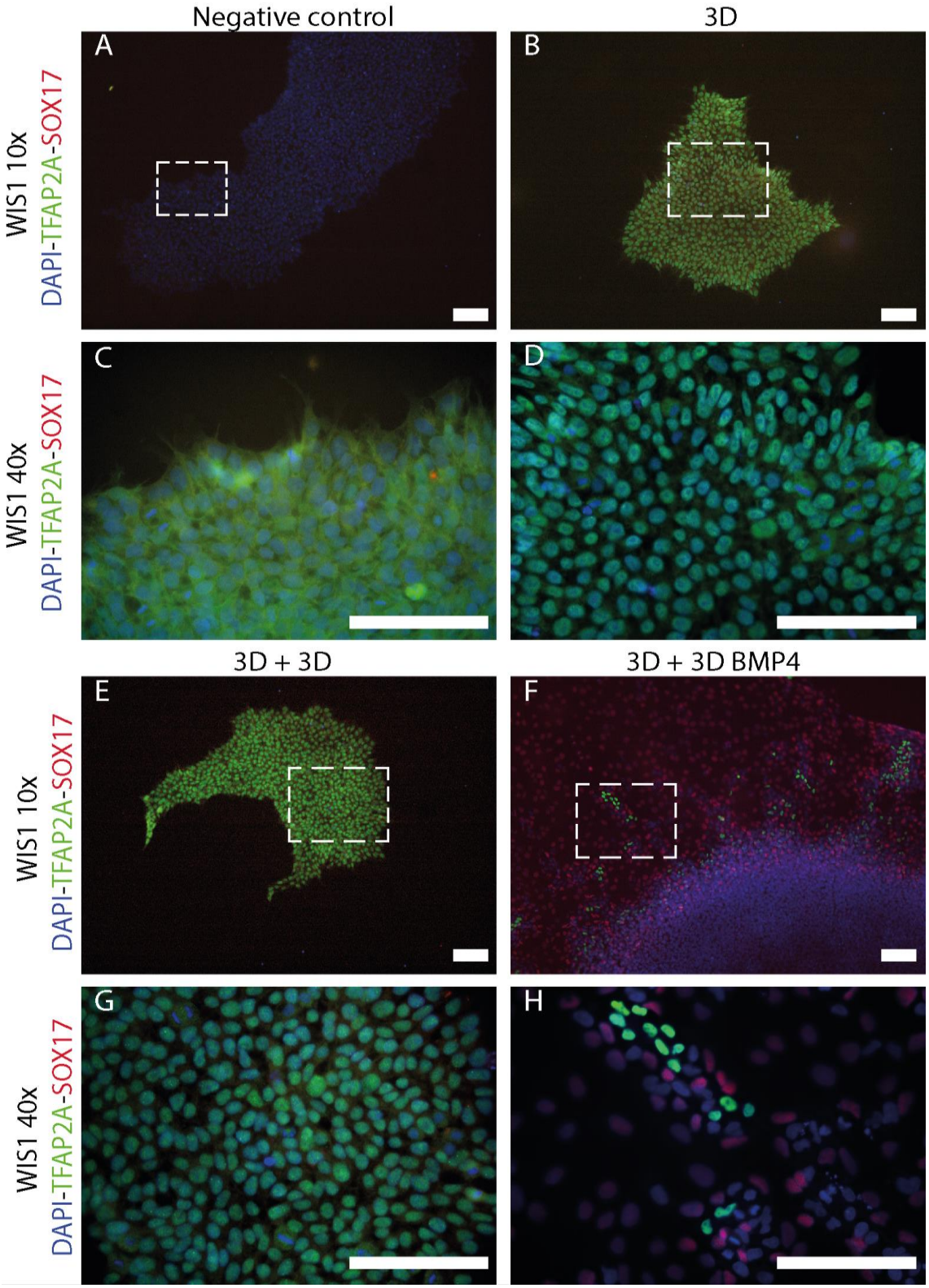
3.3.1– Staining hESCs in search off PGCs: Potential PGCs or regular PSCs found?

Figure 3.3.1. WIS1 colonies immunostained for *PDPN, POU5F1* and counterstained with DAPI. WIS1 cells stained for combination 1 at 10x (A, B, E, F) and 40x (C, D, G, H) magnifications. Conditions are shown at the top of each 10x-40x set. Selection areas indicate where the in 10x the 40x was taken. Scale bar = 50 µm.

The WIS1 hESCs were imaged at a widefield fluorescence microscope at several magnifications. For clarity purposes only 10x and 40x of representative colonies from each condition are shown.

3D condition cells were used to control for the remaining conditions, referred to as, 3D+3D and 3D+BMP4. These cells show consistently homogeneity in the signal from *POU5F1* staining antibodies (Abs), as is the case for *PDPN*. In 3D condition cells, the former is much more defined and stronger, while the latter is more diffuse and appears to be weaker. The previous observations differ from what can be seen in 3D+3D colonies, which were simply allowed to grow for 3 extra days. The 3D+3D condition colonies have patches where the intensity of *POU5F1* in cell nuclei appears to be much lower and where *PDPN* is much more visible. In addition, the 3D+3D condition colonies were mostly irregularly shaped, just as in the 3D condition (Fig. 3.1.1 B, C, H and I). Although rarer, roundish colonies were also found in either. Both conditions mentioned so far have had around two colonies that were fully differentiated and flat, these were barely visible in all channels at the microscope´s ocular in every coverslip. On the other hand, coverslips to which BMP4 was added, in condition 3D+BMP4, patches and dome-like cell clusters with a noticeably higher intensity of *POU5F1* and *PDPN*. Rare fully differentiated colonies were observed as well, similar to the extent of the 3D+3D condition. These were found in the majority areas from overgrown colonies, whose borders were lost upon fusion. Such patches, long extensions or domes are described as groupings of cells found above a monolayer characterized by much lower intensities of either protein.

This staining can distinguish differentiated and double positive (DP) cells. The antibodies used cannot tell which cell types the differentiated cells belong to. This means that, confirmation that the DP cells are potential PGCs or primed hESCs is not possible with these genes alone, due to shared expression between PSCs and PGCs (Fig. 3.1.1 and 3.1.7). Moreover, this meant that a different combination would be necessary to confirm the identity of these cells (see 3.3.1 section of Discussion for further information).



3.3.2 – Staining hESCs in search off PGCs: Immunoluorescence misses potential PGCs

Figure 3.3.2. WIS1 colonies immunostained for *SOX17*, *TFAP2A* and counterstained with DAPI. WIS1 cells stained for combination 3 at 10x(a-d) and 40x(e-i) magnifications. Conditions are shown at the top of each 10x-40x set. Selection areas indicate where the in 10x the 40x was taken. Scale bar = 50 µm.

Unfortunately, the vial for *TFAP2A* was mistaken for *TFAP2C* and as such, the signal for the 488(green) 2º Ab will not be informative about the possible presence of PGCs.

In 3D+3D and 3D+3DB4 conditions the cells show some degree of homogeneity in the signal from *TFAP2A* Ab staining. In 3D condition cells, *TFAP2A* was much more defined and stronger, as is the *SOX17*. The previous observations were similar to what can be seen in 3D+3D colonies in this combination as well. The 3D+3D condition colonies were more inconsistent in intensity of *TFAP2A* in cell nuclei. In addition, the 3D+3D condition colonies were also mostly irregularly shaped, just as in the 3D condition and in combination 1 (Fig. 3.3.2B, C, H and I; Fig. 3.3.1). On the other hand, coverslips to which BMP4 was added, in condition 3D+BMP4, elevated cell clusters with a noticeably higher intensity of *TFAP2A* and where *SOX17* was also seen. These were found in overgrown colonies, which contained multilayer areas with finger like shapes on their perimeter.

This staining did not confirm the identity of the DP cells found previously (Fig. 3.3.1). However, it allowed the identification of structures resembling the *SOX17* stripes radiating from the inside of colonies in Warmflash et al. (2014) [94; see Figure 2].

3.3.3 – Staining hESCs in search off PGCs: Immunofluorescence highlights potential PGCs

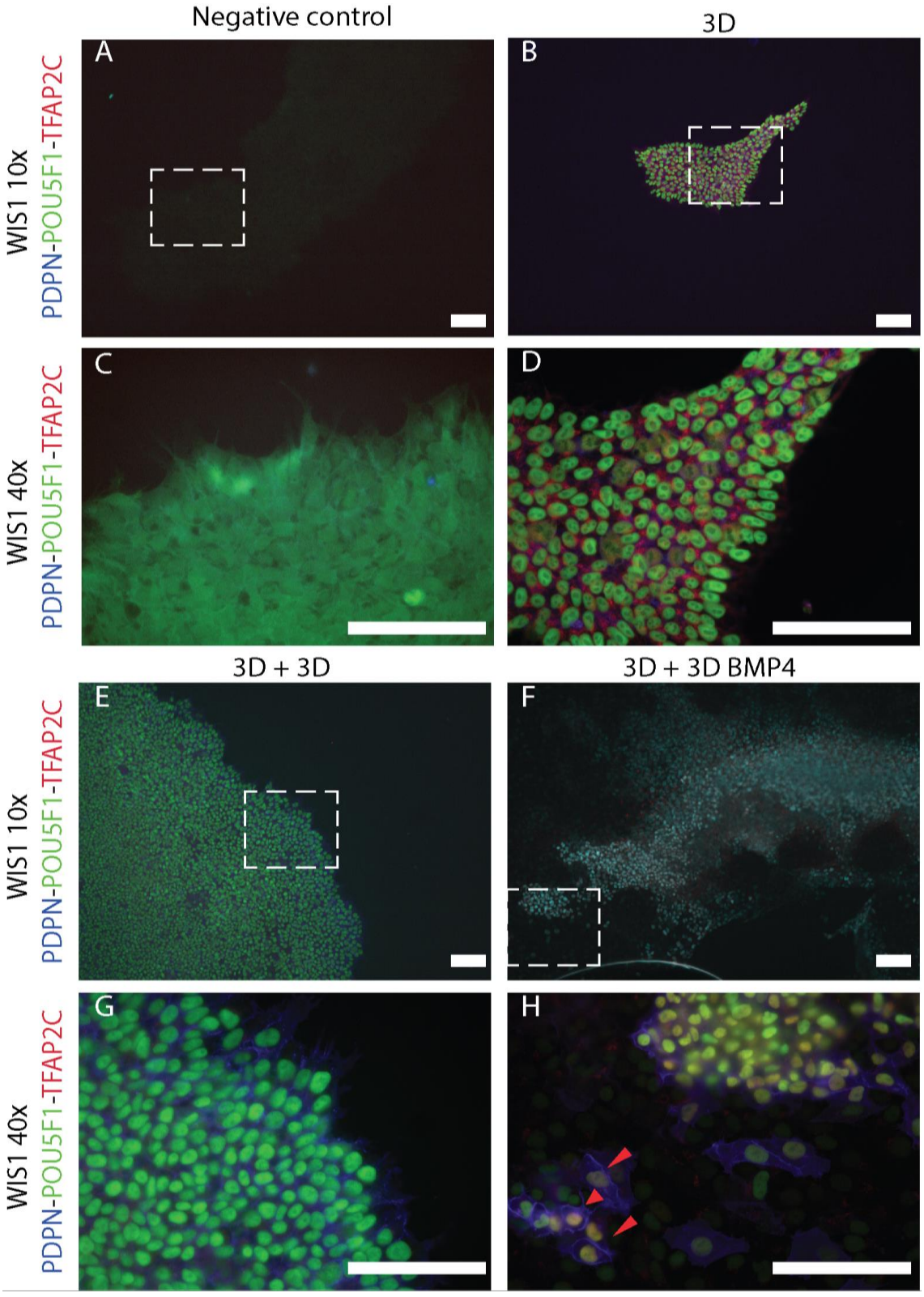


Figure 3.3.3. WIS1 colonies immunostained for *PDPN, POU5F1, TFAP2C* and counterstained with DAPI. WIS1 cells stained for combination 3 at 10x(a-d) and 40x(e-i) magnifications. Conditions are shown at the top of each 10x-40x set. Selection areas indicate where the in 10x the 40x was taken. Arrows indicate triple stained cells, which are potential PGCs. Scale bar = 50 µm.

An additional staining was performed to determine if the cells stained with higher intensity could indeed be PGCs. This was also motivated by reports that *PDPN* and *POU5F1* are also expressed by PSCs[1]. For these reasons a combination of three proteins were targeted, with the combination being: *PDPN+POU5F1+TFAP2C* (Fig. 3.3.3). Colony morphologies mentioned previously were also observed on the coverslips stained with this combination in all conditions. This means that, irregular shapes were predominant in the 3D and 3D+3D conditions (Fig. 3.3.3B, D, E and G). The 3D+BMP4 condition also contained overgrown cell areas, with dome-shaped clusters and separate cell groupings usually surrounded by weaker stained cells. Such features were observed in all staining, to similar extents.

This staining allowed the detection of potential PGCs (Fig. 3.3.3H, red arrows), these cells showed colocalization of *POU5F1* and *TFAP2C*. Additionally, a PGC morphology was visible in these cells. This was particularly evident in one of them (Fig. 3.3.3H, shorter red arrow).

This result opens the way for a new method that can be used for investigating PGCs *in vitro* (Fig. 3.3.3). Through this method, more can be learned about how these cells are specified, aspects of their biology and potentially contribute to assisted reproduction.

4. Discussion

4.1 – Bioinformatics analysis as a tool for the detection of germ cells *in vitro*

This study initially set out to detect potential PGCs by using the protocol of Warmflash et al. (2014) as a basis[98]. To do so, a way that can consider information from multiple model organisms and combine it with previous analysis of human cells is necessary. Multiple types of computational analysis that could, in some way, provide usable information to gain insight on cell identity and refine other tools to more efficiently detect them exist. One of the simpler and most used currently is gene expression analysis by plotting absolute read/transcript counts or normalized count values as heatmaps. This method allows users and non-bioinformaticians to locate distinctive patterns of expression and extract information from column/row clustering. In turn, such information can help in understanding what characterizes test samples, what that means in the context they are looked at and to notice experimental design flaws. The latter skill requires having more extensive experience with computational biology, while the previous ones do not require as much experience. Of note is that generating gene expression from minute amounts of starting material(cells) diminishes the capacity for detecting low-abundance transcripts after sequencing and impairs accurate estimation of expression levels[2,118,119].

Despite the unsuccessful outcome using the micropatterned chips, the results obtained by resorting to culturing of hESCs on Matrigel coated coverslips appear quite promising. The referred results regarding the hESCs will be discussed below.

4.2 – scRNA-seq analysis for selection of markers

The preliminary analysis of the Li et al. dataset (2017), reinforces the general view of the expression of specific genes as markers[104]. The genes fitting the description of markers show high relative expression in the germ cells plotted in comparison to somatic cells of the same age. This means that they constitute reliable tools to detect PGCs with a high degree of confidence *in vivo* as *in vitro*.

One feature observed throughout almost all plots is that a separate clustering branch is assembled, which specifically connects somatic cells, arising from similar gene expression. Furthermore, laminin genes show a peculiar branching of somatic cells into two clusters (Fig.2B). The same is effectively observed for the germ cells. This means, certain combinations of specific genes consistently highly expressed or lowly expressed produce a high Euclidean correlation between cells. This correlation is the criterion that gathers cells inside a cluster or sets them apart and characterizes them as separate cell types. Such cell types are visually identifiable as separate clusters consisting of cells exhibiting similarly expressed genes.

4.2.1 – Most established PGC markers are not a consistent as previous studies claim

Genes classically/widely accepted as markers in the scientific community were compared regarding relative expression level (Fig.3.1.1A and 3.1.1B). A number of these were confirmed to be expressed exclusively in germ cells, in particular, at 4wpf (data from Li et al (2017)). Observations discussed in this section are from plots made using Li et al. (2017) data, which compares only somatic and germ cells. *ALPL*, *NANOG*, *SALL4*, *SOX17*, *TFAP2C*, *PRDM1* are such genes. However, *SALL4*, *SOX17*, *TFAP2C*, *PRDM1* show surprising and noticeable heterogeneity in their expression levels, which increases at 5wpf. This is unexpected given the consistence/reliability with which they are used to detect and sort cells for/after

differentiation experiments[17,19,103]. *ALPL* cannot be used to distinguish PGCs and hESC/hiPSCs cells, as it is common to primed PSCs and PGCs[85,120,121]. The pluripotency gene *NANOG*, is not reliable for distinguishing PGCs from undifferentiated stem cells. This is due to previous knowledge that *NANOG* also shares expression in PSCs and is verified in the heatmaps (Fig. 3.1.1 and 3.1.7)[85,122].

PRDM1 shows an expression in germ cells unlike from what was described previously by others in mouse, from early stages, through migration, up to later stages. Especially since ablation of the gene ultimately leads to PGC death[123–125]. Considering this, the expression observed across samples is quite striking, due to the impairment of specification and migration events when *PRDM1* is not expressed[126]. *SOX17*, another gene critical for specification is shown with an expression level among germ cells far from what was described previously by others in cells from the same organism[34,42]. *NANOS3* was used as a reference to identify PGCs in a prominent publication describing specification *in vitro*, as well as, multiple other studies confirm this[1,27,84,103] . Although, expression of *NANOS3* is somewhat variable in the PGC marker plots (Fig. 3.1.1). *CD38*, a protein determined to be a marker presented only by hPGCS and the *in vitro* generated counterparts, aside from TCam-2 cancer cells, given the absence from hESCs or somatic cells. Ultimately, this characteristic expression allowed the use of this molecule as a marker to sort between hPGCs and other cells[103]. Indeed, *CD38* is uniquely expressed by germ cells at 4wpf. Although this characteristic was less obvious at 5wpf, due to an increased number of cells without *CD38* positive relative expression (Fig. 3.1.1B). In the case of *IFITM3*, the expression observed throughout both cell types is expected, as it was considered to signal the acquisition of germ cell fate[108]. In addition, this gene has been shown to be clearly more expressed in mPGCs instead of soma, despite having greater variability around specification E6.25-6.75. Contrary to what is observed in the plots, consisting of human cells, mouse cells show a much more consistent expression of *IFITM3* in germ rather than in somatic cells[108].

Selection of genes for *in vitro* detection of PGCs depended on how specific genes were supported by the heatmaps or not. This means that, their selection needed to be based on whether they were exclusive to the germ cell cluster. Furthermore, having noticeably higher expression across germ cells than in somatic cells was also a criterion. Some genes mentioned above show inconsistency, such as *PRDM1* and *SOX17*(Fig. 3.1.1A and 3.1.1B). *POU5F1* and *PDPN* are more homogeneous compared to *PRDM1* and *SOX17* and are supported by the plot alone quite strongly. *PDPN* should, however be combined with another gene such as *TFAP2C*, due to being expected in the primed PSCs used, contrary to *TFAP2C* (Fig. 3.1.7). On the other hand, genes like *SOX2*, GATA6 showed almost no expression whatsoever in PGCs, as expected. This matches the described profile of these genes. Gata6 is a somatic endodermal gene, and so, is not expected to be expressed in germ cells, although low expression is visible in two cells. More specifically, *SOX2* is expressed in mPGCs contrary to hPGCs and linked to pluripotency PSCs. The complete *SOX2* absence confirms this expectation, according to previous studies[38]. *EOMES* induces *SOX17*, which in turn, induces *PRDM1*. Additionally, *EOMES* needs to last for a short period, for hGC competence to be acquired by cells with a relatively greater degree[1,27,42]. As *EOMES* is not detected in the 4-5wpf analysed, this confirms the requirement that this gene needs to be expressed for a short time for PGCs to be specified adequately. This means that, the absence of the gene from plots(Fig. 3.1.1A and 3.1.1B) is according to expectation, but provides no information of use, by itself, on whether *EOMES* can

be used as a marker[40 Fig. 5D,99]. *CDX2* can viewed in a similar way as *SOX2*, since expression is specific to TE/ExE and not detected during this period.

4.2.2 – PGCs multiply while migrating: hPGC proliferation is asynchronous

Most reported genes to be involved in migration did not provide any support to genes described in the literature as markers. Among them, *CXCR4*, a receptor to the cytokine *CXCL12*, showed specificity towards germ cells at 4wpf and maintained this at 5wpf. However, the cytokine *CXCL12* was less specific at 4wpf and lost further specificity in at 5wpf, while remaining mostly expressed in germ. *CXCL12* and *CXCR4*, is one of two regulatory cytokine-receptor interactors pairs/signalling systems described to coordinate mPGC migration. The second one being the cytokine *KITLG* and receptor KIT[126–128]. The latter showed zero levels of relative expression at 4wpf and not enough expression on either cell type to be considered specific of germ or somatic cells. This was highly unexpected as the analysed period corresponds to the migratory phase of hPGCs[44]. Since proliferation has been described to occur during this time span(migration), *MKI67* was included in the expression comparison as well. It was found to be expressed in roughly half of the germ cells at 4wpf and in a lesser proportion of germ cells at 5wpf (see Tang, 2016 for a general Review). To the extent of current knowledge any information on the number of cells, what accompanies this indication that ~50% of the germ cells are proliferative and what is connected to the reduction of cells expressing *MKI67* going into 5wpf. This is a newly observed expression "landscape" in germ cells and motivated further analysis.

As such, after plotting cell cycle related genes alongside some of the genes from the migration plot (Fig.3.1.3C and 3.1.3D). It was observed that not all of these cells were in the same stage of the cell cycle, different combinations of cyclins and *CDK*s, together with other cell cycle pointed to different phases. In particular, one cell at 5wpf (Fig.3.1.3D) does not show visible relative expression of *PCNA*, *MCM2* or *MKI67*. Additionally, four cells were found to express particularly high levels of CCNB1, indicating that they were entering or already in M phase. The remaining *MKI67*+ and *CCNB1*- were then, in a different phase of the cell cycle, but still "cycling"[129]. However, CASP3 was detected.

4.2.3 – Most adhesion molecules are not expressed by germ cells

The ECM components expressed by somatic cells are known to correlate with the path PGCs take and are said to impact their survival during migration, including how they migrate[68,130]. Among the ECM molecules found on the route there were laminins, fibronectin, collagen type IV. Furthermore, fibronectin has been shown to be enriched on somatic mesenchymal cells in Pereda et al. (2006). In the cells compared in the dataset, fibronectin (*FN1*) is also far more expressed in somatic cells compared to germ cells. Genes encoding adhesion molecules carried fewer potential markers, as most of the genes plotted showed no specificity towards cell type. These gene families were not informative in most of their members. This was unexpected based on reports that point to a role in guiding PGCs in their route to the gonadal primordium[34,71,131]. One gene that stood out is *ITGA6*, which is shown as germ cell specific at 4wpf. *ITGB3* was expected to be expressed in this period, but was not found to be expressed[71,103,126]. The remaining integrins presented no cell type specific expression, rather, their expression was ubiquitous. Such an expression profile could

due to Li et al. (2017) not producing or making a greater degree of depth to their analysis available, for example mRNA splice form counting. Instead, all mRNAs that could be mapped to a gene, based on homology, were assigned to the same gene identity. At the same time, cell positioning cannot be factored in when scRNA-seq is performed in cells that were dissociated from whole embryos and sorted to maximize finding the GC fate. In addition, studying cellular location in the embryo and matching it to a cells expression profile to infer response to signalling would also be impossible[104]. Any post-translation modifications that could be associated to proteins in the route that PGCs travel through with the Li et al. (2017) dataset, even when adding the information of studies that identified the proteins. This relates to the choice of antibodies in those studies being directed to detect the adhesion molecules, including their subunits indiscriminately and thus, ruling out splice variants[66,68]. This makes the expression profiles not unexpected in face of insufficient information.

Although *LAMC1* could be considered, according to the plot with Li et al. (2017) cells and due to very high expression, perhaps even with the Messmer et al. (2019) cells plot (Fig 3.1.4C and 3.1.4D, Fig. 3.1.7). Unfortunately, no antibody was available to use, for both *LAMC1* and *TET1*. According to these observations, only *ITGA6* would constitute a good marker.

4.2.4 – Most chromatin and histone modifiers do not provide good PGC markers

Concerning DNA methylation, expression levels were not informative due to lack of restriction of a great number these genes to ether germ or soma cell types. A consistent difference between clusters in relative expression was not observed either. Apart from *DNMT1* and *HDAC4*, no other genes showed an expression profile according to what is described for the 4-5wpf period[49]. *HDAC4* in particular, displayed expression levels(Fig. 3.1.6A and 3.1.6B), matching the expectation that this molecule is repressed by *POU5F1* according to a previous report[132]. Likewise, *POU5F1* displays high expression levels, while *HDAC4* is barely expressed at 5wpf, Although, there were three cells expressing *HDAC4* at medium to low expression levels. Expression of DNMTs at 4wpf is consistent with their suppression in PGCs, present before they start migrating and with global DNA demethylation[44].

TETs are linked to the conversion of methylation marks on the genome into forms that are more easily removed and later, these marks are not detected on the genome (see Fig.3B 9wpf of Tang et al., 2015)[44]. *TET1* seems to be the only enzyme active specifically in germ cells and the expression at 5wpf suggests that methylation removal is still ongoing. According to Seisenberger et al. (2012) mPGCs start losing their genomic methylation upon departure from the specification site, the base of the allantois, going into the hindgut[133]. These studies on DNA demethylation in both mouse and humans concur in this point, further supporting *TET1* as a candidate for a marker gene. When these studies are considered with gene expression in plots, they allow the assumption that 5-mc will be minimal, as opposed to 5-hmc[44,133]. Moreover, this means that TET proteins could already have a small increase in expression in comparison to surrounding somatic cells. These proteins are probably difficult to detect by IF, as pixel intensities would be less noticeable closer to specification. This assumption is corroborated by an already low intensity of 5-mc observed in hPGCs (see Fig.3B, from Tang et al., 2015) by Tang et al. (2015) at 4wpf. This fluorescence can, moreover, become detectable reliably if combined with *POU5F1*, according to Tang et al. (2015)[44]. Unfortunately, no *TET1* antibody

was available to use for this study, which was one reason for it not to have been used. Even though it was considered at first, for the potential it had and because it was regarded as ubiquitous in the plot of candidates other than established markers (fig. 3.1.7).

The ubiquitous expression profile noticed for most of these genes is likely due to their pervasive expression in a developing embryo, in which cells are differentiating everywhere. This is linked to chromatin changes in somatic cells furthering their differentiation, moving and dying, not only in germ cells[134,135].

Female PGCs would have a H3K27me3 methylation pattern outlining one active X chromosome (Xa) and one inactive X chromosome (Xi) upon specification. This also signifies that PSCs of the same sex would share the same Xa and Xi pattern[85]. As the cell line that in the end was used to attempt differentiation of hPGCs by induction is female, greater importance was assigned to observations made from 5wpf plots. Which led to the most consistently and highest expressed genes at that age being contemplated as candidates. The 5wpf age was also used as a tie breaker if a gene showed equal specificity at 4wpf but different specificity at 5wpf The definitive selection was made after the inclusion of genes specific for the germ cell fate expressed in the plots and that were consistently found to be determined as PGC markers throughout the literature by being effectively proven to detect PGCs.

All heatmap figures encompass the developmental period from 4 to 5wpf, which matches the migratory phase in humans and E8-11 in mouse, information shown might not closely mimic specification events[1,51]. Although, the most homogeneously expressed genes, such as, *POU5F1*, *ITGA6* and *PDPN* already have support as markers to identify PGCs by methods like immunocytochemistry[1,44].

4.3 – Testing micropatterns and supported marker genes

4.3.1 –Unsuccessful reproduction of the micropatterning methods

In the previous section it was observed that micropatterns were never fully covered. Only a few circles in each CL002 seeded chip are almost completely covered. In some areas, clumps can be observed above the shape. Such clumps seem to protrude from incomplete attachment clumps on passaging. Suboptimal attachment was observed in all diameters of the chips (Fig. 3.2.1B, 3.2.1E and 3.2.1J; Fig. 3.2.2). The absence of attachment to fibronectin by WIS1 cells can be explained as due to a possible incompatibility between cell line and coating. Such an incompatibility can arise if cells do not have the response of producing the proteins or other factors necessary to attach.

4.3.2 – hiPSC wells do not present different colony diameter variability between passages.

Colonies displayed a size heterogeneity common in cell culture. Taking this into consideration, most colonies still measured over 500µm after four days. Their diameters before passaging did not vary significantly (p-value= 0,053). For this reason, a growth period of three days prior to differentiation by BMP4 addition was chosen to allow colonies to reach the ideal expected size during the differentiation period. Taking advantage of this, an experiment with the

goal of differentiating hiPSCs with BMP4 but without restricting colony borders was attempted. In this experiment, cells were allowed to differentiate for the same interval as the chips.

Considering the aforementioned reasons, the use of hiPSCs culture to study PSC growth and choosing to use such adjustments makes this type of culture a good proxy for trying a different protocol. This choice also seemed to have not had too much of a negative impact since it was mostly in conditions where BMP4 was supplemented that overgrown colonies were found. The outcome of latter mentioned protocol will be further discussed and in greater detail in the following subsection.

4.3.3 – Staining of hESCs on matrigel coated coverslips

Conditions referred to as, 3D(control) and 3D+3D controlled for the effects of BMP4 supplementation in the remaining condition. The latter condition appears to have induced the differentiation of PGCs (Fig. 3.1.3J). The consistent homogeneity in the signal from *POU5F1* staining antibodies (Abs), as for *PDPN* seen in the WIS1 as mentioned in the results section seen in BMP4- conditions contrasts with this. Such observations were made in 3D condition cells, where the signal for the former is much more defined and stronger, while the latter is more diffuse and weaker. These cells appear to be regular stem cells. In 3D+3D colonies, the differences in *POU5F1* are attributable to prolonged culture. The 3D+3D condition colony patches outline where differentiation had started to occur and allow for greater confidence on observations regarding the differentiation found in BMP4+ condition. Colony shapes made any possible measurement on colony diameter statistically unreliable, as round colonies were insufficient in number and due to loss of one coverslip while imaging (Fig. 3.3.2, 3.3.2, 3.3.3). Both BMP4- conditions mentioned so far, on either staining were mostly of irregular colonies. This removes any reliability from these results, concerning the diameter measurements that could have been made, as statistical robustness would not be possible. However, this further adds to the reliability of results when comparing BMP4- and BMP4+ conditions, specifically to differences arising from BMP4 supplementation. Contrary to what was found in BMP- conditions, coverslips to which BMP4 was added(3D+BMP4), isolated groupings, patches and dome-like cell clusters with noticeably higher intensity of *POU5F1* and *PDPN* were observed. Rare fully differentiated colonies and the 3D+3D patches provide a good reference as to what cells not differentiated into the PGC phenotype may look like. Found only in BMP4+ condition, such patches, long extensions or domes where cells with stronger intensities of *POU5F1* and *PDPN* (double positive or DP) are tempting for claiming PGCs have been generated. However, these two genes are also expressed by PSCs. Of note is also the fact that a more complete combination could test if DP cells are potential PGCs by adding a third marker, *TFAP2C*. This marker was also supported by the heatmaps.

A second staining was performed (Fig. 3.3.2), however, due to a mistake the wrong vial was used and a *SOX17+TFAP2C* could not be tested. Although, the *SOX17+TFAP2A* allowed to see that *SOX17* patterns resembling the ones from Warmflash et al. (2014) were possible without micropatterns.

The additional staining performed to further test if DP cells were indeed PGCs bore fruit as cells with PGC morphology were found, in addition to staining for all three markers[136]. This supports reports that human germ cells can be identified by markers composing the latter

combination[27,44]. For these reasons, the combination *PDPN+POU5F1+TFAP2C* can be considered a potentially strong combination when identifying PGCs (Fig. 3.3.3). Further combinations could have been used, such as, anti-*SOX17* + anti- Oct4. Although, due to time and logistic constraints, combination staining attempts were limited.

There is also a possibility that the Matrigel coating, which originates from a mouse tumour had a role in favouring the PGC fate, thought to have been found in this study. It has been shown that these tumours originate from parietal endoderm tissue[137–141]. One other reason that supports this is the influence that ECM-cell interactions on cultured cells paths of differentiation[142].

5. Conclusions and future perspectives.

5.1 – Transcriptome analysis for the selection of combinations targeting PGCs

Fortunately, the combinations used allowed the main objective to be met and *in vitro* differentiated hPGCs could be detected. However, caution is necessary when choosing the markers to be stained and when differentiation spans cell types with common links in their phenotype (ex: pluripotency genes). One example that stood out in this study were the genes *PDPN* and *POU5F1*, which are common to both the starting and the target cell type (PSCs and PGCs).

Particularly important as well, is to make sure that the information relied upon for performing computation expression analysis is well understood. This means that, DFs must be well annotated, as to consistently visualize cells with correct identities. Complementing a good annotation and proper use of the correct tools(functions) in the chosen software(s) is fundamental for accurately combining multiple conditions according to determined sets of genes ultimately enabling a truthful and reliable analysis of the data. In cases when possible cell type markers intersect between cell lines, as mentioned above, such measures are not to be ignored.

For the future, a better understanding and improvement of scRNA-seq experimental design would lead to the development of better algorithms for processing scRNA-seq outputs.

Only after, fulfilling the aforementioned conditions, were the combinations selected for immunostaining. This reinforces the conclusion that the cells shown in the bottom last image from section 3.3 of the results are most likely hPGCs generated *in vitro*.

5.2 – Micropatterned chips protocol requires further optimization for the cell lines used

The micropatterned chips experiment lack of success did not allow for a search for PGCs *in vitro*. Most importantly, an inability to fully cover the micropatterns and reach a stage at which the three germ layer phenotypes coexisted in an organized fashion, as rings, was not possible. Of note is despite this lab´s large efforts to optimize the Deglicerti et al. (2016) protocol, prior to the experiment shown in this study. For these reasons, further developments of this platform could be extremely beneficial for PGC development research. This will be crucial the increase of cell attachment efficiency or refinement of medium composition. Nonetheless, hPGCs could be identified using hESCs seeded on Matrigel coated coverslips and by immunostaining cells with the selected genes.

5.3 – Imaging

Widefield fluorescence was able to identify potential PGCs according to the combinations obtained from the bioinformatics analysis with success. This was due to the combination of well chosen datasets and a sound immunostaining protocol. Moreover, using the DM6B-Z, a microscope that allowed imaging of the cells with adequate resolution and image processing proved sufficient for this purpose.

Although, confocal imaging would allow for a better understanding on whether some 3D cell structures can be used to find potential PGCs. Including, where could these cells preferentially differentiate in terms of spatial location around or in these structures. The IMARIS software would also be applicable in this instance and could locate and count potential PGCs quickly on an entire coverslip. This capability of the program in combination with confocal microscopy would easily enable quantitative methods to access to efficiency of the differentiation.

BIBLIOGRAPHY

[1] Irie N, Tang WWC, Azim Surani M. Germ cell specification and pluripotency in mammals: a perspective from early embryogenesis. Reprod Med Biol 2014;13:203–15. https://doi.org/10.1007/s12522-014-0184-2.

[2] Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, et al. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. Dev Cell 2015;35:366–82. https://doi.org/10.1016/j.devcel.2015.10.011.

[3] Cooke HJ, Saunders PTK. Mouse models of male infertility. Nat Rev Genet 2002;3:790–801. https://doi.org/10.1038/nrg911.

[4] Ferlin A, Raicu F, Gatta V, Zuccarello D, Palka G, Foresta C. Male infertility: role of genetic background. Reprod Biomed Online 2007;14:734–45. https://doi.org/10.1016/s1472-6483(10)60677-3.

[5] Hwang K, Yatsenko AN, Jorgez CJ, Mukherjee S, Nalam RL, Matzuk MM, et al. Mendelian genetics of male infertility. Ann N Y Acad Sci 2010;1214:E1–17. https://doi.org/10.1111/j.1749-6632.2010.05917.x.

[6] Pellicer A, Ruiz A, Castellvi RM, Calatayud C, Ruiz M, Tarin JJ, et al. Is the retrieval of high numbers of oocytes desirable in patients treated with gonadotrophin-releasing hormone analogues (GnRHa) and gonadotrophins? Hum Reprod 1989;4:536–40. https://doi.org/10.1093/oxfordjournals.humrep.a136940.

[7] Macklon NS, Stouffer RL, Giudice LC, Fauser BCJM. The Science behind 25 Years of Ovarian Stimulation for in Vitro Fertilization. Endocr Rev 2006;27:170–207. https://doi.org/10.1210/er.2005-0015.

[8] Peters BA, Kermani BG, Alferov O, Agarwal MR, McElwain MA, Gulbahce N, et al. Detection and phasing of single base de novo mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing. Genome Res 2015;25:426–34. https://doi.org/10.1101/gr.181255.114.

[9] Cavaliere G. A 14-day limit for bioethics: the debate over human embryo research. BMC Med Ethics 2017;18:38. https://doi.org/10.1186/s12910-017-0198-5.

[10] Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 2006;126:663–76. https://doi.org/10.1016/j.cell.2006.07.024.

[11] Guillomot M. Cellular interactions during implantation in domestic ruminants. J Reprod Fertil Suppl 1995;49:39–51.

[12] Viebahn C. The anterior margin of the mammalian gastrula: comparative and phylogenetic aspects of its role in axis formation and head induction. Curr Top Dev Biol 1999;46:63–103.

[13] Schatten H, Sun Q-Y. The role of centrosomes in mammalian fertilization and its significance for ICSI. Mol Hum Reprod 2009;15:531–8. https://doi.org/10.1093/molehr/gap049.

[14] Dean W, Ferguson-Smith A. Genomic imprinting: Mother maintains methylation marks. Curr Biol 2001;11:R527–30. https://doi.org/10.1016/S0960-9822(01)00311-6.

[15] Hue I, Renard JP, Viebahn C. Brachyury is expressed in gastrulating bovine embryos well ahead of implantation. Dev Genes Evol 2001;211:157–9.

[16] Chuva de Sousa Lopes SM, Roelen BAJ. Primordial germ cell specification: the importance of being "blimped". Histol Histopathol 2008;23:1553–61. https://doi.org/10.14670/HH-23.1553.

[17] Sasaki K, Nakamura T, Okamoto I, Yabuta Y, Iwatani C, Tsuchiya H, et al. The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. Dev Cell 2016;39:169–85. https://doi.org/10.1016/j.devcel.2016.09.007.

[18] Boroviak T, Stirparo GG, Dietmann S, Hernando-Herraez I, Mohammed H, Reik W, et al. Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. Development 2018;145:dev167833. https://doi.org/10.1242/dev.167833.

[19] Kobayashi T, Zhang H, Tang WWC, Irie N, Withey S, Klisch D, et al. Principles of early human development and germ cell program from conserved model systems. Nature 2017;546:416–20. https://doi.org/10.1038/nature22812.

[20] Lawson KA, Pedersen RA. Clonal analysis of cell fate during gastrulation and early neurulation in the mouse. Ciba Found Symp 1992. https://doi.org/10.1002/9780470514221.ch2.

[21] Lawson KA, Hage WJ. Clonal analysis of the origin of primordial germ cells in the mouse. Ciba Found Symp 1994;182:68–84; discussion 84-91.

[22] Ginsburg M, Snow M, McLaren A. Primordial germ cells in the mouse embryo during gastrulation. Development 1990;110.

[23] Lawson KA, Hage WJ. Clonal analysis of the origin of primordial germ cells in the mouse. Ciba Found Symp 1994;182:68–84; discussion 84-91.

[24] Lawson KA, Hage WJ. Clonal Analysis of the Origin of Primordial Germ Cells in the Mouse. Ciba Found. Symp., vol. 182, 2007, p. 68–91. https://doi.org/10.1002/9780470514573.ch5.

[25] Bertocchini F, Chuva de Sousa Lopes SM. Germline development in amniotes: A paradigm shift in primordial germ cell specification. BioEssays 2016;38:791–800. https://doi.org/10.1002/bies.201600025.

[26] Saitou M, Miyauchi H. Gametogenesis from Pluripotent Stem Cells. Cell Stem Cell 2016;18:721–35. https://doi.org/10.1016/j.stem.2016.05.001.

[27] Tang WWC, Kobayashi T, Irie N, Dietmann S, Surani MA. Specification and epigenetic programming of the human germ line. Nat Rev Genet 2016;17:585–600. https://doi.org/10.1038/nrg.2016.88.

[28] Nagamatsu G, Hayashi K. Stem cells, in vitro gametogenesis and male fertility.

Reproduction 2017;154:F79–91. https://doi.org/10.1530/REP-17-0510.

[29] Nichols J, Smith A. Pluripotency in the embryo and in culture. Cold Spring Harb Perspect Biol 2012;4:1–14. https://doi.org/10.1101/cshperspect.a008128.

[30] Tam PPL, Behringer RR. Mouse gastrulation: The formation of a mammalian body plan. Mech Dev 1997;68:3–25. https://doi.org/10.1016/S0925-4773(97)00123-8.

[31] Aramaki S, Hayashi K, Kurimoto K, Ohta H, Yabuta Y, Iwanari H, et al. A Mesodermal Factor, T, Specifies Mouse Germ Cell Fate by Directly Activating Germline Determinants. Dev Cell 2013;27:516–29. https://doi.org/10.1016/J.DEVCEL.2013.11.001.

[32] Yamaji M, Seki Y, Kurimoto K, Yabuta Y, Yuasa M, Shigeta M, et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat Genet 2008;40:1016–22. https://doi.org/10.1038/ng.186.

[33] Ohinata Y, Ohta H, Shigeta M, Yamanaka K, Wakayama T, Saitou M. A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. Cell 2009;137:571–84. https://doi.org/10.1016/j.cell.2009.03.014.

[34] Irie N, Weinberger L, Tang WWC, Kobayashi T, Viukov S, Manor YS, et al. Article SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. Cell 2015;160:253–68. https://doi.org/10.1016/j.cell.2014.12.013.

[35] Nakaki F, Hayashi K, Ohta H, Kurimoto K, Yabuta Y, Saitou M. Induction of mouse germ-cell fate by transcription factors in vitro. Nature 2013;501:222–6. https://doi.org/10.1038/nature12417.

[36] Saitou M, Payer B, O'Carroll D, Ohinata Y, Surani MA. Blimp1 and the Emergence of the Germ Line during Development in the Mouse. Cell Cycle 2005;4:1736–40. https://doi.org/10.4161/cc.4.12.2209.

[37] Weber S, Eckert D, Nettersheim D, Gillis AJM, Schäfer S, Kuckenberg P, et al. Critical Function of AP-2gamma/TCFAP2C in Mouse Embryonic Germ Cell Maintenance1. Biol Reprod 2010;82:214–23. https://doi.org/10.1095/biolreprod.109.078717.

[38] de Jong J, Stoop H, Gillis A, van Gurp R, van de Geijn G-J, Boer M de, et al. Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications. J Pathol 2008;215:21–30. https://doi.org/10.1002/path.2332.

[39] Campolo F, Gori M, Favaro R, Nicolis S, Pellegrini M, Botti F, et al. Essential Role of Sox2 for the Establishment and Maintenance of the Germ Cell Line. Stem Cells 2013;31:1408–21. https://doi.org/10.1002/stem.1392.

[40] Irie N, Sybirna A, Surani MA. What Can Stem Cell Models Tell Us About Human Germ Cell Biology? Curr. Top. Dev. Biol., vol. 129, 2018, p. 25–65. https://doi.org/10.1016/bs.ctdb.2018.02.010.

[41] Yokobayashi S, Okita K, Nakagawa M, Nakamura T, Yabuta Y, Yamamoto T, et al. Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate†. Biol Reprod 2017;96:1154–66. https://doi.org/10.1093/biolre/iox038.

[42] Kojima Y, Sasaki K, Yokobayashi S, Sakai Y, Nakamura T, Yabuta Y, et al. Evolutionarily Distinctive Transcriptional and Signaling Programs Drive Human Germ Cell Lineage Specification from Pluripotent Stem Cells. Cell Stem Cell 2017;21:517-532.e5. https://doi.org/10.1016/j.stem.2017.09.005.

[43] Sasaki K, Yokobayashi S, Nakamura T, Okamoto I, Yabuta Y, Kurimoto K, et al. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells Article Robust

In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. Stem Cell 2015;17:178–94. https://doi.org/10.1016/j.stem.2015.06.014.

[44] Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, et al. A unique gene regulatory network resets the human germline epigenome for development. Cell 2015;161:1453–67. https://doi.org/10.1016/j.cell.2015.04.053.

[45] Fujimoto T, Miyayama Y, Fuyuta M. The origin, migration and fine morphology of human primordial germ cells. Anat Rec 1977;188:315–29. https://doi.org/10.1002/ar.1091880305.

[46] Matsui Y, Zsebo K, Hogan BLM. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell 1992;70:841–7. https://doi.org/10.1016/0092-8674(92)90317-6.

[47] Resnick JL, Bixler LS, Cheng L, Donovan PJ. Long-term proliferation of mouse primordial germ cells in culture. Nature 1992;359:550–1. https://doi.org/10.1038/359550a0.

[48] Labosky PA, Barlow DP, Hogan BLM. Embryonic Germ Cell Lines and Their Derivation from Mouse Primordial Germ Cells. Ciba Found. Symp., vol. 182, 2007, p. 157–78. https://doi.org/10.1002/9780470514573.ch9.

[49] Hargan-Calvopina J, Taylor S, Cook H, Chiang Y-S, Chen P-Y, Correspondence ATC. Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. Dev Cell 2016;39:75–86. https://doi.org/10.1016/j.devcel.2016.07.019.

[50] Eguizabal C, Herrera L, De Oñate L, Montserrat N, Hajkova P, Izpisua Belmonte JC. Characterization of the Epigenetic Changes During Human Gonadal Primordial Germ Cells Reprogramming. Stem Cells 2016;34:2418–28. https://doi.org/10.1002/stem.2422.

[51] Fernandes MG, Bialecka M, Salvatori DCF, Chuva De Sousa Lopes SM. Characterization of migratory primordial germ cells in the aorta-gonad-mesonephros of a 4.5-week-old human embryo: A toolbox to evaluate in vitro early gametogenesis. Mol Hum Reprod 2018;24:233–43. https://doi.org/10.1093/molehr/gay011.

[52] Kagiwada S, Kurimoto K, Hirota T, Yamaji M, Saitou M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J 2012;32:340–53. https://doi.org/10.1038/emboj.2012.331.

[53] Saitou M, Yamaji M. Primordial Germ Cells in Mice n.d.:1–20.

[54] Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 2012;48:849–62. https://doi.org/10.1016/j.molcel.2012.11.001.

[55] Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 2006;442:199–202. https://doi.org/10.1038/nature04917.

[56] Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 2008;31:785–99. https://doi.org/10.1016/j.molcel.2008.09.003.

[57] Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev 2008;22:908–17. https://doi.org/10.1101/gad.1640708.

[58] Roovers EF, Rosenkranz D, Mahdipour M, Han C-T, He N, Chuva de Sousa Lopes SM,

et al. Piwi proteins and piRNAs in mammalian oocytes and early embryos. Cell Rep 2015;10:2069–82. https://doi.org/10.1016/j.celrep.2015.02.062.

[59] Fujimoto T, Yoshinaga K, Kono I. Distribution of fibronectin on the migratory pathway of primordial germ cells in mice. Anat Rec 1985;211:271–8. https://doi.org/10.1002/ar.1092110307.

[60] De Felici M. Regulation of primordial germ cell development in the mouse. Int J Dev Biol 2000;44:575–80.

[61] Alvarez-Buylla A, Merchant-Larios H. Mouse primordial germ cells use fibronectin as a substrate for migration. Exp Cell Res 1986;165:362–8. https://doi.org/10.1016/0014-4827(86)90590-2.

[62] Heasman J, Hynes RO, Swan AP, Thomas V, Wylie CC. Primordial germ cells of Xenopus embryos: The role of fibronectin in their adhesion during migration. Cell 1981;27:437–47. https://doi.org/10.1016/0092-8674(81)90385-8.

[63] Clark JM, Eddy EM. Fine structural observations on the origin and associations of primordial germ cells of the mouse. Dev Biol 1975;47:136–55. https://doi.org/10.1016/0012-1606(75)90269-9.

[64] García-Castro MI, Anderson R, Heasman J, Wylie C. Interactions between germ cells and extracellular matrix glycoproteins during migration and gonad assembly in the mouse embryo. J Cell Biol 1997;138:471–80. https://doi.org/10.1083/jcb.138.2.471.

[65] Wylie, C.C., D. Stott and PJD. Primordial germ cell migration. In The Cellular Basis of Morphogenesis. In: Browder L, editor. Primordial germ cell Migr. 2nd ed., New York: Plenum Press, New York; 1986, p. 433–450.

[66] Soto-Suazo M, Abrahamsohn PA, Pereda J, Zorn TMT. Distribution and space-time relationship of proteoglycans in the extracellular matrix of the migratory pathway of primordial germ cells in mouse embryos. Tissue Cell 1999;31:291–300. https://doi.org/10.1054/tice.1999.0041.

[67] García-Castro MI, Anderson R, Heasman J, Wylie C. Interactions between germ cells and extracellular matrix glycoproteins during migration and gonad assembly in the mouse embryo. J Cell Biol 1997;138:471–80. https://doi.org/10.1083/jcb.138.2.471.

[68] Pereda J, Zorn T, Soto-Suazo M. Migration of human and mouse primordial germ cells and colonization of the developing ovary: An ultrastructural and cytochemical study. Microsc Res Tech 2006;69:386–95. https://doi.org/10.1002/jemt.20298.

[69] De Felici M, Klinger FG, Farini D, Scaldaferri ML, Iona S, Lobascio M. Establishment of oocyte population in the fetal ovary: primordial germ cell proliferation and oocyte programmed cell death. Reprod Biomed Online 2005;10:182–91. https://doi.org/10.1016/S1472-6483(10)60939-X.

[70] Wakayama S, Cibelli JB, Wakayama T. Effect of Timing of the Removal of Oocyte Chromosomes Before or After Injection of Somatic Nucleus on Development of NT Embryos. Cloning Stem Cells 2003;5:181–9. https://doi.org/10.1089/153623003769645848.

[71] De Felici M. In vitro adhesion of mouse fetal germ cells to extracellular matrix components. Cell Differ Dev 1998;26:87–96.

[72] Anderson R, Fassler R, Georges-Labouesse E, Hynes RO, Bader BL, Kreidberg JA, et al. Mouse primordial germ cells lacking beta1 integrins enter the germline but fail to migrate normally to the gonads. Development 1999;126.

[73] Anderson R, Copeland TK, Schöler H, Heasman J, Wylie C. The onset of germ cell migration in the mouse embryo. Mech Dev 2000;91:61–8. https://doi.org/10.1016/S0925-4773(99)00271-3.

[74] Fassler R, Meyer M. Consequences of lack of beta 1 integrin gene expression in mice. Genes Dev 1995;9:1896–908. https://doi.org/10.1101/gad.9.15.1896.

[75] Stephens LE, Sutherland AE, Klimanskaya I V, Andrieux A, Meneses J, Pedersen RA, et al. Deletion of beta 1 integrins in mice results in inner cell mass failure and peri-implantation lethality. Genes Dev 1995;9:1883–95. https://doi.org/10.1101/gad.9.15.1883.

[76] Hill MA. Embryology Primordial Germ Cell Migration Movie. Dr Mark Hill 2019, UNSW Embryol ISBN 978 0 7334 2609 4 - UNSW CRICOS Provid Code No 00098G 2019. https://embryology.med.unsw.edu.au/embryology/index.php/Primordial_Germ_Cell_Migration_Movie#Reference.

[77] Morgani SM, Canham MA, Nichols J, Sharov AA, Migueles RP, Ko MSH, et al. Totipotent Embryonic Stem Cells Arise in Ground-State Culture Conditions. Cell Rep 2013;3:1945–57. https://doi.org/10.1016/J.CELREP.2013.04.034.

[78] Cahan P, Daley GQ. Origins and implications of pluripotent stem cell variability and heterogeneity. Nat Rev Mol Cell Biol 2013;14:357–68. https://doi.org/10.1038/nrm3584.

[79] Kolodziejczyk AA, Kim JK, Tsang JCH, Ilicic T, Henriksson J, Natarajan KN, et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell 2015;17:471–85. https://doi.org/10.1016/j.stem.2015.09.011.

[80] Singh AM. Cell Cycle-Driven Heterogeneity: On the Road to Demystifying the Transitions between "Poised" and "Restricted" Pluripotent Cell States. Stem Cells Int 2015;2015:219514. https://doi.org/10.1155/2015/219514.

[81] Torres-Padilla M-E, Chambers I. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. Development 2014;141:2173–81. https://doi.org/10.1242/DEV.102624.

[82] Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 2015;161:1187–201. https://doi.org/10.1016/j.cell.2015.04.044.

[83] Gkountela S, Zhang KX, Shafiq TA, Liao W-W, Hargan-Calvopiña J, Chen P-Y, et al. DNA Demethylation Dynamics in the Human Prenatal Germline. Cell 2015;161:1425–36. https://doi.org/10.1016/j.cell.2015.05.012.

[84] Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell 2011. https://doi.org/10.1016/j.cell.2011.06.052.

[85] Nichols J, Smith A. Naive and Primed Pluripotent States. Cell Stem Cell 2009;4:487–92. https://doi.org/10.1016/j.stem.2009.05.015.

[86] Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature 2007;448:196–9. https://doi.org/10.1038/nature05972.

[87] Messmer T, von Meyenn F, Savino A, Santos F, Mohammed H, Lun ATL, et al. Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. Cell Rep 2019;26:815-824.e4.

https://doi.org/10.1016/j.celrep.2018.12.099.

[88] Brimble SN, Sherrer ES, Uhl EW, Wang E, Kelly S, Merrill AH, et al. The Cell Surface Glycosphingolipids SSEA-3 and SSEA-4 Are Not Essential for Human ESC Pluripotency. Stem Cells 2006;25:54–62. https://doi.org/10.1634/stemcells.2006-0232.

[89] De Felici M. Origin, Migration, and Proliferation of Human Primordial Germ Cells. Oogenesis, London: Springer London; 2013, p. 19–37. https://doi.org/10.1007/978-0-85729-826-3_2.

[90] O'Connor MD, Kardel MD, Iosfina I, Youssef D, Lu M, Li MM, et al. Alkaline Phosphatase-Positive Colony Formation Is a Sensitive, Specific, and Quantitative Indicator of Undifferentiated Human Embryonic Stem Cells. Stem Cells 2008;26:1109–16. https://doi.org/10.1634/stemcells.2007-0801.

[91] Tilgner K, Atkinson SP, Golebiewska A, Stojković M, Lako M, Armstrong L. Isolation of Primordial Germ Cells from Differentiating Human Embryonic Stem Cells. Stem Cells 2008;26:3075–85. https://doi.org/10.1634/stemcells.2008-0289.

[92] Park TS, Galic Z, Conway AE, Lindgren A, van Handel BJ, Magnusson M, et al. Derivation of Primordial Germ Cells from Human Embryonic and Induced Pluripotent Stem Cells Is Significantly Improved by Coculture with Human Fetal Gonadal Cells. Stem Cells 2009;27:783–95. https://doi.org/10.1002/stem.13.

[93] Bucay N, Yebra M, Cirulli V, Afrikanova I, Kaido T, Hayek A, et al. A Novel Approach for the Derivation of Putative Primordial Germ Cells and Sertoli Cells from Human Embryonic Stem Cells. Stem Cells 2009;27:68–77. https://doi.org/10.1634/stemcells.2007-1018.

[94] Kerr CL, Hill CM, Blumenthal PD, Gearhart JD. Expression of pluripotent stem cell markers in the human fetal ovary. Hum Reprod 2008;23:589–99. https://doi.org/10.1093/humrep/dem411.

[95] Sugawa F, Araúzo- Bravo MJ, Yoon J, Kim K, Aramaki S, Wu G, et al. Human primordial germ cell commitment *in vitro* associates with a unique PRDM14 expression profile. EMBO J 2015;34:1009–24. https://doi.org/10.15252/embj.201488049.

[96] Harrison SE, Sozen B, Christodoulou N, Kyprianou C, Zernicka-Goetz M. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. Science 2017;356:eaal1810. https://doi.org/10.1126/science.aal1810.

[97] Sozen B, Amadei G, Cox A, Wang R, Na E, Czukiewska S, et al. Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. Nat Cell Biol 2018;20:979–89. https://doi.org/10.1038/s41556-018-0147-7.

[98] Warmflash A, Sorre B, Etoc F, Siggia ED, Brivanlou AH. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nat Methods 2014;11:847–54. https://doi.org/10.1038/nmeth.3016.

[99] Deglincerti A, Etoc F, Guerra MC, Martyn I, Metzger J, Ruzo A, et al. Self-organization of human embryonic stem cells on micropatterns. Nat Protoc 2016;11:2223–32. https://doi.org/10.1038/nprot.2016.131.

[100] Deglincerti A, Croft GF, Pietila LN, Zernicka-Goetz M, Siggia ED, Brivanlou AH. Self-organization of the in vitro attached human embryo. Nature 2016;533:251–4. https://doi.org/10.1038/nature17948.

[101] Martyn I, Kanno TY, Ruzo A, Siggia ED, Brivanlou AH. Self-organization of a human organizer by combined Wnt and Nodal signalling. Nature 2018;558:132–5.

https://doi.org/10.1038/s41586-018-0150-y.

[102] Etoc F, Metzger J, Ruzo A, Kirst C, Yoney A, Ozair MZ, et al. A Balance between Secreted Inhibitors and Edge Sensing Controls Gastruloid Self-Organization. Dev Cell 2016;39:302–15. https://doi.org/10.1016/j.devcel.2016.09.016.

[103] Irie N, Weinberger L, Tang WWC, Kobayashi T, Viukov S, Manor YS, et al. SOX17 is a critical specifier of human primordial germ cell fate. Cell 2015;160:253–68. https://doi.org/10.1016/j.cell.2014.12.013.

[104] Li L, Dong J, Yan L, Yong J, Liu X, Hu Y, et al. Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. Cell Stem Cell 2017;20:891–2. https://doi.org/10.1016/j.stem.2017.05.009.

[105] Li L, Dong J, Yan L, Yong J, Liu X, Hu Y, et al. Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. Cell Stem Cell 2017;20:858-873.e4. https://doi.org/10.1016/j.stem.2017.03.007.

[106] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 2012;9:676–82. https://doi.org/10.1038/nmeth.2019.

[107] Messmer T, von Meyenn F, Savino A, Santos F, Mohammed H, Lun ATL, et al. Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. Cell Rep 2019;26:815-824.e4. https://doi.org/10.1016/j.celrep.2018.12.099.

[108] Yabuta Y, Kurimoto K, Ohinata Y, Seki Y, Saitou M. Gene Expression Dynamics During Germline Specification in Mice Identified by Quantitative Single-Cell Gene Expression Profiling1. Biol Reprod 2006;75:705–16. https://doi.org/10.1095/biolreprod.106.053686.

[109] Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev 2008;22:908–17. https://doi.org/10.1101/gad.1640708.

[110] Chen ZX, Riggs AD. DNA methylation and demethylation in mammal. J Biol Chem 2011;286:18347–53. https://doi.org/10.1074/jbc.R110.205286.

[111] Miller I, Min M, Yang C, Tian C, Gookin S, Carter D, et al. Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence. Cell Rep 2018;24:1105-1112.e5. https://doi.org/10.1016/j.celrep.2018.06.110.

[112] Nasmyth K. Segregating sister genomes: The molecular biology of chromosome separation. Science (80-) 2002;297:559–65. https://doi.org/10.1126/science.1074757.

[113] Ma Y-Q, Qin J, Wu C, Plow EF. Kindlin-2 (Mig-2): a co-activator of beta3 integrins. J Cell Biol 2008;181:439–46. https://doi.org/10.1083/jcb.200710196.

[114] Qu H, Tu Y, Shi X, Larjava H, Saleem MA, Shattil SJ, et al. Kindlin-2 regulates podocyte adhesion and fibronectin matrix deposition through interactions with phosphoinositides and integrins. J Cell Sci 2011;124:879–91. https://doi.org/10.1242/JCS.076976.

[115] Hargan-Calvopina J, Taylor S, Cook H, Hu Z, Lee SA, Yen M-R, et al. Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. Dev Cell 2016;39:75–86. https://doi.org/10.1016/j.devcel.2016.07.019.

[116] Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell

2013;49:825–37. https://doi.org/10.1016/j.molcel.2013.01.038.

[117] Yokobayashi S, Okita K, Nakagawa M, Nakamura T, Yabuta Y, Yamamoto T, et al. Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate†. Biol Reprod 2017;96:1154–66. https://doi.org/10.1093/biolre/iox038.

[118] Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, et al. Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 2013;10:1093–5. https://doi.org/10.1038/nmeth.2645.

[119] Grün D, Kester L, Van Oudenaarden A. Validation of noise models for single-cell transcriptomics. Nat Methods 2014;11:637–40. https://doi.org/10.1038/nmeth.2930.

[120] Ye H, Wang Q. Efficient Generation of Non-Integration and Feeder-Free Induced Pluripotent Stem Cells from Human Peripheral Blood Cells by Sendai Virus. Cell Physiol Biochem 2018;50:1318–31. https://doi.org/10.1159/000494589.

[121] Laursen SB, Møllgard K, Olesen C, Oliveri RS, Brøchner CB, Byskov AG, et al. Regional differences in expression of specific markers for human embryonic stem cells. Reprod Biomed Online 2007;15:89–98. https://doi.org/10.1016/S1472-6483(10)60697-9.

[122] Weinberger L, Ayyash M, Novershtern N, Hanna JH. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol 2016;17:155–69. https://doi.org/10.1038/nrm.2015.28.

[123] Yamashiro C, Hirota T, Kurimoto K, Nakamura T, Yabuta Y, Nagaoka SI, et al. Persistent Requirement and Alteration of the Key Targets of PRDM1 During Primordial Germ Cell Development in Mice1. Biol Reprod 2016;94. https://doi.org/10.1095/biolreprod.115.133256.

[124] Mitani T, Yabuta Y, Ohta H, Nakamura T, Yamashiro C, Yamamoto T, et al. Principles for the regulation of multiple developmental pathways by a versatile transcriptional factor, BLIMP1. Nucleic Acids Res 2017;45:12152–69. https://doi.org/10.1093/nar/gkx798.

[125] Ohinata Y, Payer B, O'Carroll D, Ancelin K, Ono Y, Sano M, et al. Blimp1 is a critical determinant of the germ cell lineage in mice. Nature 2005;436:207–13. https://doi.org/10.1038/nature03813.

[126] Richardson BE, Lehmann R. Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nat Rev Mol Cell Biol 2010;11:37–49. https://doi.org/10.1038/nrm2815.

[127] Salomonsson A, Jönsson M, Isaksson S, Karlsson A, Jönsson P, Gaber A, et al. Histological specificity of alterations and expression of *KIT* and *KITLG* in non-small cell lung carcinoma. Genes, Chromosom Cancer 2013;52:1088–96. https://doi.org/10.1002/gcc.22103.

[128] Teicher BA, Fricker SP. CXCL12 (SDF-1)/CXCR4 Pathway in Cancer. Clin Cancer Res 2010;16:2927–31. https://doi.org/10.1158/1078-0432.CCR-09-2329.

[129] Lim S, Kaldis P. Cdks, cyclins and CKIs: Roles beyond cell cycle regulation. Dev 2013;140:3079–93. https://doi.org/10.1242/dev.091744.

[130] Heyn R, Makabe S, Motta PM. Ultrastructural dynamics of human testicular cords from 6 to 16 weeks of embryonic development. Study by transmission and high resolution scanning electron microscopy. Ital J Anat Embryol 1998;103:17–29.

[131] Richardson BE, Lehmann R. Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nat Rev Mol Cell Biol 2010;11:37–49.

https://doi.org/10.1038/nrm2815.

[132] Addis RC, Prasad MK, Yochem RL, Zhan X, Sheets TP, Axelman J, et al. OCT3/4 regulates transcription of histone deacetylase 4 (Hdac4) in mouse embryonic stem cells. J Cell Biochem 2010;111:391–401. https://doi.org/10.1002/jcb.22707.

[133] Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell 2012;48:849–62. https://doi.org/10.1016/j.molcel.2012.11.001.

[134] O'Neill C. The epigenetics of embryo development. Anim Front 2015;5:42–9. https://doi.org/10.2527/af.2015-0007.

[135] Solnica-Krezel L, Sepich DS. Gastrulation: Making and Shaping Germ Layers. Annu Rev Cell Dev Biol 2012;28:687–717. https://doi.org/10.1146/annurev-cellbio-092910-154043.

[136] Donovan PJ, Stott D, Cairns LA, Heasman J, Wylie CC. Migratory and postmigratory mouse primordial germ cells behave differently in culture. Cell 1986;44:831–8. https://doi.org/10.1016/0092-8674(86)90005-X.

[137] Vukicevic S, Kleinman HK, Luyten FP, Roberts AB, Roche NS, Reddi AH. Identification of multiple active growth factors in basement membrane matrigel suggests caution in interpretation of cellular activity related to extracellular matrix components. Exp Cell Res 1992;202:1–8. https://doi.org/10.1016/0014-4827(92)90397-Q.

[138] Kleinman HK, McGarvey ML, Liotta LA, Robey PG, Tryggvason K, Martin GR. Isolation and Characterization of Type IV Procollagen, Laminin, and Heparan Sulfate Proteoglycan from the EHS Sarcoma. Biochemistry 1982;21:6188–93. https://doi.org/10.1021/bi00267a025.

[139] Kleinman HK, McGarvey ML, Hassell JR, Star VL, Cannon FB, Laurie GW, et al. Basement membrane complexes with biological activity. Biochemistry 1986;25:312–8. https://doi.org/10.1021/bi00350a005.

[140] Mackay AR, Gomez DE, Cottam DW, Rees RC, Nason AM, Thorgeirsson UP. Identification of the 72-kDa (MMP-2) and 92-kDa (MMP-9) gelatinase/type IV collagenase in preparations of laminin and Matrigel. Biotechniques 1993;15:1048–51.

[141] Kleinman HK, Martin GR. Matrigel: Basement membrane matrix with biological activity. Semin Cancer Biol 2005;15:378–86. https://doi.org/10.1016/j.semcancer.2005.05.004.

[142] Villa-Diaz LG, Ross AM, Lahann J, Krebsbach PH. The evolution of human pluripotent stem cell culture: from feeder cells to synthetic coatings. Stem Cells 2013;31:1. https://doi.org/10.1002/STEM.1260.