Universiteit
Leiden
The Netherlands

Organ-specific barcodes in human fetal development and stem cell differentiation : the pancreas in the spotlight

Roost, M.S.

Citation
Roost, M. S. (2017, March 22). *Organ-specific barcodes in human fetal development and stem cell differentiation : the pancreas in the spotlight*. Retrieved from https://hdl.handle.net/1887/47069

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/47069

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/47069 holds various files of this Leiden University dissertation.

Author: Roost, M.S.
Title: Organ-specific barcodes in human fetal development and stem cell differentiation : the pancreas in the spotlight
Issue Date: 2017-03-22

ORGAN-SPECIFIC BARCODES

IN HUMAN FETAL DEVELOPMENT AND STEM CELL DIFFERENTIATION: THE PANCREAS IN THE SPOTLIGHT

MATTHIAS SHONA ROOST

ORGAN-SPECIFIC BARCODES

IN HUMAN FETAL DEVELOPMENT AND STEM CELL DIFFERENTIATION: THE PANCREAS IN THE SPOTLIGHT

MATTHIAS SHONA ROOST

Organ-Specific Barcodes in Human Fetal Development and Stem Cell Differentiation: The Pancreas in the Spotlight

2017, Matthias Shona Roost

All rights are reserved. No part of this thesis may be reproduced, stored, or transmitted in any form or by any means, without permission of the copyright owners.

ISBN: 978-3-033-06079-1
Layout by: Silvia Hugi Lory (henrygehtzummond.ch)
Photography and drawing by: Urs Gehbauer
Printed by: Druckerei Odermatt AG, Dallenwil, Switzerland
Font: Proxima Nova

The research presented in this thesis was funded by the Bontius Foundation, the Netherlands Organisation for Scientific Research (NWO, ASPASIA 015.007.037) and the Interuniversity Attraction Poles (IAP).

ORGAN-SPECIFIC BARCODES

IN HUMAN FETAL DEVELOPMENT AND STEM CELL DIFFERENTIATION: THE PANCREAS IN THE SPOTLIGHT

PROEFSCHRIFT
TER VERKRIJGING VAN
DE GRAAD VAN DOCTOR AAN DE UNIVERSITEIT LEIDEN,
OP GEZAG VAN RECTOR MAGNIFICUS PROF. MR. C.J.J.M. STOLKER,
VOLGENS BESLUIT VAN HET COLLEGE VOOR PROMOTIES
TE VERDEDIGEN OP WOENSDAG 22 MAART 2017
KLOKKE 16.15 UUR

door

MATTHIAS SHONA ROOST
geboren te Bern, Zwitserland
in 1983

Promotores	Prof. dr. C.L. Mummery
	Prof. dr. E.J.P. de Koning
Copromotor	Dr. S.M. Chuva de Sousa Lopes
Promotiecommissie	Prof. dr. P.C.W. Hogendoorn
	Prof. dr. J.J. Goeman
	Prof. dr. P. Slagboom
	Prof. dr. A. Sonnenberg (Netherlands Cancer Institute)
	Prof. dr. N. Hanley (University of Manchester)

CONTENTS

9	CHAPTER 1	**General introduction**
47	CHAPTER 2	**Lymphangiogenesis and angiogenesis during human fetal pancreas development**
93	CHAPTER 3	**KeyGenes, a tool to probe tissue differentiation using a human fetal transcriptional atlas**
141	CHAPTER 4	**DNA methylation landscapes of human fetal development**
197	CHAPTER 5	**DNA methylation and transcriptional trajectories during human development and reprogramming of isogenic induced pluripotent stem cells**
243	CHAPTER 6	**General discussion**
277	APPENDIX	**Summary** **Nederlandse samenvatting** **List of publications** **Curriculum vitae** **Acknowledgements**

CHAPTER 1

GENERAL INTRODUCTION

STEM CELL BACKGROUND IN DEVELOPMENT

The ultimate goal of stem cell differentiation protocols is the generation of mature, fully functional cell types from stem cells that mimic the actual *in vivo* differentiation of a particular cell type. In order to do so, it is essential to gain knowledge about the interactions of the cell type of interest with its surrounding cells and tissues as well as the extracellular matrix, anatomically called a «niche», during development. Besides functional testing, it is also necessary to compare and eventually match the gene expression and epigenetic profiles of the differentiated cells generated from stem cells to those of the cells and tissues of interest during their *in vivo* development to properly characterize and benchmark them.

The fusion of an egg and a sperm at fertilization is the first step of many highly regulated steps in the development of a new individual. Two features of fertilization that make it such a remarkable process are: (1) the acquisition of a totipotent state, which involves the nearly complete (epigenetic) reprogramming of the parental genomes [1, 2] **and (2) the fusion of the male and female pronuclei that results in remodeling of the chromatin and the entire chromosomal machinery, which eventually leads to the switch from meiosis to mitosis** [2, 3]**. After fertilization, the fertilized egg, or zygote, undergoes a series of cell cleavage divisions as it migrates towards the uterus resulting in similar cells called blastomeres. These then undergo a process called «compaction» in which the individual borders of the cells become less distinct in a structure called the morula** [4]**. This leads to the first lineage separation: the outside cells become trophectoderm (TE) and the inside cells the inner cell mass (ICM). After further cell divisions, another morphological process called «cavitation» takes place as the**

TE cells pump water inwards, which results in the formation of a fluid-filled cavity. At this stage the embryo is known as blastocyst and consists of an outer layer of TE cells that gives rise to cells of the placenta; and the ICM cells that will give rise to the embryo but also contribute to cells of the extraembryonic lineages [5]**. The blastocyst then physically attaches and invades the uterine wall, a process called implantation** [6]**. Due to implantation, the embryo is now able to connect to the maternal circulation that will provide the required nutrients and oxygen and transport the embryo's waste material. During pre-implantation development, many mammalian species show common developmental features as the most widely studied mammal, the laboratory mouse, but after implantation the developmental strategies of each mammalian species begin to diverge.**

Although the particular focus of the work in this thesis is human development, the mouse remains the framework on which most of the studies in the literature were based. In the mouse, the inner cell mass of the blastocyst undergoes a next lineage separation resulting in epiblast cells and primitive endoderm cells. During gastrulation, the epiblast develops into the three embryonic germ layers, namely the ectoderm, definitive endoderm and embryonic mesoderm, in addition to the extraembryonic mesoderm and the germ line; and during this developmental phase the embryo gains orientation («body axes») as determined by the position of the primitive streak [7]**. The three embryonic germ layers are distinct populations of cells with specific molecular signatures and give rise to all somatic cells of the embryo: the ectoderm is the outer layer and forms the neural system and skin, the endoderm is the inner layer and gives rise to many epithelial organs and the gastrointestinal tract, whilst the mesoderm forms the middle compartment and primarily gives rise to mesenchyme, blood, bones and muscle tissue. Once the initial body plan is specified by the three germ layers, the formation of mature functional tissues**

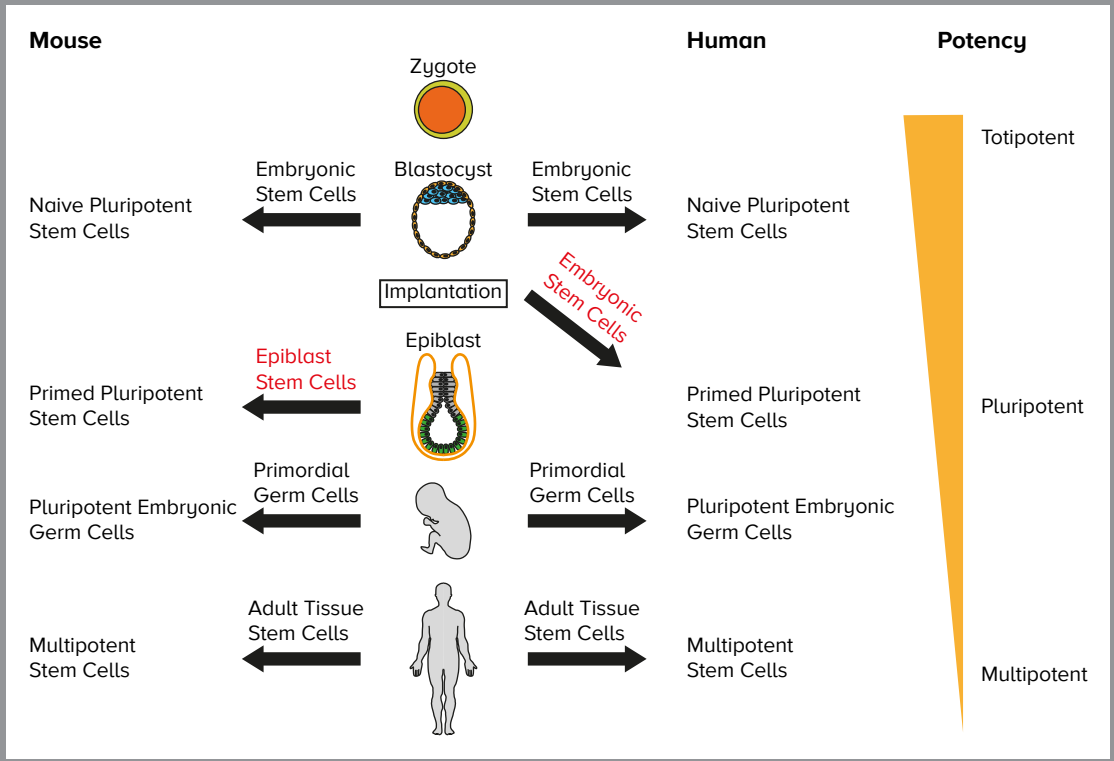
and organs is orchestrated during the complex processes of organ development.

The pancreas is one of the organs that are formed initially from the endoderm with a contribution from mesoderm during organogenesis. In the research of this thesis, the pancreas was used as a paradigm to study human organ development since so much is known about its development in mice and what causes its abnormal development and malfunction. The pancreas is an extraordinarily complex organ harboring two morphologically and functionally different compartments originating from the same progenitors in mouse and other species: the endocrine and the exocrine pancreas [8]. The exocrine pancreas, consisting of acinar and duct cells, produces the digestive fluids, which are eventually secreted into the duodenum [9]. On the other hand, the endocrine pancreas is organized in morphologically distinct compartments, the islets of Langerhans that contain 5 types of endocrine cells. The β- and the α-cells that play a key role in glucose metabolism, and δ-, ε- and PP cells [10]. Research on postimplantation human fetal stages has, for obvious ethical and legal reasons, been extrapolated from other species as stated earlier, and this is also the case for the pancreas [11]. Nevertheless, and despite limited availability of human embryos, recent studies have shed light on human pancreas development [12]. At Carnegie Stage 10 (CS), which equals 25–27 days post conception, the anterior intestinal portal emerges from the foregut endoderm, where four days later (CS12) pancreas specification is induced and is typically marked by the expression of the transcription factor PDX1 [11, 13]. This is followed by the appearance of the dorsal and ventral pancreatic buds at CS13, which eventually fuse and constitute the endocrine and exocrine compartments of the pancreas [14]. This process is accompanied by extensive proliferation of the pancreatic progenitor cells, which divide into two different cell populations that leads to a tip-trunk segregation at CS19 [13, 15]. Whereas the tip progenitor cells differentiate into the

acinar cells, the progenitor cells in the trunk are bipotent giving rise to the ductal and endocrine cells [11, 16]**. The first endocrine cells, predominantly β-cells, emerge from the ducts around 9 weeks of gestation marked by the transient upregulation of the transcription factor NGN3** [17–20]**. Further development to the mature pancreas then involves proper establishment of the blood vascular networks as well as the neural innervation** [21, 22]**. Work in this thesis has added new information to the understanding of the development of the human pancreas, which has contributed to setting the stage for deriving functional pancreas cells from pluripotent stem cells** [23]**. It is hoped that in the future these cells can be transplanted into patients with diabetes with regard to providing them with new insulin-secreting β-cells destroyed by the disease. Protocols for deriving pancreatic cells from pluripotent stem cells rely on deep understanding of pancreas development, but also on how cells in the embryo move from a pluripotent state, as in the inner cells mass, to fully differentiated somatic cells.**

PLURIPOTENCY AND PLURIPOTENT STEM CELLS

Stem cells exhibit two hallmark features, namely self-renewal, which means maintaining one cell in an undifferentiated state after cell division, and their ability to differentiate [24, 25]**. The potency of stem cells in culture and *in vivo*, i.e. to what range of cells they can differentiate, is essentially determined by the location and stage during development they are derived from** [Fig. 1]**. Although totipotency is conceptually the most interesting state, pluripotent stem cells (PSCs) represent the most promising source of cells not only for basic research, but also for applications such as cell replacement therapies, drug screening, toxicity testing and disease modeling** [26–28]**. The concept of pluripotency was initially defined by the potential of the cells to differentiate towards all three embryonic germ layers, namely ectoderm, mesoderm and endoderm, and emerged with the *in vitro* derivation of pluripotent cells from teratocarcinomas** [29]**.**



[Fig. 1]

Different states of stem cells.

Stem cells differ in potency, culture conditions, morphology, signaling pathways and their (epi)genetic makeup. Differences between murine and human stem cells are depicted in red.

The first pluripotent cell types identified were embryonal carcinoma cells, but those with embryonic origin were mouse embryonic stem cells [30, 31]**. Knowledge of the identity of these cells eventually led to the derivation of the first human embryonic stem cell lines** [32]**.**

Recent progress in developmental and cell biology linked to new technological possibilities led to the concept of different shades of pluripotency. Cell lines that are derived from the ICM of the blastocyst before implantation are classically known as embryonic stem cells (ESCs), whereas cell lines derived early during implantation from the epiblast are called epiblast stem cells (EpiSCs) [Fig. 1] [33, 34]**. Although those different cells share the global features of pluripotency, they differ in culture condition requirements, morphology, signaling circuits, genes expression programs and epigenetic status** [25, 26]**. Furthermore, it has been shown that there are species-specific differences regarding the appearance and culture conditions of pluripotent stem cells derived from the same stage, which led to introduction of the terms «naïve» and «primed» pluripotent states** [Fig. 1] [35–37]**. Whereas mouse ESCs exhibit a *bona fide* naïve, «ground-state» pluripotency, human ESCs are considered to acquire a rather EpiSC phenotype when compared to the mouse** [34]**. Recently, several groups have reported the generation of naïve human ESCs, which would provide a valuable resource considering the advantages of naïve mouse ESCs in the range of cell types they can form** [38–42]**. Other advantages of naïve ESCs include their homogenous character, high clonogenicity and the lack of differentiation bias and, therefore, they are easier to use for gene editing** [35, 43, 44]**.**

SOMATIC REPROGRAMMING

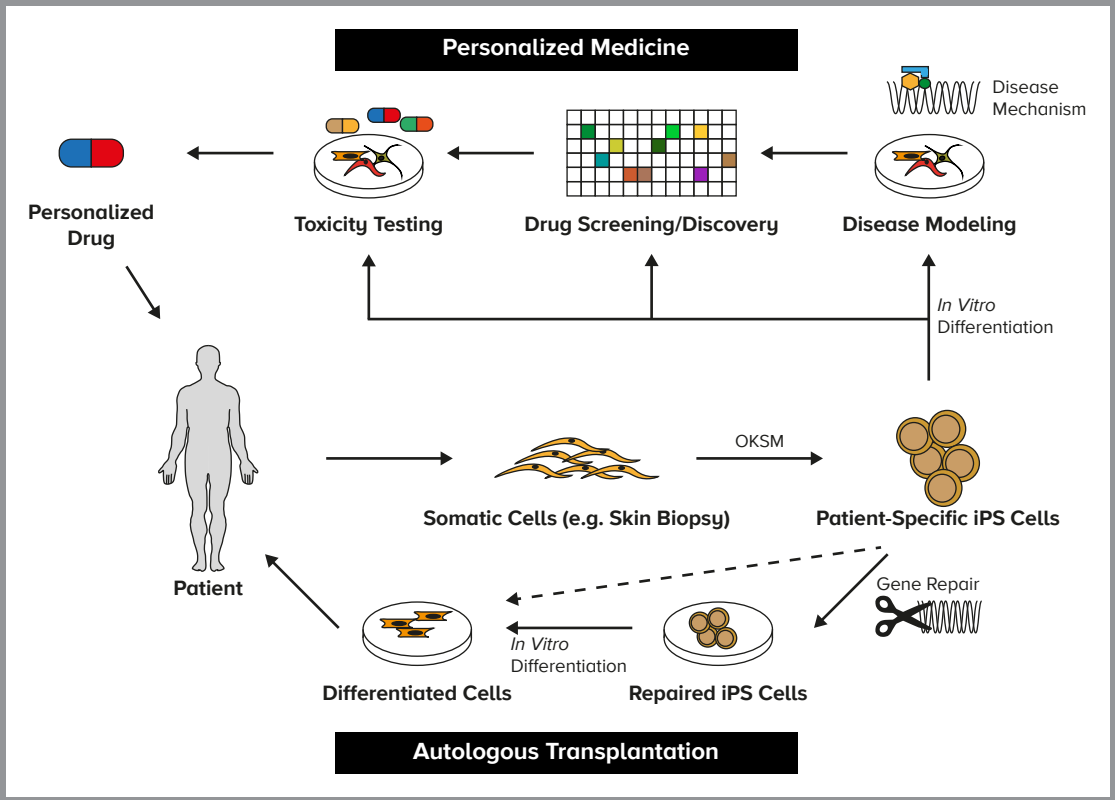
As the result of groundbreaking work of Sir John B. Gurdon, it became clear that the differentiation state of adult somatic cells was not fixed but could be reversed. Gurdon and later others showed that somatic cell nuclear transfer into an enucleated egg

cell or oocyte could lead to a reprogrammed (epi)genome similar to the one in the pluripotent state [45, 46]**. This concept eventually led to generation of the first induced PSCs (iPSCs) in mice and humans** [47–49]**. The ectopic expression of only four transcription factors, namely Oct4, Sox2, Klf4 and c-Myc, was sufficient to reset the differentiation state of fully differentiated somatic cells to a pluripotent state similar to that of ESCs. Since somatic reprogramming with the classical Yamanaka factors is a very inefficient process, many later studies have focused on identifying reprogramming enhancers including genes associated with pluripotency and cell cycle but also epigenetic modifiers** [50]**. Because the initial methods to deliver the four factors involved integrating vectors such as retro- and lentiviruses, recent efforts have focused on methods that do not integrate into the genome including episomal vectors, Sendai viruses and synthetic mRNAs** [50]**. Since research as well as therapeutic applications of human ESCs have been always accompanied by moral and ethical concerns, the discovery of induced pluripotency raised hope for an unencumbered source of pluripotent cells** [51]**. Initially, it was reported that iPSCs harbor an epigenetic state very similar to ESCs** [52]**, but over the years, several studies suggested that iPSCs have distinct gene expression and DNA methylation signatures distinguishing them from ESCs** [53–55]**. The observation that iPSCs retain the expression of genes that can be traced back to the reprogrammed donor cells** [56, 57] **led to the idea of an «epigenetic memory» of the cell of origin** [58]**. Furthermore, it has been speculated that an epigenetic memory may not only bias differentiation towards the fate of the donor cell** [59–66]**, but also improve the functional characteristics of the differentiated cell** [67]**. Although the presence of an epigenetic memory has been observed in murine and human iPSCs, a specific murine characteristic appears to be that the memory is restricted to early passage iPSCs and disappears upon continuous passaging** [62]**. However, it was suggested that the reprogram-**

ming method, prolonged culture but also genetic background could bring iPSCs closer to ESCs [68, 69]**. Indeed, recent studies taking advantage of technological progress and higher numbers of analyzed cell lines have shown that most variation originates from differences in the genetic background, i.e. if the genetic background is similar, then ESCs and iPSCs are almost indistinguishable** [70–72]**.**

Since their discovery, iPSCs have clearly shown that they show promise to revolutionize the field of regenerative and personalized medicine, but also other medical and pharmacological applications [Fig. 2] [28, 73, 74]**. Human ESCs harbor the intrinsic problem of an immune rejection after transplantation since they are not «self», while iPSCs derived from one's own cells do not have this problem. Although there are clinical trials ongoing with human ESCs particularly for diseases related to the eye, brain and kidney, but also in diabetes** [75]**, iPSCs are believed to resolve the majority of these immune issues** [27, 51]**. However, the bottleneck to date is that it remains very expensive to make an individual patient-specific iPSC-based therapy. Besides cell-based therapies, iPSCs generated from cells of patients affected by genetic disease provide a versatile platform to investigate the mechanism of pathologies *in vitro*, a process called disease modelling** [76]**. Since it has been shown that various disease-specific human iPSC lines can be produced and show a phenotype** [77]**, these cell lines can be used to screen for new, if needed patient-specific, drugs and perform toxicity tests of drug candidates** [Fig. 2] [76]**. The manipulation of the genome, called genome editing, opened a huge variety of new opportunities in basic research but also regenerative medicine. It now comprises methods such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and particularly CRISPR/Cas9** [78]**. This progress in combination with iPSCs holds extraordinary potential, because it allows the correction of patient-specific differentiated cells *ex vivo*, which then would eventually be transplanted**

back to the patient, but more immediately can serve as a proper control population in disease modelling [Fig. 2] [79, 80]**.**



[Fig. 2]

Applications of human iPSCs.

Human iPSCs have significant potential in personalized medicine (top) and autologous cell transplantation (bottom). Illustrations were adapted from Bellin *et al.*; Robinton and Daley [28, 76].

GENE EXPRESSION AND EPIGENETICS

GENERAL

Although all the cells in the body share the same genome, there are more than 200 different cell types in the human body. The type and function of a cell is determined by their protein makeup, which is a product of gene expression: each cell specifically regulates its transcriptional activity in a spatiotemporal manner mainly mediated by transcription factors binding to regulatory elements known as promoters and enhancers [81, 82]**. Understanding this type of regulation is crucial for proper control of the differentiation state of stem cells and their differentiated derivatives. In order to allow gene regulation, DNA in eukaryotic cells is organized as chromatin, a condensed complex consisting of DNA, proteins and RNA. The basic elements of chromatin are the nucleosomes: they are composed of the four canonical histone types H2A, H2B, H3, and H4 arranged in repetitive octamers, but there also exist alternative histone variants** [83, 84]**. A further important constituent of nucleosomes is the histone H1, which is not part of the core, but is on top and maintains the higher order of the chromatin. Chromatin can then be remodeled to alter the accessibility of the DNA, which in turn can regulate gene expression without changing the DNA sequence itself, a process called epigenetic regulation. Amongst others, chemical covalent modifications to DNA and/or histones are prominent and well-studied mechanisms of chromatin remodeling.**

Enzymes can chemically modify histones on many different amino acid residues by adding various chemical groups, whereof acetylation and methylation are the best-studied processes [85]**. It has been shown that acetylation and methylation regulate transcription at various regulatory regions including promoters, enhancers and gene bodies** [86, 87]**.**

Whereas acetylation, particularly at lysine 27 of histone H3 (H3K27), has mainly an activating effect on transcription, methylation of lysines and arginines mark either active or repressive chromatin regions [88, 89]**. Mono-, di- and trimethyation at lysine 4 of histone H3 (H3K4) are for instance indicators of active promoters and enhancers, whereas, amongst others, trimethylation at the lysins of H3K27, H3K9, and H3K79 have the opposite effect and mark repressed promoters and enhancers** [90]**. Increasing knowledge about histone modifications combined with advancing genome-wide technologies, such as ChIP-seq, allows new insights into regulatory elements of a wide variety of human tissues and cells** [91]**. It has been shown that the combinatorial pattern of a set of histone modifications is sufficient to identify various functional elements in the genome** [92, 93]**. An interesting phenomenon, and of paramount importance for the plasticity of ESCs, are bivalent domains that are marked by active and repressive histone marks and are often associated with promoters** [94]**. Those poised promoters allow ESCs to regulate gene expression accurately during the complex process of cellular differentiation.**

Another epigenetic modification of paramount importance for transcriptional regulation is DNA methylation, particularly cytosine methylation in the context of CpG dinucleotides [95, 96]**. It has been shown that DNA methylation is a major determinant in fundamental biological processes such as embryogenesis, X-chromosome inactivation, genomic imprinting and tumorigenesis** [97–101]**. The majority of the CpGs in the genome are methylated, however, there are DNA regions up to a length of 2 kb with a high CpG density, called CpG islands (CGIs)** [102]**. Interestingly, those CGIs remain hypomethylated and are present in a majority of promoters** [103, 104]**. Although DNA methylation of promoters was historically associated with repression of gene expression, it has been shown recently that the CpG density also plays a role in their activity** [105]**. For promoters with a high and intermediate density of CpGs this**

paradigm holds true, in contrast, methylated promoters with a low CpG density can be still transcriptionally active and there is no correlation with gene expression. The same negative correlation between DNA methylation and gene expression has been observed for enhancers, which are distal regulatory sequences interacting with the promoters [106, 107]. By contrast, DNA methylation of gene bodies marks active transcription and this is a much less studied relationship [95, 108]. It has been shown that this intragenic DNA methylation could be related to the use of alternative promoters and splicing sites since those intragenic CpGs are often found at intron-exon boundaries [109, 110].

DNA METHYLATION IN DEVELOPMENT AND STEM CELLS

Comparative analyses of DNA methylation landscapes between different tissues have shown that human adult tissues harbor specific DNA methylation patterns that are associated with genes mediating tissue-specific functions [111, 112]. A very common distinguishable feature is tissue-specific hypomethylation [113, 114], which has been found in regions with a low CpG content [115, 116], but also in regulatory regions such as enhancers [117, 118]. Interestingly, intragenic DNA methylation has been shown to be involved in the use of alternative promoters leading to the tissue-specific expression of alternative transcripts [109]. The DNA methylation landscape of human preimplantation embryos is extensively reprogrammed after conception [119–121]. Briefly, the genome is almost completely demethylated after fertilization, with the exception of all imprinted regions that retain their parent-specific DNA methylation, followed by remethylation upon implantation of the blastocyst. However, how tissue-specific DNA methylation patterns observed in adult tissues are established during fetal development is still unclear. Specific sets of hypomethylated CpGs, a tissue-specific feature, have been already observed in several second trimester

human fetal tissues [113]**. This tissue-specific demethylation establishing tissue-specific hypomethylated regions was at least confirmed in the human fetal liver and brain and seems to be the predominant distinguishing feature** [122, 123]**. However, the presence of unmethylated regions (UMRs) that become tissue-specifically *de novo* methylated emphasizes that not only demethylation plays an important role during organ development** [124]**. Furthermore, it has been shown that tissue-specific methylation is often associated with *HOX* and *PAX* genes known to be involved in development** [125]**. Interestingly, only a small portion of hypo- but also hypermethylated tissue-specific differentially methylated regions (DMRs) identified in five different second trimester tissues were found in their adult counterparts suggesting a dynamic DNA methylation landscape during the later stages of organ development** [126]**. Our work in this thesis has expanded the knowledge about the significance of tissue-specific hypomethylation as well as extensive DNA methylation remodeling during first and second trimester development** [127]**.**

Differentiation of ESCs is characterized by compaction of the chromatin that is achieved by epigenetic remodeling [128]. It has been shown that human ESCs have a specific DNA methylation signature and are globally slightly hypermethylated compared to differentiated cells [69, 110, 129]**. The high density CpG promoters of pluripotency-associated genes such as *Oct4* and *Nanog* are hypomethylated in ESCs regulating their own expression, whereas low CpG promoters of tissue-specific genes have been found to be methylated and, therefore, repressed in ESCs** [105, 130]**. The global hypermethylation, i.e. the increase of DNA methylation, in undifferentiated stem cells compared to differentiated cells seems to confirm previous observations of demethylation during *in vivo* development** [110, 122, 123]**. It has been shown that the differentiation of ESCs towards the three embryonic germ layers, particularly towards ectoderm, is indeed accompanied by demethylation** [113, 131]**. However, it has been shown that *in vitro* differentiations**

of ESCs do not always reflect the actual developmental steps: for instance *de novo* DNA methylation, e.g. of bivalent domains or low CpG promoters, occurs to a greater extent and regulates differential expression of tissue-specific genes [123, 132]**. Human ESCs have the hypermethylated state in common with mouse EpiSCs, however, it is noteworthy that naïve mouse ESCs are globally hypomethylated, a feature also observed in human naïve ESCs recently** [42, 133]**. The regulatory functions of DNA methylation are always complemented with other epigenetic processes particularly histone modifications, which together maintain the pluripotent state as well as drive differentiation** [134, 135]**.**

Pluripotent stem cells harbor a more «active» chromatin with an open, accessible confirmation compared to the compact chromatin of differentiated cells [136]**. Therefore, reprogramming of a somatic cell into a pluripotent iPSC is mainly an epigenetic process and involves extensive remodeling of the epigenome** [137]**. Like ESCs, iPSCs are globally slightly DNA hypermethylated suggesting suppression of somatic, tissue-specific genes mediated by DNA methylation** [53, 69, 138]**. Although the majority of stem cell-specific DMRs are hypermethylated, a key step of reprogramming is the demethylation of the promoter regions of pluripotency-associated genes including *Oct4* and *Nanog*** [47]**. Interestingly, it has been shown that the establishment of the pluripotent methylome, in contrast to histone modifications, occurs rather late during the reprogramming process suggesting that DNA methylation and demethylation is the rate-limiting step of somatic reprogramming** [139, 140]**. This may also explain why intermediate, partly reprogrammed cell populations during reprogramming are unstable and lose pluripotency upon removal of the reprogramming factors** [141]**. Underlining the importance of DNA methylation, particularly demethylation of pluripotency-associated promoters, during the reprogramming process, it has been observed that reducing DNA methylation chemically**

or by inhibiting DNA (cytosine-5)-methyltransferase 1 (Dnmt1), which maintains DNA methylation, is beneficial for reprogramming [142]**. In contrast, somatic cells with a knocked out *de novo* methyltransferases (Dnmt3a/b) did not show any reprogramming deficiency** [143]**.**

For their different applications, particularly for therapeutic use in regenerative medicine, it is of paramount importance that the (epi)genetic state of iPSCs and the role it may play in directed differentiation is completely understood. As mentioned above, initially it was shown that iPSCs are epigenetically very similar to ESCs [52]**. However, observations of transcriptional patterns retained from the reprogrammed cells of origin raised questions on the existence of an epigenetic memory** [56, 57]**. Indeed, a somatic memory has been detected in iPSCs, and it was mainly attributed to retained DNA methylation marks from the somatic cells** [60, 62, 64, 144]**. Interestingly, in contrast to mouse iPSCs, human iPSCs retained such a memory only during early passages** [62]**. Furthermore, it has been shown that different reprogramming methods and culture conditions also determined to what extent the epigenetic variability was present** [68, 69]**. However, using a large sample set of different ESCs and iPSCs revealed that the differences actually rather depended on the individual cell line, and not on global differences between these two types of pluripotent cells** [145]**. Due to the contradictory evidence on the presence of an epigenetic memory, it has remained controversial whether an epigenetic memory actually exists - and if yes, if it should be removed from the iPSCs or if it could be beneficial considering the hypothesized differentiation bias. Although still having limitations, recent studies using genetically matched iPSCs and ESCs suggested that the driving force of variation was actually the genetic background and ESCs and iPSCs were indistinguishable** [70–72]**. Illustrating the contradictory evidence, it is noteworthy that the generation of human ESCs via somatic cell nuclear transfer (SCNT) has been reported recently, and they were**

epigenetically closer to ESCs than iPSCs [144]**. Although they share the same advantages with iPSCs in terms of applications, their generation is ethically still delicate and more research is needed.**

AIM AND OUTLINE OF THE THESIS

The pancreas plays a key role in the development of diabetes mellitus, a metabolic disorder characterized by elevated blood glucose levels, which has reached epidemic proportions affecting more than 200 million people worldwide. Type 1 diabetes is caused by an almost complete elimination of pancreatic β-cells due to T-cell-mediated autoimmune destruction [146] **and insulin is required to survive. Type 2 diabetes results from impaired function of the β-cells in combination with insulin resistance in peripheral organs** [147]**. Initially patients with type 2 diabetes can be treated with lifestyle modifications, later combined with insulin secretagogues and sensitizers, both anti-hyperglycemic drugs lowering the glucose levels. Later in the course of the disease, many patients with type 2 depend on insulin replacement via injection or an insulin pump. The only ultimate cure to date for patients that require insulin replacement is transplantation of a vascularized pancreas or of isolated, purified islets** [148]**. However, due to the limited availability of cadaveric donors, alternative sources for β-cell replacement need to be found. Since both forms of diabetes result from deficiency of β-cells, human ESCs as well as iPSCs provide a valuable source to generate insulin-producing cells used for cell-based therapies to restore the β-cell mass** [149]**. It has been shown that spontaneous differentiation of human ESCs leads to the formation of cells able to secrete insulin, although not in response to glucose** [150]**. Since this proof of principle, a wide variety of protocols have been developed to generate insulin-secreting cells from human ESCs and iPSCs, but they all depend on *in vivo* transplantation into mice to undergo the final steps of functional maturation, which is the capability of glucose-stimulated insulin secretion (GSIS)** [151–156]**. A closer look at the human fetal development of the pancreas may explain the**

difficulties to direct the differentiation of stem cells towards a functional β-cell phenotype: although partly matured endocrine cells have been detected in the second trimester of human pregnancy, the final steps towards a completely glucose-responsive β-cell occurs neonatally [157–160]**. Indeed, transcriptional analyses have been shown β-cells derived from human pluripotent stem cells are closer to fetal than adult β-cells** [161]**. Although there have been recent advances towards the generation of mature β-cells *in vitro*, research is still needed in order to generate fully characterized cells that are equivalent to their *in vivo* counterparts, before these cells can be used for clinical applications** [162–164]**. As stated earlier, an interesting aspect of iPSCs is the presence of an epigenetic memory from the cell of origin, which may introduce a differentiation bias as well as increased functionality of the differentiated cells** [59, 61–66]**. As a proof of concept, it has been reported that iPSCs generated from β-cells show indeed increased efficiency when differentiated towards insulin-producing cells** [60]**.**

Our access to rare first and second trimester human fetal material derived from elective abortions determined the outline of this thesis. First, it allowed us to investigate human pancreas organ development with good quality specimens at various time points. Second, since it has been shown that the genetic background may be the driving force of the variability between iPSCs [70–72]**, we could generate isogenic iPSC lines derived from different organs using the same culture conditions and reprogramming method in order to assess the existence of an epigenetic memory. These lines provide a valuable platform to investigate in the future if the iPSC lines derived from pancreatic cells represent a source for more or more functional β-cells compared to the other cell lines.**

Understanding the various aspects of organ development during human fetal development is necessary to draw conclusions regarding the required niche to achieve adequate differentiation towards a particular cell type from stem cells

***in vitro*. It has been shown that signals from the blood vasculature are important for the generation and functionality of β-cells** [165, 166]**. On the other hand, the establishment and role of the lymphatic vasculature during pancreas development is largely unknown. In chapter 2, we investigated lymphangiogenesis and angiogenesis in the human fetal pancreas between 9 and 22 weeks of gestation.**

Since it has been shown that the differentiated derivatives from pluripotent stem cells are still immature and have a fetal phenotype, the field is in need of tools and methods to qualitatively assess stem cell differentiation protocols [161, 167]**. In chapter 3, we have generated the transcriptional profiles of 21 fetal organs and the maternal endometrium at different time points. Using these data combined with an algorithm, we have developed a tool, KeyGenes, which is capable of staging transcriptional data of the cells of interest. Furthermore, we provide a homepage (www.keygenes.nl) with an easy-to-use and freely available web application.**

In order to investigate an epigenetic memory in human iPSCs, the epigenetic landscapes during the fetal development of individual organs need to be determined. In chapter 4, we present an in-depth analysis of the DNA methylation patterns of four fetal organs (pancreas, amnion, muscle, adrenal gland).

Using the transcriptional data generated in chapter 3 combined with genome-wide DNA methylation data of 21 human fetal organs and the maternal endometrium, we provide a roadmap of the gene expression and DNA methylation dynamics between 9 and 22 weeks of gestation in chapter 5. Furthermore, we generated isogenic human iPSC lines and investigated the presence of an epigenetic memory by comparing their DNA methylation profiles with their *in vivo* counterparts.

Chapter 6 summarizes the findings and puts the thesis into perspective in the context of past, current and future research.

REFERENCES

[1] Seisenberger S, Peat JR, Hore TA, Santos F, Dean W, Reik W: **Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers**. *Philos Trans R Soc Lond B Biol Sci* 2013, **368**(1609):20110330.

[2] Clift D, Schuh M: **Restarting life: fertilization and the transition from meiosis to mitosis**. *Nature reviews Molecular cell biology* 2013, **14**(9):549–562.

[3] Courtois A, Schuh M, Ellenberg J, Hiiragi T: **The transition from meiotic to mitotic spindle assembly is gradual during early mammalian development**. *J Cell Biol* 2012, **198**(3):357–370.

[4] Niakan KK, Han J, Pedersen RA, Simon C, Pera RA: **Human pre-implantation embryo development**. *Development* 2012, **139**(5):829–841.

[5] Cockburn K, Rossant J: **Making the blastocyst: lessons from the mouse**. *The Journal of clinical investigation* 2010, **120**(4):995–1003.

[6] Cha J, Sun X, Dey SK: **Mechanisms of implantation: strategies for successful pregnancy**. *Nature medicine* 2012, **18**(12):1754–1767.

[7] Solnica-Krezel L, Sepich DS: **Gastrulation: making and shaping germ layers**. *Annu Rev Cell Dev Biol* 2012, **28**:687–717.

[8] Slack JM: **Developmental biology of the pancreas**. *Development* 1995, **121**(6):1569–1580.

[9] Cleveland MH, Sawyer JM, Afelik S, Jensen J, Leach SD: **Exocrine ontogenies: on the development of pancreatic acinar, ductal and centroacinar cells**. *Semin Cell Dev Biol* 2012, **23**(6):711–719.

[10] Collombat P, Hecksher-Sorensen J, Serup P, Mansouri A: **Specifying pancreatic endocrine cell fates**. *Mechanisms of development* 2006, **123**(7):501–512.

[11] Pan FC, Wright C: **Pancreas organogenesis: from bud to plexus to gland**. *Developmental dynamics : an official publication of the American Association of Anatomists* 2011, **240**(3):530–565.

[12] Jennings RE, Berry AA, Strutt JP, Gerrard DT, Hanley NA: **Human**

pancreas development. *Development* 2015, **142**(18):3126–3137.

[13] Jennings RE, Berry AA, Kirkwood-Wilson R, Roberts NA, Hearn T, Salisbury RJ, Blaylock J, Piper Hanley K, Hanley NA: **Development of the human pancreas from foregut to endocrine commitment**. *Diabetes* 2013, **62**(10):3514–3522.

[14] Piper K, Brickwood S, Turnpenny LW, Cameron IT, Ball SG, Wilson DI, Hanley NA: **Beta cell differentiation during early human pancreas development**. *The Journal of endocrinology* 2004, **181**(1):11–23.

[15] Zhou Q, Law AC, Rajagopal J, Anderson WJ, Gray PA, Melton DA: **A multipotent progenitor domain guides pancreatic organogenesis**. *Developmental cell* 2007, **13**(1):103–114.

[16] Castaing M, Duvillie B, Quemeneur E, Basmaciogullari A, Scharfmann R: **Ex vivo analysis of acinar and endocrine cell development in the human embryonic pancreas**. *Developmental dynamics : an official publication of the American Association of Anatomists* 2005, **234**(2):339–345.

[17] Polak M, Bouchareb-Banaei L, Scharfmann R, Czernichow P: **Early pattern of differentiation in the human pancreas**. *Diabetes* 2000, **49**(2):225–232.

[18] Lyttle BM, Li J, Krishnamurthy M, Fellows F, Wheeler MB, Goodyer CG, Wang R: **Transcription factor expression in the developing human fetal endocrine pancreas**. *Diabetologia* 2008, **51**(7):1169–1180.

[19] Sarkar SA, Kobberup S, Wong R, Lopez AD, Quayum N, Still T, Kutchma A, Jensen JN, Gianani R, Beattie GM *et al*: **Global gene expression profiling and histochemical analysis of the developing human fetal pancreas**. *Diabetologia* 2008, **51**(2):285–297.

[20] Piper Hanley K, Hearn T, Berry A, Carvell MJ, Patch AM, Williams LJ, Sugden SA, Wilson DI, Ellard S, Hanley NA: **In vitro expression of NGN3 identifies RAB3B as the predominant Ras-associated GTP-binding protein 3 family member in human islets**. *The Journal of endocrinology* 2010, **207**(2):151–161.

[21] Amella C, Cappello F, Kahl P, Fritsch H, Lozanoff S, Sergi C: **Spatial and temporal dynamics of innervation during the development of fetal human pancreas**. *Neuroscience* 2008, **154**(4):1477–1487.

[22] Cleaver O, Dor Y: **Vascular instruction of pancreas development**.

Development 2012, **139**(16):2833–2843.

[23] Roost MS, van Iperen L, de Melo Bernardo A, Mummery CL, Carlotti F, de Koning EJ, Chuva de Sousa Lopes SM: **Lymphangiogenesis and angiogenesis during human fetal pancreas development**. *Vascular cell* 2014, **6**:22.

[24] He S, Nakada D, Morrison SJ: **Mechanisms of stem cell self-renewal**. *Annu Rev Cell Dev Biol* 2009, **25**:377–406.

[25] De Los Angeles A, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, Hochedlinger K, Jaenisch R, Lee S, Leitch HG *et al*: **Hallmarks of pluripotency**. *Nature* 2015, **525**(7570):469–478.

[26] Wu J, Izpisua Belmonte JC: **Dynamic Pluripotent Stem Cell States and Their Applications**. *Cell Stem Cell* 2015, **17**(5):509–525.

[27] Trounson A, DeWitt ND: **Pluripotent stem cells progressing to the clinic**. *Nature reviews Molecular cell biology* 2016, **17**(3):194–200.

[28] Robinton DA, Daley GQ: **The promise of induced pluripotent stem cells in research and therapy**. *Nature* 2012, **481**(7381):295–305.

[29] Stevens LC: **Studies on transplantable testicular teratomas of strain 129 mice**. *J Natl Cancer Inst* 1958, **20**(6):1257–1275.

[30] Evans MJ, Kaufman MH: **Establishment in culture of pluripotential cells from mouse embryos**. *Nature* 1981, **292**(5819):154–156.

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

[32] Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM: **Embryonic stem cell lines derived from human blastocysts**. *Science* 1998, **282**(5391):1145–1147.

[33] Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA *et al*: **Derivation of pluripotent epiblast stem cells from mammalian embryos**. *Nature* 2007, **448**(7150):191–195.

[34] Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD: **New cell lines from mouse epiblast share defining features with human embryonic stem cells**. *Nature* 2007, **448**(7150):196–199.

[35] Nichols J, Smith A: **Naive and primed pluripotent states**. *Cell Stem Cell* 2009, **4**(6):487–492.

[36] Hanna J, Cheng AW, Saha K, Kim J, Lengner CJ, Soldner F, Cassady JP, Muffat J, Carey BW, Jaenisch R: **Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(20):9222–9227.

[37] Greber B, Wu G, Bernemann C, Joo JY, Han DW, Ko K, Tapia N, Sabour D, Sterneckert J, Tesar P *et al*: **Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells**. *Cell Stem Cell* 2010, **6**(3):215–226.

[38] Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, Kalma Y, Viukov S, Maza I, Zviran A *et al*: **Derivation of novel human ground state naive pluripotent stem cells**. *Nature* 2013, **504**(7479):282–286.

[39] Ware CB, Nelson AM, Mecham B, Hesson J, Zhou W, Jonlin EC, Jimenez-Caliani AJ, Deng X, Cavanaugh C, Cook S *et al*: **Derivation of naive human embryonic stem cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**(12):4484–4489.

[40] Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, Oxley D, Santos F, Clarke J, Mansfield W *et al*: **Resetting transcription factor control circuitry toward ground-state pluripotency in human**. *Cell* 2014, **158**(6):1254–1269.

[41] Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, Fan ZP, Maetzel D, Ganz K, Shi L *et al*: **Systematic identification of culture conditions for induction and maintenance of naive human pluripotency**. *Cell Stem Cell* 2014, **15**(4):471–487.

[42] Duggal G, Warrier S, Ghimire S, Broekaert D, Van der Jeught M, Lierman S, Deroo T, Peelman L, Van Soom A, Cornelissen R *et al*: **Alternative Routes to Induce Naive Pluripotency in Human Embryonic Stem Cells**. *Stem Cells* 2015, **33**(9):2686–2698.

[43] Yang Y, Zhang X, Yi L, Hou Z, Chen J, Kou X, Zhao Y, Wang H, Sun XF, Jiang C *et al*: **Naive Induced Pluripotent Stem Cells Generated From beta-Thalassemia Fibroblasts Allow Efficient Gene Cor-**

rection With CRISPR/Cas9. *Stem cells translational medicine* 2016, **5**(1):8–19.

[44] Dodsworth BT, Flynn R, Cowley SA: **The Current State of Naive Human Pluripotency**. *Stem Cells* 2015, **33**(11):3181–3186.

[45] Gurdon JB: **The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles**. *J Embryol Exp Morphol* 1962, **10**:622–640.

[46] Tachibana M, Amato P, Sparman M, Gutierrez NM, Tippner-Hedges R, Ma H, Kang E, Fulati A, Lee HS, Sritanaudomchai H *et al*: **Human embryonic stem cells derived by somatic cell nuclear transfer**. *Cell* 2013, **153**(6):1228–1238.

[47] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors**. *Cell* 2007, **131**(5):861–872.

[48] Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors**. *Cell* 2006, **126**(4):663–676.

[49] Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R *et al*: **Induced pluripotent stem cell lines derived from human somatic cells**. *Science* 2007, **318**(5858):1917–1920.

[50] Takahashi K, Yamanaka S: **A decade of transcription factor-mediated reprogramming to pluripotency**. *Nature reviews Molecular cell biology* 2016, **17**(3):183–193.

[51] Power C, Rasko JE: **Will cell reprogramming resolve the embryonic stem cell controversy? A narrative review**. *Ann Intern Med* 2011, **155**(2):114–121.

[52] Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R *et al*: **Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution**. *Cell Stem Cell* 2007, **1**(1):55–70.

[53] Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S *et al*: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and**

fibroblasts. *Nat Genet* 2009, **41**(12):1350–1353.

[54] Pick M, Stelzer Y, Bar-Nur O, Mayshar Y, Eden A, Benvenisty N: **Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells**. *Stem Cells* 2009, **27**(11):2686–2690.

[55] Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiuwu O, Richter L, Zhang J *et al*: **Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures**. *Cell Stem Cell* 2009, **5**(1):111–123.

[56] Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR: **Transcriptional signature and memory retention of human-induced pluripotent stem cells**. *PLoS One* 2009, **4**(9):e7076.

[57] Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC: **Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells**. *PLoS One* 2010, **5**(2):e8975.

[58] Firas J, Liu X, Polo JM: **Epigenetic memory in somatic cell nuclear transfer and induced pluripotency: evidence and implications**. *Differentiation* 2014, **88**(1):29–32.

[59] Lee SB, Seo D, Choi D, Park KY, Holczbauer A, Marquardt JU, Conner EA, Factor VM, Thorgeirsson SS: **Contribution of hepatic lineage stage-specific donor memory to the differential potential of induced mouse pluripotent stem cells**. *Stem Cells* 2012, **30**(5):997–1007.

[60] Bar-Nur O, Russ HA, Efrat S, Benvenisty N: **Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells**. *Cell Stem Cell* 2011, **9**(1):17–23.

[61] Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T *et al*: **Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells**. *Nat Biotechnol* 2010, **28**(8):848–855.

[62] Kim K, Zhao R, Doi A, Ng K, Unternaehrer J, Cahan P, Huo H, Loh YH, Aryee MJ, Lensch MW *et al*: **Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent**

stem cells. *Nat Biotechnol* 2011, **29**(12):1117–1119.

[63] Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S *et al*: **Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells**. *Nature* 2011, **471**(7336):68–73.

[64] Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ *et al*: **Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells**. *Nat Cell Biol* 2011, **13**(5):541–549.

[65] Quattrocelli M, Swinnen M, Giacomazzi G, Camps J, Barthelemy I, Ceccarelli G, Caluwe E, Grosemans H, Thorrez L, Pelizzo G *et al*: **Mesodermal iPSC-derived progenitor cells functionally regenerate cardiac and skeletal muscle**. *The Journal of clinical investigation* 2015, **125**(12):4463–4482.

[66] Nukaya D, Minami K, Hoshikawa R, Yokoi N, Seino S: **Preferential gene expression and epigenetic memory of induced pluripotent stem cells derived from mouse pancreas**. *Genes to cells : devoted to molecular & cellular mechanisms* 2015, **20**(5):367–381.

[67] Rizzi R, Di Pasquale E, Portararo P, Papait R, Cattaneo P, Latronico MV, Altomare C, Sala L, Zaza A, Hirsch E *et al*: **Post-natal cardiomyocytes can generate iPS cells with an enhanced capacity toward cardiomyogenic re-differentation**. *Cell death and differentiation* 2012, **19**(7):1162–1174.

[68] Chin MH, Pellegrini M, Plath K, Lowry WE: **Molecular analyses of human induced pluripotent stem cells and embryonic stem cells**. *Cell Stem Cell* 2010, **7**(2):263–269.

[69] Nishino K, Toyoda M, Yamazaki-Inoue M, Fukawatase Y, Chikazawa E, Sakaguchi H, Akutsu H, Umezawa A: **DNA methylation dynamics in human induced pluripotent stem cells over time**. *PLoS Genet* 2011, **7**(5):e1002085.

[70] Choi J, Lee S, Mallard W, Clement K, Tagliazucchi GM, Lim H, Choi IY, Ferrari F, Tsankov AM, Pop R *et al*: **A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs**. *Nat Biotechnol* 2015.

[71] Tiemann U, Wu G, Marthaler AG, Scholer HR, Tapia N: **Epigenetic**

Aberrations Are Not Specific to Transcription Factor-Mediated Reprogramming. *Stem Cell Reports* 2016, **6**(1):35–43.

[72] Rouhani F, Kumasaka N, de Brito MC, Bradley A, Vallier L, Gaffney D: **Genetic background drives transcriptional variation in human induced pluripotent stem cells**. *PLoS Genet* 2014, **10**(6):e1004432.

[73] Csobonyeiova M, Polak S, Koller J, Danisovic L: **Induced pluripotent stem cells and their implication for regenerative medicine**. *Cell and tissue banking* 2015, **16**(2):171–180.

[74] Inoue H, Nagata N, Kurokawa H, Yamanaka S: **iPS cells: a game changer for future medicine**. *EMBO J* 2014, **33**(5):409–417.

[75] Trounson A, McDonald C: **Stem Cell Therapies in Clinical Trials: Progress and Challenges**. *Cell Stem Cell* 2015, **17**(1):11–22.

[76] Bellin M, Marchetto MC, Gage FH, Mummery CL: **Induced pluripotent stem cells: the new patient?** *Nature reviews Molecular cell biology* 2012, **13**(11):713–726.

[77] Park IH, Arora N, Huo H, Maherali N, Ahfeldt T, Shimamura A, Lensch MW, Cowan C, Hochedlinger K, Daley GQ: **Disease-specific induced pluripotent stem cells**. *Cell* 2008, **134**(5):877–886.

[78] Maggio I, Goncalves MA: **Genome editing at the crossroads of delivery, specificity, and fidelity**. *Trends in biotechnology* 2015, **33**(5):280–291.

[79] Hockemeyer D, Jaenisch R: **Induced Pluripotent Stem Cells Meet Genome Editing**. *Cell Stem Cell* 2016, **18**(5):573–586.

[80] Deleidi M, Yu C: **Genome editing in pluripotent stem cells: research and therapeutic applications**. *Biochemical and biophysical research communications* 2016, **473**(3):665–674.

[81] Splinter E, de Laat W: **The complex transcription regulatory landscape of our genome: control in three dimensions**. *EMBO J* 2011, **30**(21):4345–4355.

[82] Danino YM, Even D, Ideses D, Juven-Gershon T: **The core promoter: At the heart of gene expression**. *Biochimica et biophysica acta* 2015, **1849**(8):1116–1131.

[83] Cutter AR, Hayes JJ: **A brief review of nucleosome structure**. *FEBS letters* 2015, **589**(20 Pt A):2914–2922.

[84] Kamakaka RT, Biggins S: **Histone variants: deviants?** *Genes Dev*

2005, **19**(3):295–310.

[85] Kouzarides T: **Chromatin modifications and their function**. *Cell* 2007, **128**(4):693–705.

[86] Zhou VW, Goren A, Bernstein BE: **Charting histone modifications and the functional organization of mammalian genomes**. *Nat Rev Genet* 2011, **12**(1):7–18.

[87] Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al*: **Mapping and analysis of chromatin state dynamics in nine human cell types**. *Nature* 2011, **473**(7345):43–49.

[88] Dong X, Greven MC, Kundaje A, Djebali S, Brown JB, Cheng C, Gingeras TR, Gerstein M, Guigo R, Birney E *et al*: **Modeling gene expression using chromatin features in various cellular contexts**. *Genome Biol* 2012, **13**(9):R53.

[89] Eberharter A, Becker PB: **Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics**. *EMBO reports* 2002, **3**(3):224–229.

[90] Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome**. *Cell* 2007, **129**(4):823–837.

[91] The ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**(7414):57–74.

[92] Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ *et al*: **Combinatorial patterns of histone acetylations and methylations in the human genome**. *Nat Genet* 2008, **40**(7):897–903.

[93] Ucar D, Hu Q, Tan K: **Combinatorial chromatin modification patterns in the human genome revealed by subspace clustering**. *Nucleic acids research* 2011, **39**(10):4063–4075.

[94] Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K *et al*: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells**. *Cell* 2006, **125**(2):315–326.

[95] Jones PA: **Functions of DNA methylation: islands, start sites, gene bodies and beyond**. *Nat Rev Genet* 2012, **13**(7):484–492.

[96] Bird A: **DNA methylation patterns and epigenetic memory**. *Genes Dev* 2002, **16**(1):6–21.

[97] Fedoriw A, Mugford J, Magnuson T: **Genomic imprinting and epigenetic control of development**. *Cold Spring Harbor perspectives in biology* 2012, **4**(7):a008136.

[98] Feinberg AP, Vogelstein B: **Hypomethylation distinguishes genes of some human cancers from their normal counterparts**. *Nature* 1983, **301**(5895):89–92.

[99] Okano M, Bell DW, Haber DA, Li E: **DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development**. *Cell* 1999, **99**(3):247–257.

[100] Reik W: **Stability and flexibility of epigenetic gene regulation in mammalian development**. *Nature* 2007, **447**(7143):425–432.

[101] Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ: **Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation**. *Hum Mol Genet* 2015, **24**(6):1528–1539.

[102] Illingworth RS, Bird AP: **CpG islands--'a rough guide'**. *FEBS letters* 2009, **583**(11):1713–1720.

[103] Deaton AM, Bird A: **CpG islands and the regulation of transcription**. *Genes Dev* 2011, **25**(10):1010–1022.

[104] Messerschmidt DM, Knowles BB, Solter D: **DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos**. *Genes Dev* 2014, **28**(8):812–828.

[105] Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB *et al*: **Genome-scale DNA methylation maps of pluripotent and differentiated cells**. *Nature* 2008, **454**(7205):766–770.

[106] Sandovici I, Smith NH, Nitert MD, Ackers-Johnson M, Uribe-Lewis S, Ito Y, Jones RH, Marquez VE, Cairns W, Tadayyon M *et al*: **Maternal diet and aging alter the epigenetic control of a promoter-enhancer interaction at the Hnf4a gene in rat pancreatic islets**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(13):5449–5454.

[107] Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A,

van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D *et al*: **DNA-binding factors shape the mouse methylome at distal regulatory regions**. *Nature* 2011, **480**(7378):490–495.

[108] Lou S, Lee HM, Qin H, Li JW, Gao Z, Liu X, Chan LL, Lam V, So WY, Wang Y *et al*: **Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation**. *Genome Biol* 2014, **15**(7):408.

[109] Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y *et al*: **Conserved role of intragenic DNA methylation in regulating alternative promoters**. *Nature* 2010, **466**(7303):253–257.

[110] Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J *et al*: **Dynamic changes in the human methylome during differentiation**. *Genome Res* 2010, **20**(3):320–331.

[111] Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, Yang AS: **Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns**. *Hum Mol Genet* 2009, **18**(24):4808–4817.

[112] Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M *et al*: **An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs)**. *Genome Res* 2008, **18**(9):1518–1529.

[113] Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Muller FJ, Wang YC, Boscolo FS *et al*: **Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives**. *Cell Stem Cell* 2012, **10**(5):620–634.

[114] Lokk K, Modhukur V, Rajashekar B, Martens K, Magi R, Kolde R, Koltsina M, Nilsson TK, Vilo J, Salumets A *et al*: **DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns**. *Genome Biol* 2014, **15**(4):r54.

[115] Nagae G, Isagawa T, Shiraki N, Fujita T, Yamamoto S, Tsutsumi S,

Nonaka A, Yoshiba S, Matsusaka K, Midorikawa Y *et al*: **Tissue-specific demethylation in CpG-poor promoters during cellular differentiation**. *Hum Mol Genet* 2011, **20**(14):2710–2721.

[116] Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, Suchiman HE, Lameijer EW, Putter H, van den Akker EB *et al*: **Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array**. *Epigenetics Chromatin* 2013, **6**(1):26.

[117] Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL *et al*: **DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape**. *Nat Genet* 2013, **45**(7):836–841.

[118] Yan L, Guo H, Hu B, Li R, Yong J, Zhao Y, Zhi X, Fan X, Guo F, Wang X *et al*: **Epigenomic Landscape of Human Fetal Brain, Heart, and Liver**. *J Biol Chem* 2016, **291**(9):4386–4398.

[119] Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J *et al*: **The DNA methylation landscape of human early embryos**. *Nature* 2014, **511**(7511):606–610.

[120] Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, Eggan K, Meissner A: **DNA methylation dynamics of the human preimplantation embryo**. *Nature* 2014, **511**(7511):611–615.

[121] Okae H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, Kikuchi H, Yoshida H, Tanaka A, Suyama M *et al*: **Genome-wide analysis of DNA methylation dynamics during early human development**. *PLoS Genet* 2014, **10**(12):e1004868.

[122] Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CC, O'Donovan MC, Bray NJ, Mill J: **Methylomic trajectories across human fetal brain development**. *Genome Res* 2015, **25**(3):338–352.

[123] Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E *et al*: **Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver**. *Genome Res* 2009, **19**(6):1044–1056.

[124] Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, Simon I, Yakhini Z, Cedar H: **Developmental program-**

ming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* 2009, **16**(5):564–571.

[125] Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J *et al*: **A novel CpG island set identifies tissue-specific methylation at developmental gene loci**. *PLoS biology* 2008, **6**(1):e22.

[126] Yuen RK, Neumann SM, Fok AK, Penaherrera MS, McFadden DE, Robinson WP, Kobor MS: **Extensive epigenetic reprogramming in human somatic tissues between fetus and adult**. *Epigenetics Chromatin* 2011, **4**:7.

[127] Slieker RC, Roost MS, van Iperen L, Suchiman HE, Tobi EW, Carlotti F, de Koning EJ, Slagboom PE, Heijmans BT, Chuva de Sousa Lopes SM: **DNA Methylation Landscapes of Human Fetal Development**. *PLoS Genet* 2015, **11**(10):e1005583.

[128] Zhu J, Adli M, Zou JY, Verstappen G, Coyne M, Zhang X, Durham T, Miri M, Deshpande V, De Jager PL *et al*: **Genome-wide chromatin state transitions associated with developmental and environmental cues**. *Cell* 2013, **152**(3):642–654.

[129] Bibikova M, Chudin E, Wu B, Zhou L, Garcia EW, Liu Y, Shin S, Plaia TW, Auerbach JM, Arking DE *et al*: **Human embryonic stem cells have a unique epigenetic signature**. *Genome Res* 2006, **16**(9):1075–1083.

[130] Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G: **Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation**. *Cell Stem Cell* 2008, **2**(2):160–169.

[131] Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R *et al*: **Transcriptional and epigenetic dynamics during specification of human embryonic stem cells**. *Cell* 2013, **153**(5):1149–1163.

[132] Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D *et al*: **Epigenomic analysis of multilineage differentiation of human embryonic stem cells**. *Cell* 2013, **153**(5):1134–1148.

[133] Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, Mansfield W, Nashun B, Knezovich JG, Smith A *et al*: **Naive pluripotency is associated with global DNA hypomethylation**. *Nat Struct Mol Biol* 2013, **20**(3):311–316.

[134] Han JW, Yoon YS: **Epigenetic landscape of pluripotent stem cells**. *Antioxidants & redox signaling* 2012, **17**(2):205–223.

[135] Boland MJ, Nazor KL, Loring JF: **Epigenetic regulation of pluripotency and differentiation**. *Circ Res* 2014, **115**(2):311–324.

[136] Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T: **Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells**. *Developmental cell* 2006, **10**(1):105–116.

[137] Nashun B, Hill PW, Hajkova P: **Reprogramming of cell fate: epigenetic memory and the erasure of memories past**. *EMBO J* 2015, **34**(10):1296–1308.

[138] Nishino K, Umezawa A: **DNA methylation dynamics in human induced pluripotent stem cells**. *Human cell* 2016.

[139] Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J *et al*: **A molecular roadmap of reprogramming somatic cells into iPS cells**. *Cell* 2012, **151**(7):1617–1632.

[140] Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE, Meissner A: **Reprogramming factor expression initiates widespread targeted chromatin remodeling**. *Cell Stem Cell* 2011, **8**(1):96–105.

[141] Papp B, Plath K: **Epigenetics of reprogramming to induced pluripotency**. *Cell* 2013, **152**(6):1324–1343.

[142] Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A: **Dissecting direct reprogramming through integrative genomic analysis**. *Nature* 2008, **454**(7200):49–55.

[143] Pawlak M, Jaenisch R: **De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state**. *Genes Dev* 2011, **25**(10):1035–1040.

[144] Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD, Hariharan M, Nery JR, Castanon R, Sabatini K *et al*: **Abnormalities in human pluripotent cells due to reprogramming mechanisms**. *Nature* 2014,

511(7508):177–183.

[145] Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH *et al*: **Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines**. *Cell* 2011, **144**(3):439–452.

[146] Bluestone JA, Herold K, Eisenbarth G: **Genetics, pathogenesis and clinical interventions in type 1 diabetes**. *Nature* 2010, **464**(7293):1293–1300.

[147] Leahy JL: **Pathogenesis of type 2 diabetes mellitus**. *Archives of medical research* 2005, **36**(3):197–209.

[148] Halban PA, German MS, Kahn SE, Weir GC: **Current status of islet cell replacement and regeneration therapy**. *The Journal of clinical endocrinology and metabolism* 2010, **95**(3):1034–1043.

[149] Weir GC, Cavelti-Weder C, Bonner-Weir S: **Stem cell approaches for diabetes: towards beta cell replacement**. *Genome medicine* 2011, **3**(9):61.

[150] Assady S, Maor G, Amit M, Itskovitz-Eldor J, Skorecki KL, Tzukerman M: **Insulin production by human embryonic stem cells**. *Diabetes* 2001, **50**(8):1691–1697.

[151] D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, Moorman MA, Kroon E, Carpenter MK, Baetge EE: **Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells**. *Nat Biotechnol* 2006, **24**(11):1392–1401.

[152] Kroon E, Martinson LA, Kadoya K, Bang AG, Kelly OG, Eliazer S, Young H, Richardson M, Smart NG, Cunningham J *et al*: **Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo**. *Nat Biotechnol* 2008, **26**(4):443–452.

[153] Wei R, Yang J, Hou W, Liu G, Gao M, Zhang L, Wang H, Mao G, Gao H, Chen G *et al*: **Insulin-producing cells derived from human embryonic stem cells: comparison of definitive endoderm- and nestin-positive progenitor-based differentiation strategies**. *PLoS One* 2013, **8**(8):e72513.

[154] Wei R, Yang J, Liu GQ, Gao MJ, Hou WF, Zhang L, Gao HW, Liu Y, Chen GA, Hong TP: **Dynamic expression of microRNAs during the**

differentiation of human embryonic stem cells into insulin-producing cells. *Gene* 2013, **518**(2):246–255.

[155] Rezania A, Bruin JE, Riedel MJ, Mojibian M, Asadi A, Xu J, Gauvin R, Narayan K, Karanu F, O'Neil JJ *et al*: **Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice**. *Diabetes* 2012, **61**(8):2016–2029.

[156] Nostro MC, Sarangi F, Ogawa S, Holtzinger A, Corneo B, Li X, Micallef SJ, Park IH, Basford C, Wheeler MB *et al*: **Stage-specific signaling through TGFbeta family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells**. *Development* 2011, **138**(5):861–871.

[157] Ashworth MA, Leach FN, Milner RD: **Development of insulin secretion in the human fetus**. *Archives of disease in childhood* 1973, **48**(2):151–152.

[158] Milner RD, Ashworth MA, Barson AJ: **Insulin release from human foetal pancreas in response to glucose, leucine and arginine**. *The Journal of endocrinology* 1972, **52**(3):497–505.

[159] Hoffman L, Mandel TE, Carter WM, Koulmanda M, Martin FI: **Insulin secretion by fetal human pancreas in organ culture**. *Diabetologia* 1982, **23**(5):426–430.

[160] Dhawan S, Tschen SI, Zeng C, Guo T, Hebrok M, Matveyenko A, Bhushan A: **DNA methylation directs functional maturation of pancreatic beta cells**. *The Journal of clinical investigation* 2015, **125**(7):2851–2860.

[161] Hrvatin S, O'Donnell CW, Deng F, Millman JR, Pagliuca FW, Dilorio P, Rezania A, Gifford DK, Melton DA: **Differentiated human stem cells resemble fetal, not adult, beta cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**(8):3038–3043.

[162] Russ HA, Parent AV, Ringler JJ, Hennings TG, Nair GG, Shveygert M, Guo T, Puri S, Haataja L, Cirulli V *et al*: **Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro**. *EMBO J* 2015, **34**(13):1759–1772.

[163] Pagliuca FW, Millman JR, Gurtler M, Segel M, Van Dervort A, Ryu JH,

Peterson QP, Greiner D, Melton DA: **Generation of functional human pancreatic beta cells in vitro**. *Cell* 2014, **159**(2):428–439.

[164] Rezania A, Bruin JE, Arora P, Rubin A, Batushansky I, Asadi A, O'Dwyer S, Quiskamp N, Mojibian M, Albrecht T *et al*: **Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells**. *Nat Biotechnol* 2014, **32**(11):1121–1133.

[165] Nikolova G, Jabs N, Konstantinova I, Domogatskaya A, Tryggvason K, Sorokin L, Fassler R, Gu G, Gerber HP, Ferrara N *et al*: **The vascular basement membrane: a niche for insulin gene expression and Beta cell proliferation**. *Developmental cell* 2006, **10**(3):397–405.

[166] Lammert E, Cleaver O, Melton D: **Induction of pancreatic differentiation by signals from blood vessels**. *Science* 2001, **294**(5542):564–567.

[167] Patterson M, Chan DN, Ha I, Case D, Cui Y, Van Handel B, Mikkola HK, Lowry WE: **Defining the nature of human pluripotent stem cell progeny**. *Cell Res* 2012, **22**(1):178–193.

CHAPTER 2

LYMPHANGIOGENESIS AND ANGIOGENESIS DURING HUMAN FETAL PANCREAS DEVELOPMENT

Matthias S. Roost[1], Liesbeth van Iperen[1], Ana de Melo Bernardo[1], Christine L. Mummery[1], Françoise Carlotti[2], Eelco J.P. de Koning[2,3], Susana M. Chuva de Sousa Lopes[1,4]

[1] Department of Anatomy and Embryology,
[2] Department of Nephrology, Leiden University Medical Center, Leiden, the Netherlands
[3] Hubrecht Institute for Developmental Biology and Stem Cell Research, University Medical Center, Utrecht, the Netherlands
[4] Department for Reproductive Medicine, Ghent University Hospital, Ghent, Belgium

Vascular Cell 2014, **6**:22

ABSTRACT

BACKGROUND

The complex endocrine and exocrine functionality of the human pancreas depends on an efficient fluid transport through the blood and the lymphatic vascular systems. The lymphatic vasculature has key roles in the physiology of the pancreas and in regulating the immune response, both important for developing successful transplantation and cell-replacement therapies to treat diabetes. However, little is known about how the lymphatic and blood systems develop in humans. Here, we investigated the establishment of these two vascular systems in human pancreas organogenesis in order to understand neovascularization in the context of emerging regenerative therapies.

METHODS

We examined angiogenesis and lymphangiogenesis during human pancreas development between 9 and 22 weeks of gestation (W9–W22) by immunohistochemistry.

RESULTS

As early as W9, the peri-pancreatic mesenchyme was populated by CD31-expressing blood vessels as well as LYVE1- and PDPN-expressing lymphatic vessels. The appearance of smooth muscle cell-coated blood vessels in the intra-pancreatic mesenchyme occurred only several weeks later and from W14.5 onwards the islets of Langerhans also became heavily irrigated by blood vessels. In contrast to blood vessels, LYVE1- and PDPN-expressing lymphatic vessels were restricted to the peri-pancreatic mesenchyme until later in development (W14.5–W17), and some of these invading lymphatic vessels contained smooth muscle cells at W17. Interestingly, between W11–W22, most large caliber lymphatic vessels were lined with a characteristic,

discontinuous, collagen type IV-rich basement membrane. Whilst lymphatic vessels did not directly intrude the islets of Langerhans, three-dimensional reconstruction revealed that they were present in the vicinity of islets of Langerhans between W17–W22.

CONCLUSION

Our data suggest that the blood and lymphatic machinery in the human pancreas is in place to support endocrine function from W17–W22 onwards. Our study provides the first systematic assessment of the progression of lymphangiogenesis during human pancreatic development.

BACKGROUND

Type 1 diabetes is an autoimmune disease that progressively destroys insulin-producing β-cells [1]**. Given the shortage of organ donors, pluripotent stem cells, such as human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs), but also adult stem cells from the pancreas may provide a valuable source of insulin-producing β-cells for cell-replacement therapies as well as for studying mechanisms underlying β-cell pathologies** [2–4]**. To date, protocols to differentiate β-cells from pluripotent stem cells *in vitro* usually generate immature endocrine cells that constitutively secrete insulin, instead of responding to exogenous glucose levels** [5–7]**. The maturation of these cells into fully functional β-cells is only augmented after transplantation into (immunocompromised) mice** [8–10] **and likely depends upon a favorable microenvironment for cell maturation and function.**

There are only a few studies addressing human pancreas development due to the restricted availability of human embryos [11–20]**. Human pancreas development starts between 26 and 35 days post conception with the emergence of dorsal and ventral buds from the foregut epithelium. At 6 weeks of gestation (equivalent to 4 weeks post conception) the two buds fuse and become a single organ formed by stratified epithelium embedded in mesenchyme. The stratified epithelium will give rise to both the exocrine and endocrine compartments of the definitive pancreas** [21]**.**

One important physiological regulator of development and normal function of the endocrine cells of the pancreas is the microcirculation through specialized sinusoidal capillaries that irrigate the islets of Langerhans [22–24]**. The endothelial cells of these capillaries are highly fenestrated to facilitate the exchange of signals. The dense network ensures**

that each endocrine cell (glucagon-producing α-cell, insulin-producing β-cell, somatostatin-producing d-cell, ghrelin-producing e-cell and pancreatic polypeptide-producing PP-cell) is in close proximity to the circulation [25]**. It makes up a considerable part of the islets and it is responsible for critical communication via blood signals between the endocrine and exocrine pancreas and also between the different cell types that populate the islets. After transplantation of islets to the pancreas, angiogenesis is key to restoring proper function** [8, 25–28]**.**

Another important physiological aspect that regulates the normal function of the pancreas is the presence of a lymphatic system [29]**. The main function of the lymphatic system in the pancreas is to drain excess fluid that can contain both leaking endocrine hormones and exocrine-produced digestive enzymes from the interstitial space** [29, 30]**. Another important function of the lymphatic system, particularly in the context of transplantation and cell-replacement therapies, is the ability to facilitate the immediate transport of immune cells and soluble antigens from the peripheral tissues to the regional lymph nodes for an appropriate immune response** [31]**. However, understanding the lymphatic vascular network remained poor due to the lack of specific markers capable of distinguishing lymphatic vessels. Two cell surface proteins, namely lymphatic vessel endothelial hyaluronan receptor 1 (LYVE1) and podoplanin (PDPN), have stimulated lymphatic vasculature research as they are expressed specifically by lymphatic endothelial cells, and not blood endothelial cells** [32–34]**.**

Together, the lymphatic and blood vascular system control pancreas homeostasis, including the transport of signals, gases, nutrients, hormones, and circulating cells. Both the blood and the lymphatic, vascular systems are highly branched tubular networks, in which vessels are formed by endothelial cells, but there are fundamental differences in terms of function, morphology, and composition [32, 35, 36]**. In contrast to the circular blood vascular network, the**

lymphatic system is blind-ended, consisting of both lymphatic capillaries, pre-collecting and larger collecting lymphatic vessels that are connected to lymph nodes. A profound understanding of the development of the blood and lymphatic vascular system during pancreas organogenesis and the interaction between these systems and human endocrine cells can provide important information on cell maturation and function both *in vitro* and *in vivo*.

In the human pancreas, the presence of CD34-positive blood endothelial cells has been observed from 8 weeks of gestation [13]**; PDPN-positive lymphatic vessels have been reported in the pancreas of a single fetus of 18 weeks of gestation** [37] **and mentioned briefly in a study of the anatomy of the mesocolon transversum at 13 and 16 weeks of gestation** [38]**. However, the timing of colonization by lymphatic vessels and its relation to the formation of blood vessels in the pancreas is not known. We therefore studied the temporal and spatial progression of lymphangiogenesis and compared this with angiogenesis during human pancreas development between 9 and 22 weeks of gestation.**

METHODS

FETAL PANCREAS COLLECTION

Fourteen human fetal pancreatic specimens between 9 and 22 weeks of gestation (W9–W12, n = 5; W14–W22, n = 9) were collected from elective abortion tissue obtained by vacuum aspiration. «Weeks of gestation» used in this study is based on the last menstrual period (LMP), to convert to «weeks post conception» one need to subtract two weeks. This study was approved by the Medical Ethics Committee of the Leiden University Medical Center (protocol 08.087). Informed consent was obtained on the basis of the Declaration of Helsinki by the World Medical Association (WMA). All pancreata were fixed in 4% (w/v) paraformaldehyde (MERCK, Darmstadt, Germany) in PBS overnight at 4°C. Fixation was followed by dehydration in ethanol, xylene and paraffin embedding using standard procedures. Embedding was performed using a Shandon Excelsior tissue processor (Thermo Scientific, Altrincham, UK).

HISTOLOGY AND IMMUNOHISTOCHEMISTRY

Paraffin-embedded tissues were sectioned (5 µm) using a RM2255 microtome (Leica Microsystems GmbH, Wetzlar, Germany) and mounted on StarFrost slides (Waldemar Knittel, Braunschweig, Germany). The sections were deparaffinized and rehydrated by standard procedures, namely in xylene and followed by a decreasing series of ethanol ending by rinsing in distilled water. To assess the morphology of the pancreatic sections, a Haematoxylin (MERCK, Darmstadt, Germany) and Eosin (MERCK, Darmstadt, Germany) staining was performed by standard procedures.

For immunohistochemistry, three methods were used for antigen retrieval depending on the primary antibodies: (1) 12 minutes at 97°C in 0.01M sodium citrate buffer (pH 6.0)

followed by cooling down; 2) 12 minutes at 97°C in Tris/EDTA buffer (pH 9.0) followed by cooling down; 3) 5–10 minutes 20 µg/ml proteinase K (Promega, Madison, USA) in TE-CaCl2 buffer (pH8.0) at room temperature (RT). After antigen retrieval, the sections were blocked with 1% bovine serum albumin, fraction V (BSA, Sigma-Aldrich, St. Louis, USA) in phosphate buffered-saline (PBS) with 0.05% Tween-20 (Promega, Madison, USA) for 1 hour at RT, and incubated with the primary antibodies diluted in the blocking solution overnight at 4°C in a humidified chamber. The primary antibodies used in this study were: Rabbit anti-alpha smooth muscle actin (1:100, ab5694, Abcam, Cambridge, UK), mouse anti-amylase (1:100, sc46657, Santa Cruz Biotechnologies, Dallas, USA), mouse anti-CD31 (1:100, M0823, Dako, Glostrup, Denmark), mouse anti-CD68 (1:1000, M0814, Dako), rabbit anti-CK19 (1:250, ab52625, Abcam, Cambridge, UK), mouse anti-CK19 (ready-to-use, M0888, Dako), rabbit anti-collagen type IV (1:50, AB748, Millipore, Bedford, USA), goat anti-endoglin (1:100, BAF1097, R&D Systems, Minneapolis, USA), rabbit anti-glucagon (1:200, VP-G806, Vector Laboratories Ltd., Peterborough, UK), rabbit anti-insulin (1:100, sc-9168, Santa Cruz Biotechnologies, Dallas, USA), rabbit anti-LYVE1 (1:100, 102-PA50AG, ReliaTech, Braunschweig, Germany) and mouse anti-podoplanin (1:100, ab77854, Abcam, Cambridge, UK). The secondary antibodies were diluted in blocking solution and applied at RT for 1 hour followed by a nuclear counterstaining with 4',6-diamidino-2-phenylindole (DAPI, Life Technologies, Carlsbad, USA). The secondary antibodies used were: Alexa Fluor 488 donkey anti-rabbit (1:500, A21206, Life Technologies, Carlsbad, USA), Alexa Fluor 594 donkey anti-mouse (1:500, A-21203, Life Technologies, Carlsbad, USA), and Alexa Fluor 594 donkey anti-goat (1:500, A-11058, Life Technologies, Carlsbad, USA). The sections were then mounted using ProLong Gold (Life Technologies, Carlsbad, USA). As antibody specificity controls, primary antibodies were omitted.

IMAGING

The sections stained for Haematoxylin and Eosin were scanned with the Panoramic MIDI digital scanner (3DHISTECH Ltd., Budapest, Hungary). Selection of the desired areas and adjustments were performed with the Panoramic Viewer (3DHISTECH Ltd., Budapest, Hungary). The immunofluorescence images were acquired with a Leica DM5500 fluorescence upright microscope (Leica, Mannheim, Germany) equipped with a Cool Snap HQ2 CCD camera (Photometrics, Tucson, Arizona, USA) or a Leica TCS SP8 upright microscope (Leica, Mannheim, Germany) operated with the Leica Application Suite Advanced Fluorescence software (LAS AF). Brightness and contrast was adjusted either by using ImageJ (NIH, Bethesda, USA) or Photoshop CS6 (Adobe Systems Inc., San Jose, USA).

3D-RECONSTRUCTION

For 3D-reconstruction, serial paraffin transversal sections (10 µm) of fetal pancreatic specimens at W17 and W21 were immunostained for glucagon and podoplanin as described in histology and immunohistochemistry. The sections were digitalized using the Panoramic MIDI digital scanner (3DHISTECH Ltd., Budapest, Hungary) and the reconstruction was conducted with the Amira 4.1 software (Visage Imaging, Berlin, Germany).

PERI-PANCREATIC AND INTRA-PANCREATIC MESENCHYME IN THE HUMAN FETAL PANCREAS

In this study, we refer to two types of mesenchyme in the pancreas [Fig. 1A]**: The peri-pancreatic mesenchyme (PPM) and the intra-pancreatic mesenchyme (IPM). The PPM is the thin layer of connective tissue surrounding the pancreas, whereas the IPM is the connective tissue in which the endocrine and exocrine cells are embedded that will form the (interlobular) septa defining the lobular structure of the adult pancreas.**

At W9–W12, the developing pancreas contained a prominent layer of PPM compared to the IPM where the pancreatic epithelium and ducts localize [Fig. 1A, left panel]**. At W17–W22, both types of mesenchyme were less prominent compared to W9–W12, due to the epithelial expansion that leads to endocrine and exocrine differentiation** [Fig. 1A, middle and right panel]**.**

SPATIAL PROGRESSION OF ANGIOGENESIS IN THE PANCREATIC MESENCHYMAL COMPARTMENTS

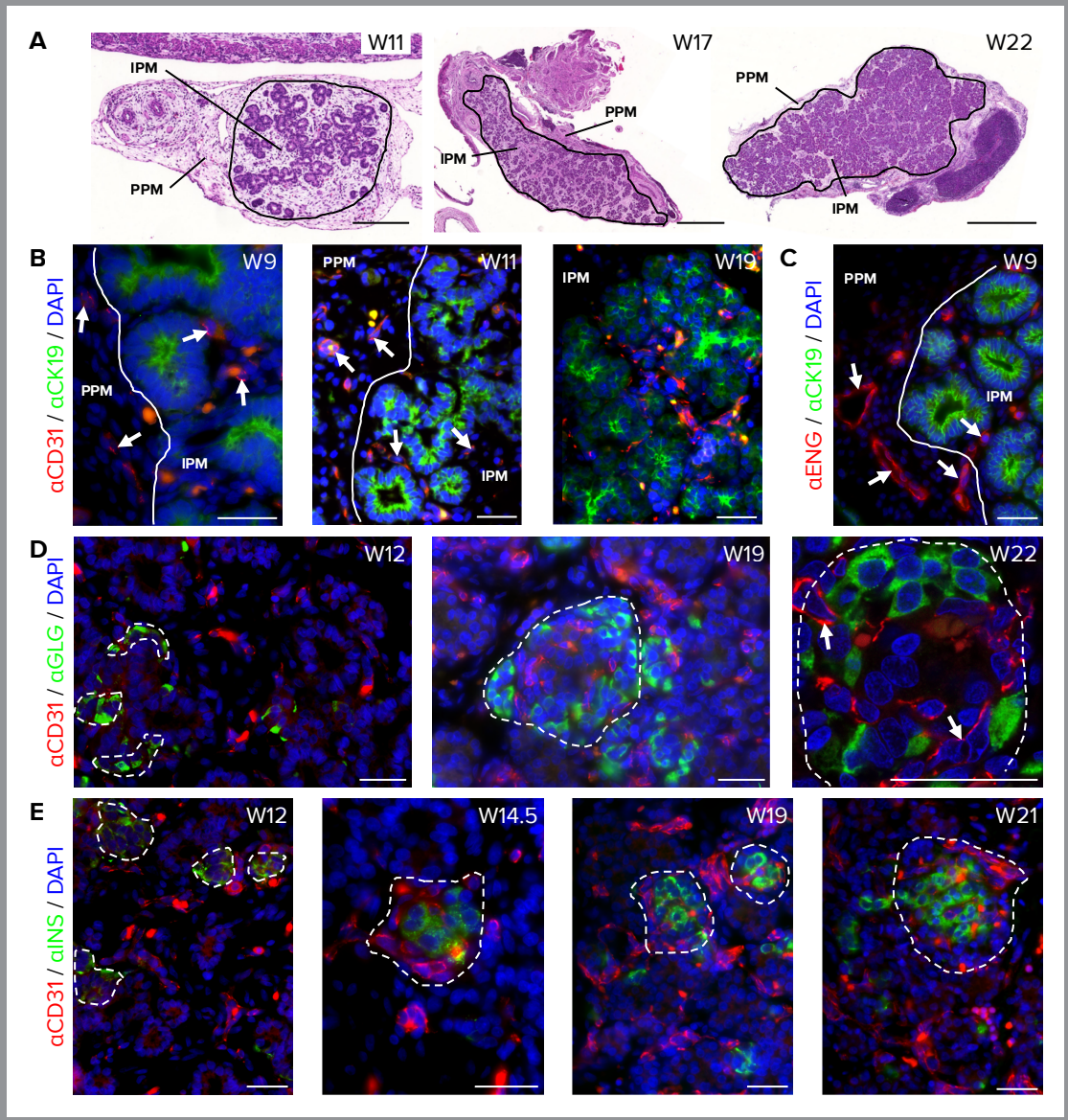
To investigate angiogenesis, we used an antibody against CD31 (or PECAM1), an established marker of endothelial cells [33, 39]**. From W9–W19, CD31-positive vessels were observed in both the PPM and IPM** [Fig. 1B]**, many close to CK19-positive pancreatic epithelial cells. To investigate whether the blood vessels at W9 (the earliest time point analyzed) resulted from angiogenesis or vasculogenesis, we used an antibody against the angiogenic marker endoglin** [40, 41]**. Already at W9, the blood vessels in both PPM and IPM were endoglin-positive** [Fig. 1C]**, suggesting that angiogenesis is the main mechanism for blood vessel formation in the pancreas during the investigated period.**

MICROCIRCULATION IN THE ISLETS OF LANGERHANS

Next, we investigated the relationship between the development of the endocrine compartment and angiogenesis in the developing pancreas. We detected insulin-positive and glucagon-positive endocrine cells within the pancreatic epithelium as early as W9 (data not shown). At W11, the glucagon-positive endocrine cells resided primarily as single cells within the pancreatic epithelium, whereas the insulin-positive endocrine cells already formed small clusters within the pancreatic epithelium (data not shown). At W12–W22, the endocrine compartment of the pancreas developed to form the islets of Langerhans, with the typical core-mantle morphology, with glucagon-producing α-cells forming the mantle [Fig. 1D] **and insulin-producing β-cells forming a compact core** [Fig. 1E] **as described** [13, 20]**. Interestingly, a network of CD31-positive capillaries was visible in proximity of the small glucagon-positive and insulin-positive cell clusters around W12 and penetrating the islets, forming the islet microcirculation, at W14.5–W22** [Fig. 1D and 1E]**. Our observations suggest that at least by W22 the islets of Langerhans and their microcirculation may form a physiological functional unit** [Fig. 1D, right panel]**.**

ASSOCIATION OF THE BLOOD VASCULATURE WITH SMOOTH MUSCLE CELLS IN THE DEVELOPING HUMAN PANCREAS

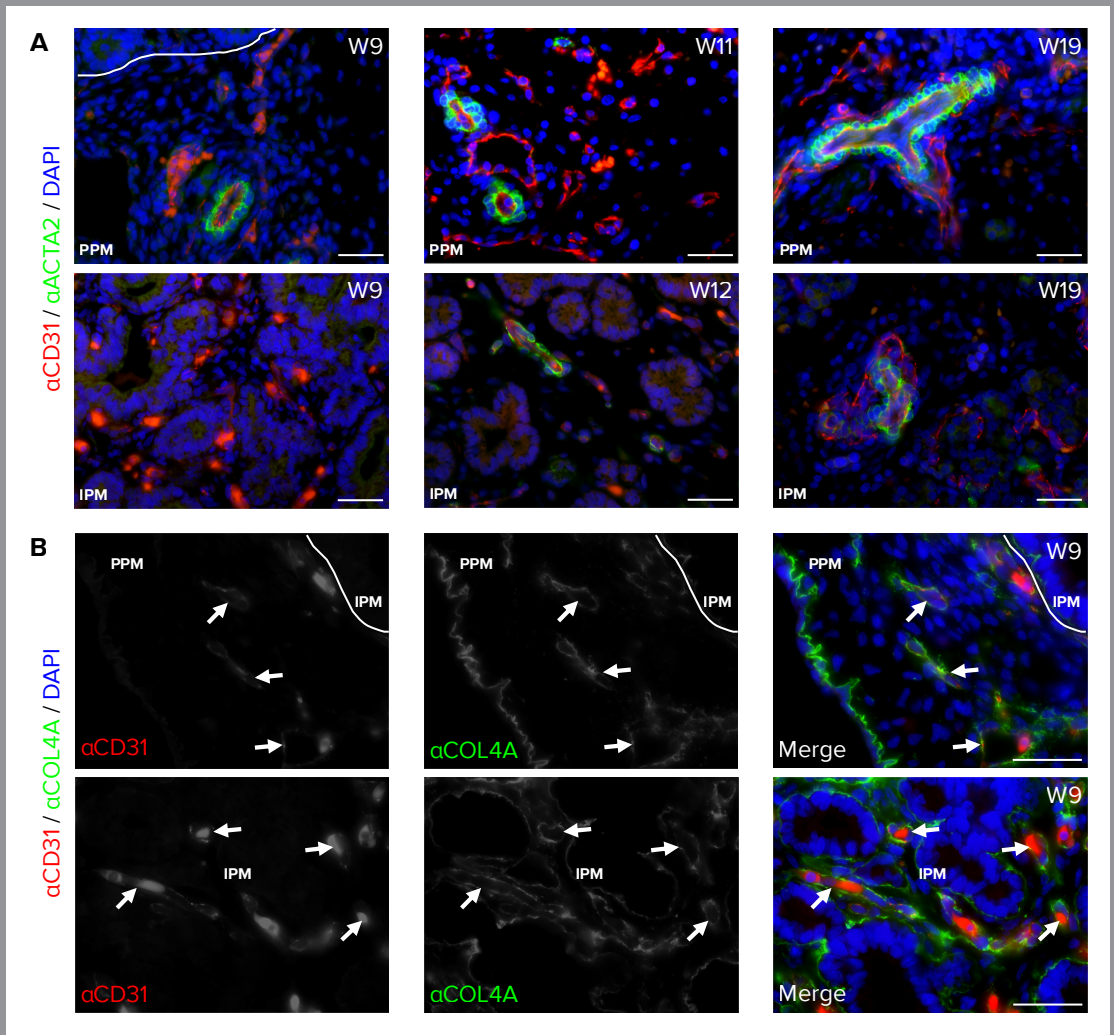
A key step in the maturation of the endothelial tubes to form arteries during vascular development is the attraction of mural cells and their subsequent differentiation to smooth muscle cells through endothelial cell association [42, 43]**. The association of smooth muscle cells, expressing alpha smooth muscle actin (ACTA2), with the blood vessels (arteries) was observed as early as W9** [Fig. 2A, left panels]**, but only in the PPM** [Fig. 2A, left panels]**. However, from W12 onwards the CD31-positive blood vessels in the IPM also started to show**


A
W11
IPM
PPM
W17
PPM
IPM
W22
PPM
IPM
B
αCD31 / αCK19 / DAPI
W9
PPM
IPM
W11
PPM
IPM
W19
IPM
C
αENG / αCK19 / DAPI
W9
PPM
IPM
D
αCD31 / αGLG / DAPI
W12
W19
W22
E
αCD31 / αINS / DAPI
W12
W14.5
W19
W21

[Fig. 1]

Angiogenesis during human pancreatic development.

[A] Haematoxylin and eosin (H&E) staining of pancreata at W11, W17 and W22 illustrating the two types of mesenchyme, the peri-pancreatic mesenchyme (PPM) and intra-pancreatic mesenchyme (IPM). The pancreatic epithelium is embedded in the IPM. **[B]** Pancreata at W9, W11 and W19 immunostained for CD31 and CK19. White line in the top panel shows the separation between the PPM and IPM. White arrows point to CD31-positive blood vessels. **[C]** Pancreas at W9 immunostained for endoglin (ENG) and CK19. **[D]** Pancreata at W12, W19 and W22 immunostained for CD31 and glucagon (GLG, α-cells). Dashed line shows the islet of Langerhans. White arrows point to capillaries. **[E]** Pancreata at W12, W14.5, W19 and W21 immunostained for CD31 and insulin (INS, β-cells). Note that autofluorescent red blood cells, as yellow/orange dots, are present in all images. Scale bars: (A, left panel) 200 µm, (A, middle and right panel) 1 mm, (B-E) 30 µm.


A
αCD31 / αACTA2 / DAPI
W9
W11
W19
PPM
W9
W12
W19
IPM
B
αCD31 / αCOL4A / DAPI
PPM
IPM
αCD31
αCOL4A
Merge
W9

[Fig. 2]

Blood vessels and their association with smooth muscle cells and basement membrane formation during human pancreatic development.

[A] Pancreata at W9, W11, W12 and W19 immunostained for CD31 and smooth muscle actin (ACTA2). Smooth muscle cell association with blood vessels in the peri-pancreatic mesenchyme (PPM, top panels) and intra-pancreatic mesenchyme (IPM, bottom panels) is shown. **[B]** Pancreas at W9 immunostained for CD31 and collagen type IV (COL4A) illustrating blood vessels in the PPM (top panels) and IPM (bottom panels). White arrows point to CD31-positive blood vessels with a continuous basement membrane. Note that autofluorescent red blood cells, as yellow/orange dots, are present in all images. Scale bars: 30 μm.

association with ACTA2-positive smooth muscle cells [Fig. 2A, middle panels]**. From W12–W22, the number of CD31-positive blood vessels associated with ACTA2-positive cells, presumably arteries, increased both in the IPM and PPM** [Fig. 2A, right panels]**, but some of the large caliber CD31-positive vessels, presumably veins, remained devoid of smooth muscle cells. Large caliber blood vessels (uncoated or coated with ACTA2-positive smooth muscle cells) were never observed in close association with islets of Langerhans. Concluding, the association of smooth muscle cells with blood vessels showed a clear spatial developmental delay of several weeks between the IPM and PPM. By contrast, at W9 blood vessels containing a collagen type IV (COL4A)-rich basement membrane, another key step in the vascular maturation, was already clearly visible in both PPM and IPM** [Fig. 2B]**.**

LYMPHANGIOGENESIS IN THE PANCREATIC MESENCHYMAL COMPARTMENTS

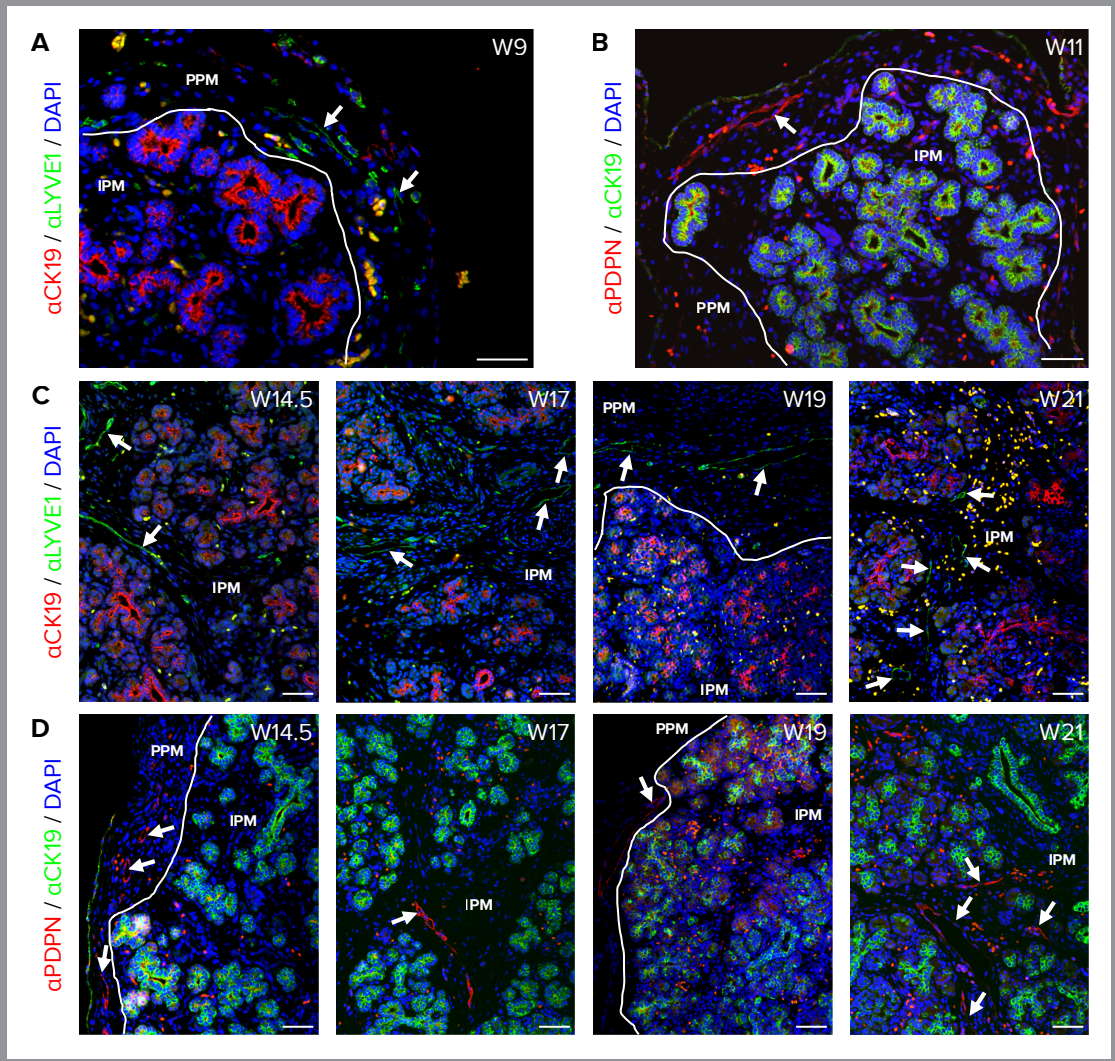
From W9 to about W14.5–W17, LYVE1-positive and PDPN-positive small lymphatic vessels were present exclusively in the PPM [Fig. 3A and 3B, white arrows]**. However, LYVE1, but not PDPN, is also known to be expressed by both CD68-positive and F4/80-positive macrophages** [44–46]**, and those were present as CD68-positive and LYVE1-positive single cells in both the PPM and IPM throughout development** [Supplementary Fig. 1A]**. By W17, LYVE1-positive and PDPN-positive lymphatic larger caliber vessels were visible in both the PPM and IPM** [Fig. 3C and 3D, white arrows]**, but were rarely or not observed penetrating the intralobular region containing the epithelial-derived tissue (the developing acinar or ductal structures and the islets of Langerhans). The first few amylase-positive acinar cells were observed at W14.5** [Supplementary Fig. 1B]**, the stage when lymphatic vessels start to colonize the IPM. However, we did not observe any direct association of lymphatic vessels with amylase-positive cells until W22 either** [Supplementary Fig. 1B]**.**

Even though there is no direct penetration, from 3D-reconstructions, we observed multiple lymphatic capillaries in close vicinity of islets of Langerhans, at least between W17–W21 [Fig. 4A and 4B; Supplementary Fig. 2 and 3]**. In summary, we observed both lymphatic and blood vessels in the PPM at W9. However, colonization of the IPM by lymphatic vessels (W14.5–W17) was delayed by several weeks compared to the colonization by blood vessels (W9–W11) and was excluded from the islets of Langerhans.**

ASSOCIATION WITH SMOOTH MUSCLE CELLS AND BASEMENT MEMBRANE FORMATION IN THE LYMPHATIC VASCULATURE

During maturation and remodeling of the lymphatic network, much like the blood vascular network, ACTA2-positive smooth muscle cells are recruited to coat the pre-collecting lymphatics rather sparsely and the collecting lymphatics more densely [47, 48]**. In contrast to the association of the blood vascular network with smooth muscle cells observed at W9 in the PPM, the association of ACTA2-positive smooth muscle cells with the PDPN-positive lymphatic vessels was observed only at about W14.5** [Fig. 5A]**, even though both lymph and blood vessels were present in the PPM from W9. In the IPM, the association of lymphatic vessels with smooth muscle cells occurred at about W17, where the largest PDPN-positive lymphatic vessels were only sparsely covered with smooth muscle cells** [Fig. 5A]**.**

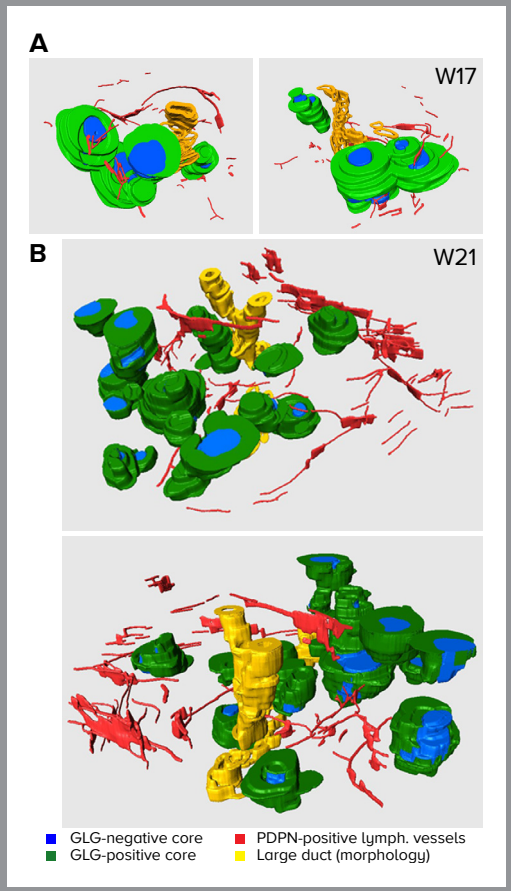
Next, we analyzed the expression of collagen type IV (COL4A), a component of the vascular basement membrane, but also a good indicator of maturation of the lymphatic vessels. COL4A forms a continuous basement membrane in all types of blood vessels and collecting lymphatic vessels, but it forms a discontinuous basement membrane in pre-collecting lymph vessels and does not form any basement membrane in the lymph capillaries [47, 49–52]**. By W11, larger-**


A
αCK19 / αLYVE1 / DAPI
W9
PPM
IPM
B
αPDPN / αCK19 / DAPI
W11
IPM
PPM
C
αCK19 / αLYVE1 / DAPI
W14.5
IPM
W17
IPM
PPM
W19
IPM
W21
IPM
D
αPDPN / αCK19 / DAPI
W14.5
PPM
IPM
W17
IPM
PPM
W19
IPM
W21
IPM

[Fig. 3]

Lymphangiogenesis during human pancreatic development.

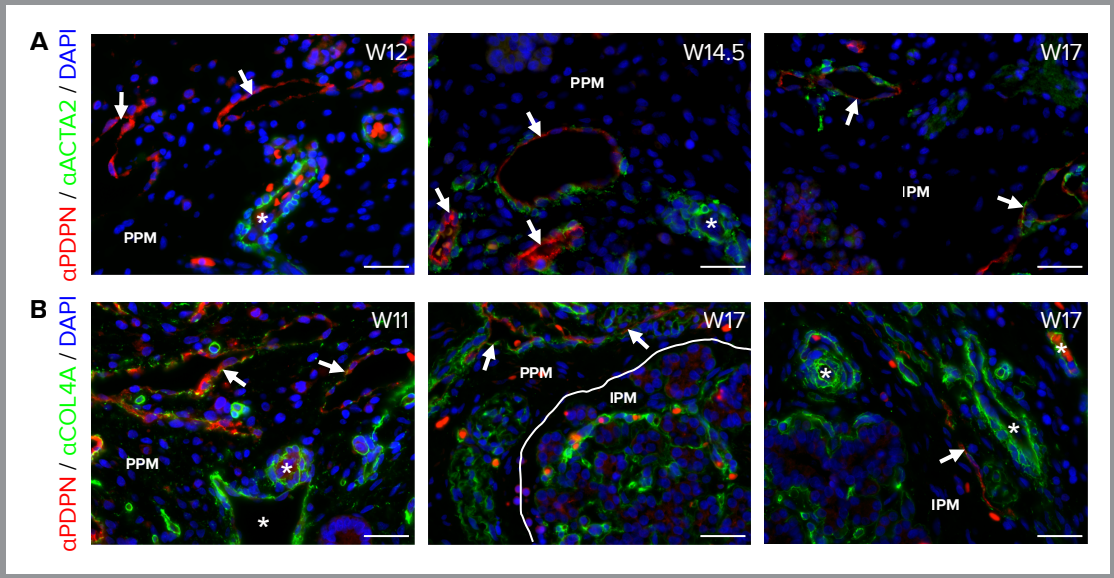
[A] Pancreas at W9 immunostained for CK19 and LYVE1. White arrows point to LYVE1-positive lymphatic vessels. White line shows the separation between the peri-pancreatic mesenchyme (PPM) and intra-pancreatic mesenchyme (IPM). **[B]** Pancreas at W11 immunostained for CK19 and podoplanin (PDPN). White arrows point to PDPN-positive lymphatic vessels. **[C]** Pancreata at W14.5, W17, W19 and W21 immunostained for CK19 and LYVE1. **[D]** Pancreata at W14.5, W17, W19 and W21 immunostained for CK19 and PDPN. Note that autofluorescent red blood cells, as yellow/orange dots, are present in all images. Scale bars: 50 µm.


A
W17
B
W21
GLG-negative core
GLG-positive core
PDPN-positive lymph. vessels
Large duct (morphology)

[Fig. 4]

Spatial arrangement of lymphatic vessels and islets of Langerhans during human pancreatic development.

[A, B] Three-dimensional reconstructions of islets of Langerhans at W17 **[A]** and W21 **[B]**, showing them from two different rotation angles. The mantle of glucagon (GLG)-positive a-cells and the GLG-negative core of the islets of Langerhans are depicted in green and blue, respectively. PDPN-positive lymphatic vessels are colored red and large ducts, identified by their morphology only, are represented in yellow.


A
αPDPN / αACTA2 / DAPI
W12
PPM
W14.5
PPM
W17
IPM
B
αPDPN / αCOL4A / DAPI
W11
PPM
W17
PPM
IPM
W17
IPM

[Fig. 5]

Association of lymphatic vessels with smooth muscle cells and basement membrane formation during human pancreatic development.

[A] Pancreata at W12, W14.5 and W17 immunostained for podoplanin (PDPN) and smooth muscle actin (ACTA2). PDPN-positive lymphatic vessels (white arrows) showed no association with smooth muscle cells up to W12 in contrast to arteries (white asterisks). From W14.5 onwards, PDPN-positive lymphatic vessels (white arrow) were sparsely covered with smooth muscle cells. **[B]** Pancreata at W11 and W17 immunostained for PDPN and COL4A. PDPN-positive lymphatic vessels (white arrows) were covered with a discontinuous basement membrane, whereas blood vessels (white asterisks) were covered by a continuous basement membrane. Scale bars: 30 µm.

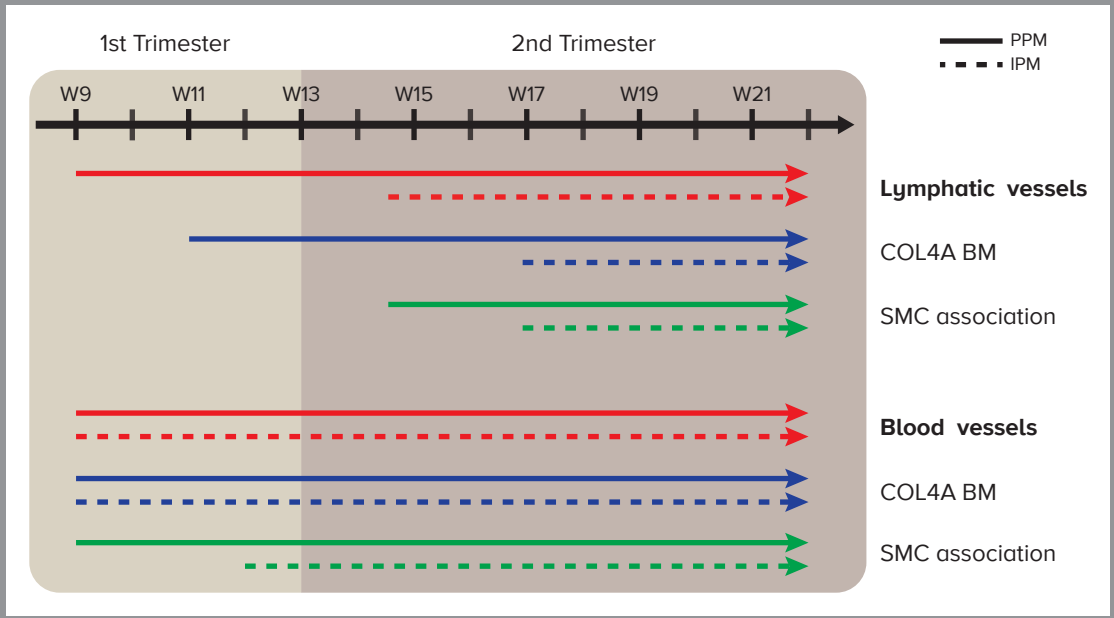
caliber PDPN-positive lymphatic vessels in the PPM showed a discontinuous COL4A-positive basement membrane [Fig. 5B, left panel]**. In the IPM, the invasion by the lymphatic vessels (W14.5–W17) seemed to occur virtually simultaneously with the appearance of a discontinuous COL4A-positive basement membrane (about W17)** [Fig. 5B, middle and right panels]**. This discontinuous COL4A-positive basement membrane was maintained at least until W22 in both the PPM and IPM. Together, our data suggest that between W9–W22 there are only lymph capillaries and pre-collecting lymphatic vessels and no collecting lymphatic vessels in the pancreas.**

During pancreas development, lymphangiogenesis occurred rather gradually in the PPM, with invasion at W9, deposition of basement membrane around W11 and coating by smooth muscle cells at W14.5; in the IPM, however, all steps occurred almost simultaneously within 3 weeks, between W14.5–W17 [Fig. 6]**.**

DISCUSSION

Our study showed that in the developing human pancreas, angiogenesis precedes lymphangiogenesis both in the PPM and the IPM [Fig. 6]**. In fact, by W9, mature smooth muscle cell-coated blood vessels were already observed in the PPM, whereas sparsely smooth muscle cell-coated lymph vessels were only observed there by W14.5 (a delay of about 6 weeks). In the IPM, we pinpointed the difference in timing of association with smooth muscle cells to be approximately 5 weeks (W12 for blood vessels and W17 for lymphatics). We conclude that the processes of angiogenesis and lymphangiogenesis followed independent developmental paths both temporally and spatially, which may perhaps be correlated with the development of the endocrine and exocrine compartments between W9–W22. Up to W22, we did not observe any PDPN-positive lymphatic vessels with a continuous layer of smooth muscle cells or a continuous COL4A-positive basement membrane, indicating that the formation of collecting lymphatic vessels may occur only later in development in the pancreas.**

We observed proximity between CD31-positive blood vessels and small clusters of glucagon-positive and insulin-positive cells at (W)12 weeks of gestation (equivalent to 10 weeks post conception) in agreement with Piper and colleagues (2004) that reported proximity to CD34-positive blood vessels at 10.5 weeks post conception [13]**. However, the following developmental stage analyzed by Piper and colleagues (2004) was week 14 post conception (equivalent to (W)16 weeks of gestation) when they observed CD34-positive blood vessels penetrating *bona fide* islets of Langerhans** [13]**. We now report penetration of the first islets of Langerhans by CD31-positive blood vessels at (W)14.5 weeks of gestation (equivalent to 12.5 weeks post conception).**


1st Trimester
2nd Trimester
PPM
IPM
W9
W11
W13
W15
W17
W19
W21
Lymphatic vessels
COL4A BM
SMC association
Blood vessels
COL4A BM
SMC association

[Fig. 6]

Model of the progression of lymphangiogenesis and angiogenesis during human pancreatic development.

Cartoon illustrating the progression of lymphangiogenesis (top) and angiogenesis (bottom) showing the timing of invasion, formation of a COL4A-rich basement membrane (BM) and the association with smooth muscle cells (SMCs) in the peri-pancreatic mesenchyme (PPM) and intra-pancreatic mesenchyme (IPM) between W9 and W22.

Furthermore, the immediate establishment of an extensive microvasculature in the islets of Langerhans while these are still being formed confirms the importance of blood vessels as integral part of the islets of Langerhans.

The islets of Langerhans were not directly invaded by lymphatic vessels, but these vessels were clearly present in the IPM in the vicinity of islets by W17. In agreement, in adult pancreas lymphatic vessels have also been observed not only in the interlobular connective tissue, but also intralobularly where the islets of Langerhans reside [53], suggesting some degree of proximity between lymphatics and islets as in the fetal pancreas. Even though the functionality of fetal islets of Langerhans regarding insulin release in response to glucose by W22 is still a matter of debate [54–57], the vascular machinery to support an insulin-glucose response by the islet of Langerhans seems to be in place.

On the overall morphology of the islets of Langerhans, we report that by W12 the majority of the developing islets of Langerhans exhibited a single core-mantle structure showing a characteristic compact core of insulin-producing β-cells partially surrounded by a thick mantle of glucagon-producing α-cells in agreement with others [13, 20]. However, in contrast to Jeon and colleagues (2009) that observed a predominantly homotypic character in the islets of Langerhans between W18–W21, we observed that the single core-mantle structure was maintained.

From the two lymphatic endothelial cell-specific antibodies used, PDPN is expressed by all lymphatic vessels, whereas LYVE1 is expressed by capillaries and pre-collecting lymphatic vessels, but not by the collecting lymphatic vessels [34, 48, 58], restricting its utility to early stages of lymphangiogenesis. Moreover, we observed that LYVE1, but not PDPN, is also expressed by CD68-positive macrophages as previously described [44–46]. Interestingly, it has been proposed that perhaps these LYVE1-positive macrophages could represent

lymphendothelial progenitors [59, 60]**. Combining the use of both LYVE1 and PDPN to mark lymphatic vessels provided a robust assessment of lymphangiogenesis in the developing human pancreas.**

CONCLUSIONS

We report here the first systematic study investigating the progression of lymphangiogenesis and angiogenesis between W9 and W22 of human pancreas development [Fig. 6]**. We show that both processes have their own dynamics of invasion and maturation, but both seem in place to provide a functional response by W22. Understanding the establishment of the two vascular systems during normal human pancreas development is of great interest to develop better protocols for transplantation of islets of Langerhans, as well as to optimize the molecular niche necessary for the differentiation of pluripotent stem cells to insulin-producing β-cells.**

ABBREVIATIONS

ACTA2: Alpha smooth muscle actin; BM: Basement membrane; CK19: Cytokeratin 19; COL4A: Collagen type IV; GLG: Glucagon; H&E: Haematoxylin and eosin; INS: Insulin; IPM: Intra-pancreatic mesenchyme; hESCs: Human embryonic stem cells; hiPSCs: Human induced pluripotent stem cells; LYVE1: Lymphatic vessel endothelial hyaluronan receptor 1; PDPN: Podoplanin; PPM: Peri-pancreatic mesenchyme; SMC: Smooth muscle cell; RT: Room temperature.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

M.S.R. designed and performed experiments, analyzed the data and wrote the manuscript. L.v.I. performed experiments. A.d.M.B. performed the 3D-reconstruction. S.M.C.d.S.L., C.L.M., F.C. and E.J.P.d.K. designed the experiments, analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

This work was supported by the Netherlands organization of Scientific Research (NWO) [ASPASIA 015.007.037] and the Interuniversity Attraction Poles-Phase VII [IUAP/PAI P7/14] to S.M.C.d.S.L., the Bontius Stichting [PANCREAS] to M.S.R., F.C., C.L.M., E.J.P.d.K., Stichting DON to E.J.P.d.K., and the Fundação para a Ciência e Tecnologia (FCT) [SFRH/BD/94387/2013] to A.d.M.B..

We would like to thank the Center for Contraception, Sexuality and Abortion (CASA) in Leiden and The Hague for collecting the human fetuses; K. Sprenkels, S. Anjie and I. Navarro for technical support with histology; J. Wiegant and A.M. van der Laan for technical support with microscopy; M. Heeren and H. Locher for advice regarding immunofluorescence;

L.J.A.C. Hawinkels for the anti-endoglin antibody; and A.C. Kraima for the fruitful discussions on the lymphatic vascular system.

REFERENCES

[1] Bluestone JA, Herold K, Eisenbarth G: **Genetics, pathogenesis and clinical interventions in type 1 diabetes**. *Nature* 2010, **464**(7293):1293–1300.

[2] Halban PA, German MS, Kahn SE, Weir GC: **Current status of islet cell replacement and regeneration therapy**. *The Journal of clinical endocrinology and metabolism* 2010, **95**(3):1034–1043.

[3] Weir GC, Cavelti-Weder C, Bonner-Weir S: **Stem cell approaches for diabetes: towards beta cell replacement**. *Genome medicine* 2011, **3**(9):61.

[4] Carlotti F, Zaldumbide A, Ellenbroek JH, Spijker HS, Hoeben RC, de Koning EJ: **beta-Cell Generation: Can Rodent Studies Be Translated to Humans?** *Journal of transplantation* 2011, **2011**:892453.

[5] D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, Moorman MA, Kroon E, Carpenter MK, Baetge EE: **Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells**. *Nature biotechnology* 2006, **24**(11):1392–1401.

[6] Wei R, Yang J, Hou W, Liu G, Gao M, Zhang L, Wang H, Mao G, Gao H, Chen G *et al*: **Insulin-producing cells derived from human embryonic stem cells: comparison of definitive endoderm- and nestin-positive progenitor-based differentiation strategies**. *PloS one* 2013, **8**(8):e72513.

[7] Wei R, Yang J, Liu GQ, Gao MJ, Hou WF, Zhang L, Gao HW, Liu Y, Chen GA, Hong TP: **Dynamic expression of microRNAs during the differentiation of human embryonic stem cells into insulin-producing cells**. *Gene* 2013, **518**(2):246–255.

[8] Hayek A, Beattie GM: **Processing, storage and experimental transplantation of human fetal pancreatic cells**. *Annals of transplantation : quarterly of the Polish Transplantation Society* 1997, **2**(3):46–54.

[9] Kroon E, Martinson LA, Kadoya K, Bang AG, Kelly OG, Eliazer S, Young H, Richardson M, Smart NG, Cunningham J *et al*: **Pancreatic endoderm derived from human embryonic stem cells generates**

glucose-responsive insulin-secreting cells in vivo. *Nature biotechnology* 2008, **26**(4):443–452.

[10] Rezania A, Bruin JE, Riedel MJ, Mojibian M, Asadi A, Xu J, Gauvin R, Narayan K, Karanu F, O'Neil JJ *et al*: **Maturation of human embryonic stem cell-derived pancreatic progenitors into functional islets capable of treating pre-existing diabetes in mice**. *Diabetes* 2012, **61**(8):2016–2029.

[11] Polak M, Bouchareb-Banaei L, Scharfmann R, Czernichow P: **Early pattern of differentiation in the human pancreas**. *Diabetes* 2000, **49**(2):225–232.

[12] Piper K, Ball SG, Keeling JW, Mansoor S, Wilson DI, Hanley NA: **Novel SOX9 expression during human pancreas development correlates to abnormalities in Campomelic dysplasia**. *Mechanisms of development* 2002, **116**(1–2):223–226.

[13] Piper K, Brickwood S, Turnpenny LW, Cameron IT, Ball SG, Wilson DI, Hanley NA: **Beta cell differentiation during early human pancreas development**. *The Journal of endocrinology* 2004, **181**(1):11–23.

[14] Wang R, Li J, Lyte K, Yashpal NK, Fellows F, Goodyer CG: **Role for beta1 integrin and its associated alpha3, alpha5, and alpha6 subunits in development of the human fetal pancreas**. *Diabetes* 2005, **54**(7):2080–2089.

[15] Lyttle BM, Li J, Krishnamurthy M, Fellows F, Wheeler MB, Goodyer CG, Wang R: **Transcription factor expression in the developing human fetal endocrine pancreas**. *Diabetologia* 2008, **51**(7):1169–1180.

[16] Sarkar SA, Kobberup S, Wong R, Lopez AD, Quayum N, Still T, Kutchma A, Jensen JN, Gianani R, Beattie GM *et al*: **Global gene expression profiling and histochemical analysis of the developing human fetal pancreas**. *Diabetologia* 2008, **51**(2):285–297.

[17] Gregg BE, Moore PC, Demozay D, Hall BA, Li M, Husain A, Wright AJ, Atkinson MA, Rhodes CJ: **Formation of a human beta-cell population within pancreatic islets is set early in life**. *The Journal of clinical endocrinology and metabolism* 2012, **97**(9):3197–3206.

[18] Riedel MJ, Asadi A, Wang R, Ao Z, Warnock GL, Kieffer TJ: **Immunohistochemical characterisation of cells co-producing insulin and glucagon in the developing human pancreas**. *Diabetologia* 2012,

55(2):372–381.

[19] Jennings RE, Berry AA, Kirkwood-Wilson R, Roberts NA, Hearn T, Salisbury RJ, Blaylock J, Piper Hanley K, Hanley NA: **Development of the human pancreas from foregut to endocrine commitment**. *Diabetes* 2013, **62**(10):3514–3522.

[20] Jeon J, Correa-Medina M, Ricordi C, Edlund H, Diez JA: **Endocrine cell clustering during human pancreas development**. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2009, **57**(9):811–824.

[21] Pan FC, Wright C: **Pancreas organogenesis: from bud to plexus to gland**. *Developmental dynamics : an official publication of the American Association of Anatomists* 2011, **240**(3):530–565.

[22] Ballian N, Brunicardi FC: **Islet vasculature as a regulator of endocrine pancreas function**. *World journal of surgery* 2007, **31**(4):705–714.

[23] Lammert E, Cleaver O, Melton D: **Induction of pancreatic differentiation by signals from blood vessels**. *Science* 2001, **294**(5542):564–567.

[24] Lammert E, Gu G, McLaughlin M, Brown D, Brekken R, Murtaugh LC, Gerber HP, Ferrara N, Melton DA: **Role of VEGF-A in vascularization of pancreatic islets**. *Current biology : CB* 2003, **13**(12):1070–1074.

[25] Lukinius A, Jansson L, Korsgren O: **Ultrastructural evidence for blood microvessels devoid of an endothelial cell lining in transplanted pancreatic islets**. *The American journal of pathology* 1995, **146**(2):429–435.

[26] Jansson L, Carlsson PO: **Graft vascular function after transplantation of pancreatic islets**. *Diabetologia* 2002, **45**(6):749–763.

[27] Brissova M, Fowler M, Wiebe P, Shostak A, Shiota M, Radhika A, Lin PC, Gannon M, Powers AC: **Intraislet endothelial cells contribute to revascularization of transplanted pancreatic islets**. *Diabetes* 2004, **53**(5):1318–1325.

[28] Cantaluppi V, Biancone L, Figliolini F, Beltramo S, Medica D, Deregibus MC, Galimi F, Romagnoli R, Salizzoni M, Tetta C *et al*: **Microvesicles derived from endothelial progenitor cells enhance neoangiogenesis of human pancreatic islets**. *Cell transplantation*

2012, **21**(6):1305–1320.

[29] O'Morchoe CC: **Lymphatic system of the pancreas**. *Microscopy research and technique* 1997, **37**(5–6):456–477.

[30] Rasio EA, Hampers CL, Soeldner JS, Cahill GF, Jr.: **Diffusion of glucose, insulin, inulin, and Evans blue protein into thoracic duct lymph of man**. *The Journal of clinical investigation* 1967, **46**(6):903–910.

[31] Girard JP, Moussion C, Forster R: **HEVs, lymphatics and homeostatic immune cell trafficking in lymph nodes**. *Nature reviews Immunology* 2012, **12**(11):762–773.

[32] Jurisic G, Detmar M: **Lymphatic endothelium in health and disease**. *Cell and tissue research* 2009, **335**(1):97–108.

[33] Ordonez NG: **Immunohistochemical endothelial markers: a review**. *Advances in anatomic pathology* 2012, **19**(5):281–295.

[34] Jackson DG: **The lymphatics revisited: new perspectives from the hyaluronan receptor LYVE-1**. *Trends in cardiovascular medicine* 2003, **13**(1):1–7.

[35] Adams RH, Alitalo K: **Molecular regulation of angiogenesis and lymphangiogenesis**. *Nature reviews Molecular cell biology* 2007, **8**(6):464–478.

[36] Choi I, Lee S, Hong YK: **The new era of the lymphatic system: no longer secondary to the blood vascular system**. *Cold Spring Harbor perspectives in medicine* 2012, **2**(4):a006445.

[37] Jin ZW, Nakamura T, Yu HC, Kimura W, Murakami G, Cho BH: **Fetal anatomy of peripheral lymphatic vessels: a D2-40 immunohistochemical study using an 18-week human fetus (CRL 155 mm)**. *Journal of anatomy* 2010, **216**(6):671–682.

[38] Jeong YJ, Cho BH, Kinugasa Y, Song CH, Hirai I, Kimura W, Fujimiya M, Murakami G: **Fetal topohistology of the mesocolon transversum with special reference to fusion with other mesenteries and fasciae**. *Clinical anatomy* 2009, **22**(6):716–729.

[39] Pusztaszeri MP, Seelentag W, Bosman FT: **Immunohistochemical expression of endothelial markers CD31, CD34, von Willebrand factor, and Fli-1 in normal human tissues**. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry*

Society 2006, **54**(4):385–395.

[40] ten Dijke P, Goumans MJ, Pardali E: **Endoglin in angiogenesis and vascular diseases**. *Angiogenesis* 2008, **11**(1):79–89.

[41] Li DY, Sorensen LK, Brooke BS, Urness LD, Davis EC, Taylor DG, Boak BB, Wendel DP: **Defective angiogenesis in mice lacking endoglin**. *Science* 1999, **284**(5419):1534–1537.

[42] Folkman J, D'Amore PA: **Blood vessel formation: what is its molecular basis?** *Cell* 1996, **87**(7):1153–1155.

[43] Wu J, Bohanan CS, Neumann JC, Lingrel JB: **KLF2 transcription factor modulates blood vessel maturation through smooth muscle cell migration**. *The Journal of biological chemistry* 2008, **283**(7):3942–3950.

[44] Attout T, Hoerauf A, Denece G, Debrah AY, Marfo-Debrekyei Y, Boussinesq M, Wanji S, Martinez V, Mand S, Adjei O *et al*: **Lymphatic vascularisation and involvement of Lyve-1+ macrophages in the human onchocerca nodule**. *PloS one* 2009, **4**(12):e8234.

[45] Harvey NL, Gordon EJ: **Deciphering the roles of macrophages in developmental and inflammation stimulated lymphangiogenesis**. *Vascular cell* 2012, **4**(1):15.

[46] Schroedl F, Brehmer A, Neuhuber WL, Kruse FE, May CA, Cursiefen C: **The normal human choroid is endowed with a significant number of lymphatic vessel endothelial hyaluronate receptor 1 (LYVE-1)-positive macrophages**. *Investigative ophthalmology & visual science* 2008, **49**(12):5222–5229.

[47] Paupert J, Sounni NE, Noel A: **Lymphangiogenesis in post-natal tissue remodeling: lymphatic endothelial cell connection with its environment**. *Molecular aspects of medicine* 2011, **32**(2):146–158.

[48] Maby-El Hajjami H, Petrova TV: **Developmental and pathological lymphangiogenesis: from models to human disease**. *Histochemistry and cell biology* 2008, **130**(6):1063–1078.

[49] Tammela T, Alitalo K: **Lymphangiogenesis: Molecular mechanisms and future promise**. *Cell* 2010, **140**(4):460–476.

[50] Sauter B, Foedinger D, Sterniczky B, Wolff K, Rappersberger K: **Immunoelectron microscopic characterization of human dermal lymphatic microvascular endothelial cells. Differential expression**

of CD31, CD34, and type IV collagen with lymphatic endothelial cells vs blood capillary endothelial cells in normal human skin, lymphangioma, and hemangioma in situ. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 1998, **46**(2):165–176.

[51] Lutter S, Xie S, Tatin F, Makinen T: **Smooth muscle-endothelial cell communication activates Reelin signaling and regulates lymphatic vessel formation**. *The Journal of cell biology* 2012, **197**(6):837–849.

[52] Norrmen C, Ivanov KI, Cheng J, Zangger N, Delorenzi M, Jaquet M, Miura N, Puolakkainen P, Horsley V, Hu J *et al*: **FOXC2 controls formation and maturation of lymphatic collecting vessels through cooperation with NFATc1**. *The Journal of cell biology* 2009, **185**(3):439–457.

[53] Rubbia-Brandt L, Terris B, Giostra E, Dousset B, Morel P, Pepper MS: **Lymphatic vessel density and vascular endothelial growth factor-C expression correlate with malignant behavior in human pancreatic endocrine tumors**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2004, **10**(20):6919–6928.

[54] Ashworth MA, Leach FN, Milner RD: **Development of insulin secretion in the human fetus**. *Archives of disease in childhood* 1973, **48**(2):151–152.

[55] Hoffman L, Mandel TE, Carter WM, Koulmanda M, Martin FI: **Insulin secretion by fetal human pancreas in organ culture**. *Diabetologia* 1982, **23**(5):426–430.

[56] Milner RD, Ashworth MA, Barson AJ: **Insulin release from human foetal pancreas in response to glucose, leucine and arginine**. *The Journal of endocrinology* 1972, **52**(3):497–505.

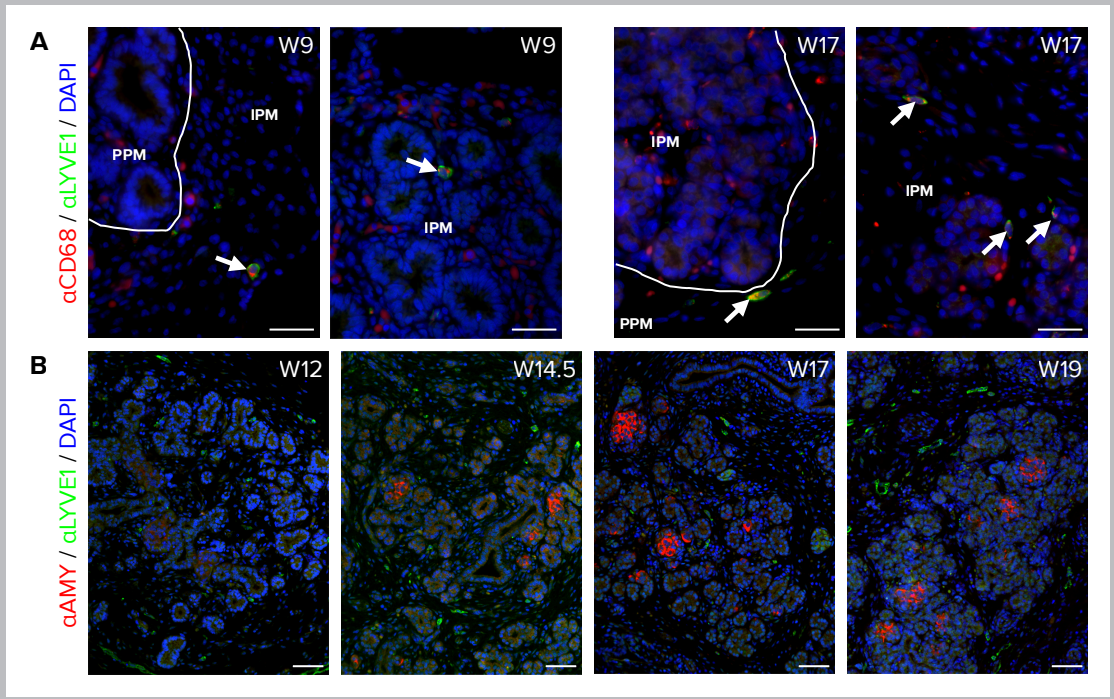
[57] Otonkoski T, Andersson S, Knip M, Simell O: **Maturation of insulin response to glucose during human fetal and neonatal development. Studies with perifusion of pancreatic isletlike cell clusters**. *Diabetes* 1988, **37**(3):286–291.

[58] Makinen T, Norrmen C, Petrova TV: **Molecular mechanisms of lymphatic vascular development**. *Cellular and molecular life sciences :*

CMLS 2007, **64**(15):1915–1929.

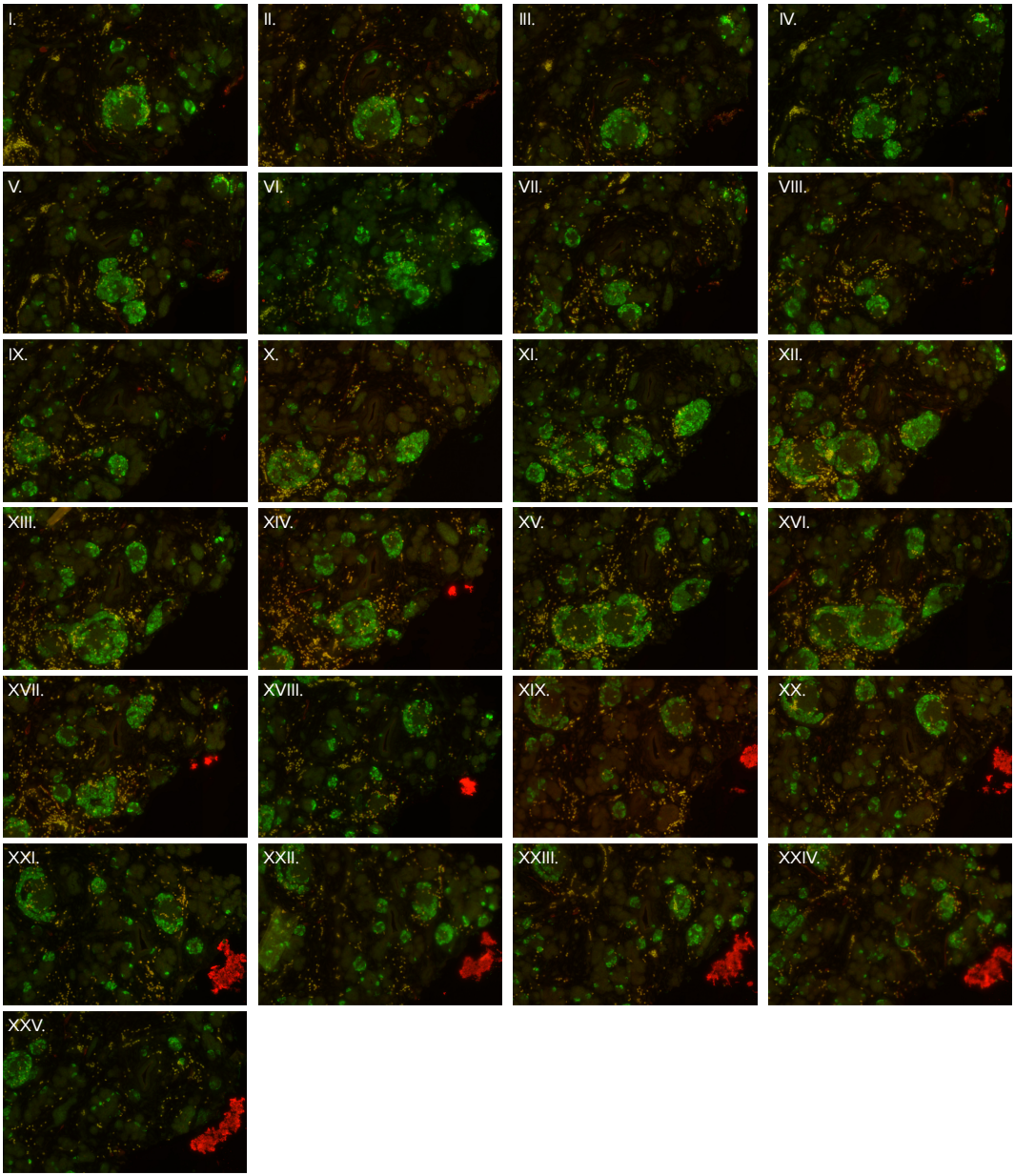
[59] Buttler K, Ezaki T, Wilting J: **Proliferating mesodermal cells in murine embryos exhibiting macrophage and lymphendothelial characteristics**. *BMC developmental biology* 2008, **8**:43.

[60] Buttler K, Kreysing A, von Kaisenberg CS, Schweigerer L, Gale N, Papoutsi M, Wilting J: **Mesenchymal cells with leukocyte and lymphendothelial characteristics in murine embryos**. *Developmental dynamics : an official publication of the American Association of Anatomists* 2006, **235**(6):1554–1562.


A
αCD68 / αLYVE1 / DAPI
W9
PPM
IPM
W9
IPM
W17
IPM
PPM
W17
IPM
B
αAMY / αLYVE1 / DAPI
W12
W14.5
W17
W19

[Supplementary Fig. 1]

LYVE1-positive macrophages and exocrine differentiation during human pancreas development. **[A]** Pancreata at W9 and W17 immunostained for LYVE1 and CD68, a marker to identify cells of the macrophage lineage (white arrows) in the peri-pancreatic mesenchyme (PPM; top panels) and intra-pancreatic mesenchyme (IPM; bottom panels). **[B]** Pancreata at W12, W14.5, W17 and W19 stained for amylase (AMY) and LYVE1. Scale bars: (A) 30 µm, (B) 50 µm.


I.
II.
III.
IV.
V.
VI.
VII.
VIII.
IX.
X.
XI.
XII.
XIII.
XIV.
XV.
XVI.
XVII.
XVIII.
XIX.
XX.
XXI.
XXII.
XXIII.
XXIV.
XXV.

[Supplementary Fig. 2]

Spatial arrangement of lymphatic vessels and islets of Langerhans in a W17 human pancreas. Consecutive sections of a W17 pancreas immunostained for PDPN and glucagon (GLG, α-cells) used for the 3D-reconstruction in Fig. 4A.

I.
II.
III.
IV.
V.
VI.
VII.
VIII.
IX.
X.
XI.
XII.
XIII.
XIV.
XV.
XVI.
XVII.
XVIII.
XIX.
XX.
XXI.
XXII.
XXIII.

[Supplementary Fig. 3]

Spatial arrangement of lymphatic vessels and islets of Langerhans in a W21 human pancreas. Consecutive sections of a W21 pancreas immunostained for PDPN and glucagon (GLG, α-cells) used for the 3D-reconstruction in Fig. 4B.

CHAPTER 3

KEYGENES, A TOOL TO PROBE TISSUE DIFFERENTIATION USING A HUMAN FETAL TRANSCRIPTIONAL ATLAS

Matthias S. Roost[1], Liesbeth van Iperen[1], Yavuz Ariyurek[2], Henk P. Buermans[2], Wibowo Arindrarto[3], Harsha D. Devalla[1], Robert Passier[1], Christine L. Mummery[1], Françoise Carlotti[4], Eelco J.P. de Koning[4, 6], Erik W. van Zwet[5], Jelle J. Goeman[7], Susana M. Chuva de Sousa Lopes[1,8]

[1] Department of Anatomy and Embryology,
[2] Leiden Genome Technology Center,
[3] Sequence Analysis Support Core,
[4] Department of Nephrology,
[5] Department of Medical Statistics and Bioinformatics,
Leiden University Medical Center, Leiden, the Netherlands
[6] Hubrecht Institute for Developmental Biology and Stem Cell Research,
University Medical Center, Utrecht, the Netherlands
[7] Department for Health Evidence, Radboud University Medical Center,
Nijmegen, the Netherlands
[8] Department for Reproductive Medicine, Ghent University Hospital,
Ghent, Belgium

Stem Cell Reports 2015, **4**(6):1112–24

SUMMARY

Differentiated derivatives of human pluripotent stem cells in culture are generally phenotypically immature compared to their adult counterparts. Their identity is often difficult to determine with certainty because little is known about their human fetal equivalents *in vivo*. Cellular identity and signalling pathways directing differentiation are usually determined by extrapolating information from either human adult tissue or model organisms, assuming conservation with humans. To resolve this, we generated a collection of human fetal transcriptional profiles at different developmental stages. Moreover, we developed an algorithm, KeyGenes, which uses this dataset to quantify the extent to which next-generation sequencing or microarray data resemble specific cell or tissue types in the human fetus. Using KeyGenes combined with the human fetal atlas, we identified multiple cell and tissue samples unambiguously on a limited set of features. We thus provide a flexible and expandable platform to monitor and evaluate the efficiency of differentiation *in vitro*.

INTRODUCTION

Detailed information on temporal and spatial patterns of gene expression during human development is essential to understand how cells establish and maintain their transcriptional identity and how they differentiate from common progenitors to form the different organs in the human body. Moreover, knowledge of the gene expression landscape in a physiological context is of paramount importance to identify aberrant patterns of transcription leading to pathological states [1, 2]**. To date, there is little information on the transcriptional profiles of organs and tissues during human development, even though optimal use of differentiated derivatives of human pluripotent stem cells (hPSCs) in regenerative medicine and disease modelling would benefit from detailed understanding of what drives and maintains the differentiated state. This knowledge is not only helpful in developing efficient differentiation protocols for hPSCs and unequivocally identifying the resultant phenotypes** [3, 4]**, but also in understanding why individual (genomic) variations between hPSCs may result in different outcomes** [5]**, and whether the wide assumption that some disease states are associated with upregulation of fetal genes is actually based on fact** [6–8]**.**

Recently, there has been great interest in computational approaches that help to quantify differences in physiological and pathological states in human adult organs and tissues [9]**, and similarities between adult human organs/tissues and differentiated cells derived from either hPSCs or by direct lineage conversion, also known as transdifferentiation** [10, 11]**. The drawback of these computational approaches so far is that they have been based on adult human organs/tissues, which limits their relevance in assigning developmental states to differentiated derivatives of hPSCs or transdifferentiated somatic cells since they are often immature and resemble fetal cells more**

than those of mature adult tissues [12, 13]**. Moreover, most current computational tools use and compare to microarray datasets only. Therefore, there is a growing need for a computational platform that can integrate next-generation sequencing (NGS) and microarray datasets and facilitate their interrogation. Here, we present an algorithm, KeyGenes, that we have used on NGS data extracted from tissues of 21 different human fetal organs, both embryonic and extraembryonic (plus the maternal endometrium), from the first and second trimesters of development to determine a panel of classifier genes that would be sufficient to confer identity to each fetal organ analysed with high confidence. We showed that the developmental classifier genes selected were largely sufficient to predict the identity of their adult organ counterparts, even when using different types of platforms (NGS and microarray). Most importantly, as proof of concept, we challenged KeyGenes to identify a series of tissues using either recently published or our own NGS datasets. These included (1) hPSCs differentiated to derivatives of the three germ lineages: endoderm (pancreas), ectoderm (brain) and mesoderm (heart); (2) tissue organoids (intestine) and (3) human fetal and adult organs/tissues. In all cases, KeyGenes accurately predicted tissue origin and, furthermore, we could use KeyGenes to assign a developmentally equivalent stage. KeyGenes is an easy-to-use, flexible and expandable tool that can be applied to identify stem cell derivatives when common marker profiles have been insufficiently informative and provide benchmarking for protocols designed to promote maturation of stem cell derivatives in culture. KeyGenes is available at www.keygenes.nl.**

EXPERIMENTAL PROCEDURES

ETHICAL STATEMENT, FETAL TISSUE PROCUREMENT

This study has been approved by the Medical Ethical Committee of the Leiden Medical University Center (P08.087). Informed consent was obtained and the study was conducted in accordance with the Declaration of Helsinki by the World Medical Association. Human fetal organ and tissue samples (n = 111) from 17 individuals representing 21 organs, and maternal endometrium, between gestational W8.2 and W22 [Supplementary Table 1] **were obtained from elective abortion material (vacuum aspiration) without medical indication. After washing with 0.9% NaCl (Fresenius Kabi, France), the organs and tissues were snap-frozen in buffer RLT (Qiagen, Germany) and stored at -80°C until further use. The organs and tissues were sex genotyped using primers for AMELOGENIN** [14] **as described** [15]**.**

NGS LIBRARY PREPARATION, SEQUENCING AND DATA PROCESSING

The organs and tissues were homogenized using a pestle or a syringe needle followed by the QIAshredder homogenizer (Qiagen, Germany). Total RNA was isolated with the RNeasy Kit (Qiagen) including on-column DNase digestion. RNA quality was assessed with the Agilent RNA 6000 Pico or Nano Kit on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). RNA concentration was determined with the Qubit RNA BR Assay Kit on a Qubit 2.0 Fluorometer (Invitrogen, USA).

The DeepSAGE libraries (fetal organs and tissues) were generated as previously described [16]**. Each library (8 pM) was loaded on a v3 flowcell and sequenced on an Illumina HiSeq2000 sequencer (Illumina, USA). Sequencing data was processed with Illumina Pipeline Software version CASAVA 1.8.2, and the FASTQ files were analyzed with a set of custom Perl**

scripts followed by aligning the sequences against the UCSC hg19 reference genome using Bowtie short read aligner 1.0.0. Finally, the tags were annotated against Ensembl gene identifiers using BIOMART.

The strand-specific RNA-Seq libraries of the NKX2.5:GFP-positive and -negative cells were generated essentially as previously described [17]. Following amplification of the strand-specific libraries, they were paired-end sequenced (2 x 75 bp) on a single NextSeq500 Mid output lane (Illumina, USA). Base quality trimming was performed on the raw sequencing files using SICKLE 1.33. Afterwards, the sequences were aligned to a custom hg19 human genome using GSNAP release 2014-12-06. We ran GSNAP by setting the following flags: --batch (set to 4), --novelsplicing (set to 1), --npaths (set to 1), and -- quiet-if-excessive. The last two flags were required so that reads map if and only if it can be mapped to a unique location in the genome. The resulting SAM file was then compressed into a BAM file, name-sorted, and indexed using SAMTOOLS 0.1.18. Finally, gene read count data were generated using HTSeq-count 0.6.1p1 using the Ensembl reference downloaded via the UCSC website (http://genome.ucsc.edu), using the non-stranded and intersectionnonempty option flags (--stranded no and --m intersection-nonempty respectively).

EXTERNAL DATA

Gene expression data were obtained either from the GEO database under the accession numbers GSE54879 [18], GSE52431 [19], GSE56930 [20] and GSE53949 [21], or from the EMBK-EBI database under the accession numbers E-MTAB-513 (Illumina Body Map) and E-MTAB-1733 [22]. The microarray data was downloaded from the Gene Expression Barcode 3.0 (http://barcode.luhs.org) [23] or from the GEO database under the accession number GSE30652 [24].

For the Illumina Bodymap dataset, we performed gene read counting using HTSeqcount version 0.6.1p1 using

the Ensembl human genes annotation set downloaded from the UCSC website (http://genome.ucsc.edu) in December 2014. We used the non-stranded and intersection-nonempty option flags (--stranded no and --m intersection-nonempty respectively) as additional arguments to the program. As the Illumina BodyMap 2.0 data were comprised of single-end and paired-end libraries, we separated them first and performed read counting on the separated library.

BIOINFORMATICS

KeyGenes algorithm: The algorithm uses a 10-fold cross validation on the basis of a LASSO (Least Absolute Shrinkage and Selection Operator) regression available in the R package «glmnet» [25]**. To use KeyGenes, to access the R scripts used here, and to access extra information on the human fetal data or the training sets used here, please go to www.keygenes.nl. We used three R scripts: (1) script 1 was used to determine the 500 most variably expressed genes across a NGS dataset (top 500), which will be saved as a .txt file. This file is then used in either script 2 or 3; (2) script 2 uses a NGS training set to predict a NGS test set. In addition, it uses the .txt file with the 500 most variably expressed genes generated with script 1; (3) Script 3 uses a NGS training set to predict a microarray test set. In addition, it uses the .txt file with the 500 most variably expressed genes generated with script 1; The script 3 is identical to script 2, but it includes a scaling step to a broad panel of 3787 housekeeper genes** [26]**. To do so, the algorithm looks for the housekeeper genes that are present in the training set and the test set, and chooses those that are expressed in at least one tissue.**

Gene expression levels: As detection limit, a cutoff value of four reads was used. To evaluate the counts per million (CPM) expression levels, our fetal data and the tested differentiation experiments were analyzed with the R package edgeR 3.2.4, using the weighted trimmed mean of M-values

(TMM) method to normalize [27, 28]**.**

Hierarchical clustering: Hierarchical clustering using complete linkage was based on the Pearson correlation of the gene expression levels (CPM) calculated using the base R package stats.

Gene ontology analyses: The enrichment of gene ontology terms for the 90 classifier genes was tested with DAVID [29]**. A FDR cutoff of 0.05 was used.**

Venn diagrams for gene expression: The datasets containing the first trimester fetal samples (1T), the second trimester fetal samples (2T) and the adult samples ([22] **and Illumina Bodymap) were compared per organ. The genes with gene expression levels higher than 10x the mean expression of the corresponding dataset (1T, 2T, adult) have been assigned either unique or common appearance across the datasets per organ and visualized in a Venn diagram** [30]**.**

Visualization*:* The data was primarily visualized using the R packages gplots and ggplot2 [31, 32]**.**

TISSUE CULTURE AND IMMUNOFLUORESCENCE

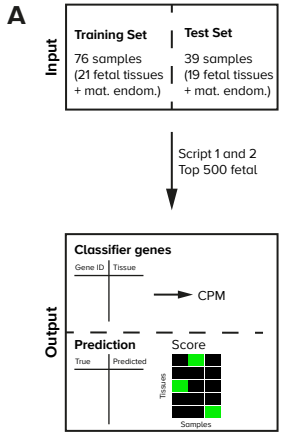
Cardiomyocytes from NKX2.5:GFP hESCs were generated as described [33] [Supplementary Fig. 4A]**. After 30 days, NKX2.5:GFP-positive cells were FACS sorted using a BD FACSAria III Cell sorter (BD Biosciences, USA) and used for library preparation. Immunofluorescence for COUP-TFII and GFP followed by imaging with confocal microscopy were described elsewhere** [33]**.**

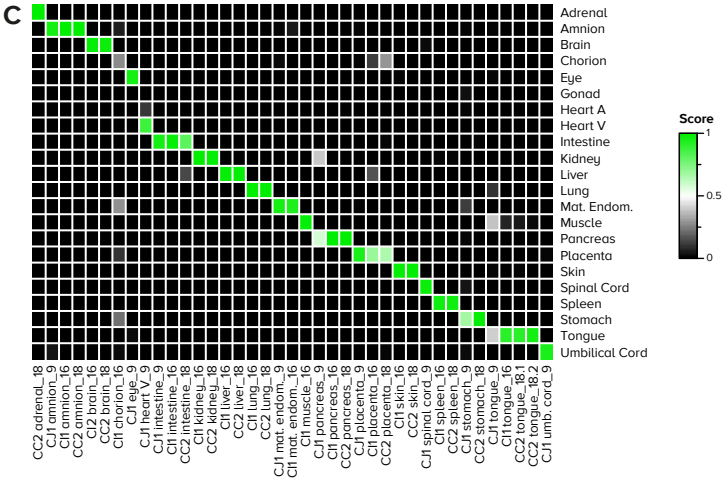
RESULTS

KEYGENES DEFINES A TRANSCRIPTIONAL BARCODE TO PREDICT HUMAN FETAL ORGAN IDENTITY

We wanted to identify a (sex-independent) transcriptional barcode that would characterize each organ primordium during human fetal development, as these are still relatively homogeneous and mainly formed by a limited number of poorly characterized lineage progenitor cells.

To do this, we first generated NGS data from 111 human organ/tissue samples from 17 individuals (n = 9 males and n = 8 females) representing 21 different fetal organs, 17 embryonic and four extraembryonic (plus the maternal endometrium), at 8.2–9.6, 16–18 and 21–22 weeks of gestation (W9, W16–18, W22) [Supplementary Table 1]**. Next, we developed an algorithm (KeyGenes) that uses a 10-fold cross validation on the basis of a LASSO (Least Absolute Shrinkage and Selection Operator) regression available in the R package «glmnet»** [25] [Fig. 1A]**. KeyGenes uses two datasets (a training set and a test set) and one table with the 500 most variably expressed genes. Then, KeyGenes determines a set of classifier genes that is sufficient to identify the samples of the training set, and finally, using the classifier genes, it predicts the identity of the samples of the test set. To validate KeyGenes, we divided our fetal dataset into a training set (n = 76) and a test set (n = 39)** [Fig. 1A and 1B]**, so that at least two samples of each organ, preferentially of both first and second trimester, were represented. Based on the classifier genes** [Supplementary Table 2] **and the identity of the samples in the training set, KeyGenes then predicted the identity of the samples in the test set** [Fig. 1C]**. KeyGenes was capable of predicting 38 of 39 samples of the test set correctly with a mean identity score of 0.92, of which 33 samples (87%) had an identity score > 0.8** [Supplementary Table 3]**.**





B

		Extraembryonic				Ectoderm				Endoderm					Mesoderm								Other
		Amnion	Chorion	Placenta	Umb. Cord	Brain	Eye	Skin	Sp. Cord	Intestine	Liver	Lung	Pancreas	Stomach	Adrenal	Gonad	Heart A	Heart V	Kidney	Sk. Muscle	Spleen	Tongue	Mat. Endom.
Training Set (n=76)	**1T**	1	0	1	2	2	1	2	2	1	2	2	2	1	1	2	0	1	2	2	0	1	1
	2T	3	2	3	0	1	4	3	0	3	2	2	2	3	1	2	4	3	2	3	2	1	1
Test Set (n=39)	**1T**	1	0	1	1	0	1	0	1	1	0	0	1	1	0	0	0	1	0	0	0	1	1
	2T	2	1	2	0	2	0	2	0	2	2	2	2	1	1	0	0	0	2	1	2	3	1

[Fig. 1]

Validation of KeyGenes using a human fetal transcriptome dataset.

[A] The fetal dataset was divided into a training set and a test set. KeyGenes used the 500 most variably expressed genes across the entire fetal dataset (Top 500 fetal) to determine a panel of classifier genes. KeyGenes then used the classifier genes to predict the identity (and provides the identity score) of the samples in the test set. **[B]** Characteristics of the human fetal samples in the training set and test set. **[C]** Identity scores for the samples in the fetal test set. The rows represent the 22 organs/tissues from the fetal training set and the columns depict the samples in the test set. The identity scores range from zero (black) to one (green). The values of all identity scores are given in Supplementary Table 3. Abbreviations: 1T, first trimester; 2T, second trimester; CPM, counts per million; heart A, heart atrium; heart V, heart ventricle; mat. endom., maternal endometrium; sk. muscle, skeletal muscle; sp. cord, spinal cord; umb. cord, umbilical cord.

KeyGenes thus provided a highly accurate prediction of fetal tissue identity for the test fetal samples.

THE TRANSCRIPTIONAL BARCODE FOR HUMAN FETAL ORGANS DURING DEVELOPMENT

A common characteristic of the 90 fetal classifier genes identified [Supplementary Table 2] **is that these are expressed at high levels throughout development in the organ(s)/tissue(s) they characterize (in general > 100 counts per million (CPM), but some > 1000 CPM), and they are either not expressed in most other tissues or highly expressed in several other tissues simultaneously, helping to define the barcode** [Fig. 2A]**. Interestingly, large differences in expression levels of some classifier genes were even sufficient to distinguish between first and second trimester organs. For example, *CYP17A1* in the adrenal is expressed > 100 CPM during first trimester, but > 1000 CPM in the second trimester** [Fig. 2A]**. Five classifier genes (*TNMD, RSPO2, LINC00514, NR5A1* and *CRABP1*) identified two different fetal organs/tissues** [Supplementary Table 2]. **The selected classifiers for each organ/tissue varied from one gene (*NPPA-AS1* for heart atrium and *MYL3* for heart ventricle) to nine genes for the umbilical cord** [Fig. 2B]**.**

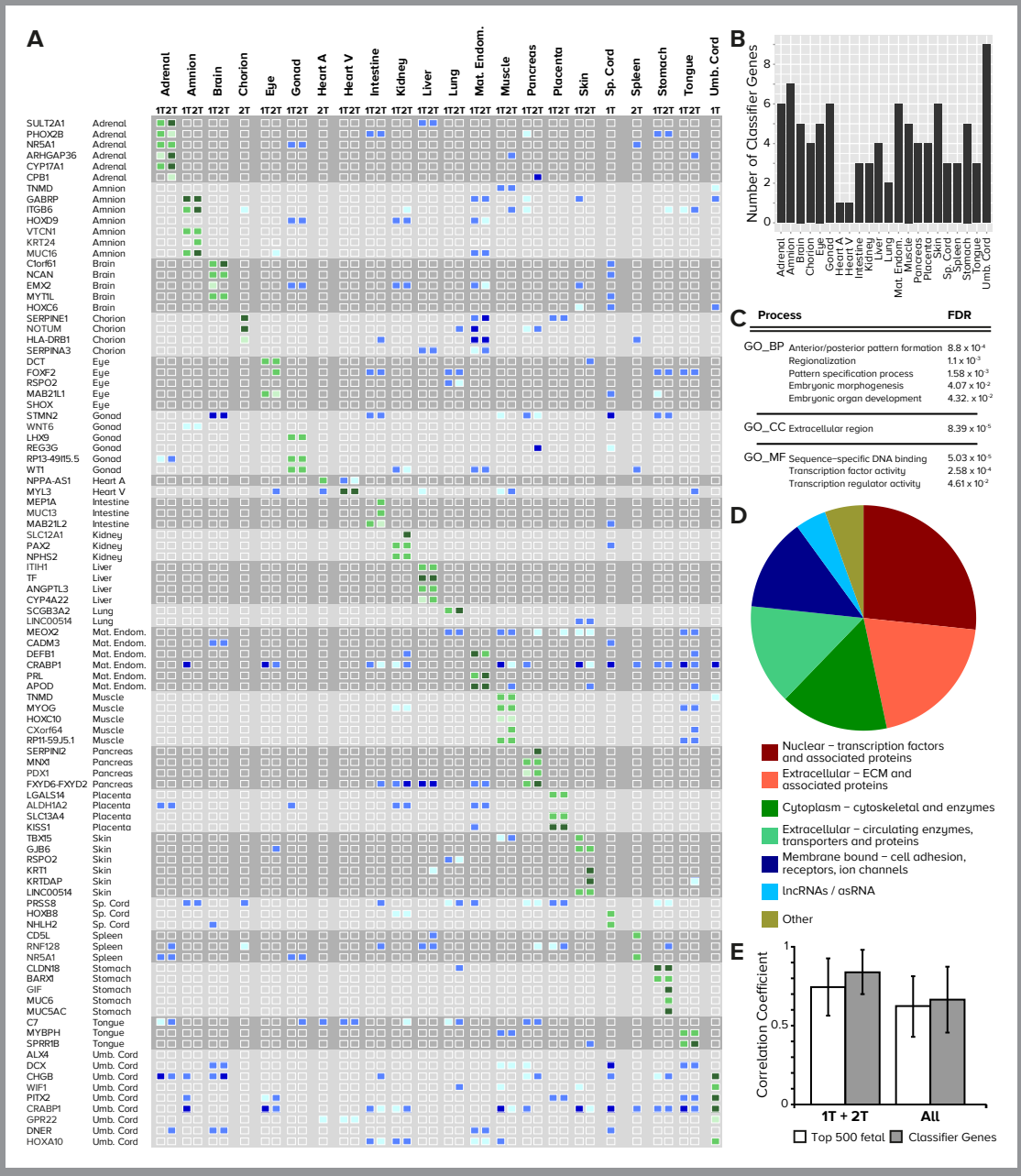
In terms of gene ontology (GO), it is noteworthy that the fetal classifier genes identified were sufficiently related to be significantly enriched (FDR < 0.05) for categories associated with biological processes («patterning», «morphogenesis» and «development»), cellular compartment («extracellular») and molecular function («transcription» and «DNA binding») [Fig. 2C; Supplementary Table 2]**. Furthermore, a substantial proportion of the fetal classifier genes are transcription factors (n = 24)** [Fig. 2D]**, many being tissue-specific and considered master regulator genes in mice** [Supplementary Table 2]**. However, it was striking that about half of the total fetal classifier genes were in fact not directly related to transcription, but were instead associated to extracellular matrix (ECM), cell adhesion and**

surface tension, or are components of the cytoskeleton and cellular transport machinery [Fig. 2D; Supplementary Table 2]**. This underscored the importance of cellular shape, structure and the niche in determining tissue identity and function. Interestingly, a few classifier genes are long non-coding RNAs (lncRNAs) or anti-sense RNAs (asRNAs) (n = 4), including the unique classifier gene for heart atrium *(NPPA-AS1)*. Our data is in agreement with the growing evidence placing lncRNAs as an emerging class of important cellular regulators in development** [34]**.**

Hierarchical clustering (Pearson correlation) of the expression values per organ and trimester (1T + 2T), as well as per organ independent of the trimester (All), resulted in a higher mean correlation coefficient using the 90 classifier genes than the mean correlation coefficient using the 500 most variably expressed genes of the entire fetal dataset (Top 500 fetal) [Fig. 2E]**. Accordingly, the clustering of the fetal samples using the 90 classifier genes was sufficient to group them in a germ layer-, tissue- and age-specific manner** [Supplementary Fig. 1A and 1B]**. Taken together, our results suggested that the classifier genes, identified as the minimal set of genes or barcode necessary to predict the identity of 21 different fetal human organs and the maternal endometrium, represent physiologically relevant genes, many of which are associated with human pathological conditions affecting the organ/tissue they characterize** [Supplementary Table 2]**.**

USING THE FETAL DATA, KEYGENES PREDICTS HUMAN ADULT TISSUE IDENTITY FROM NGS DATA

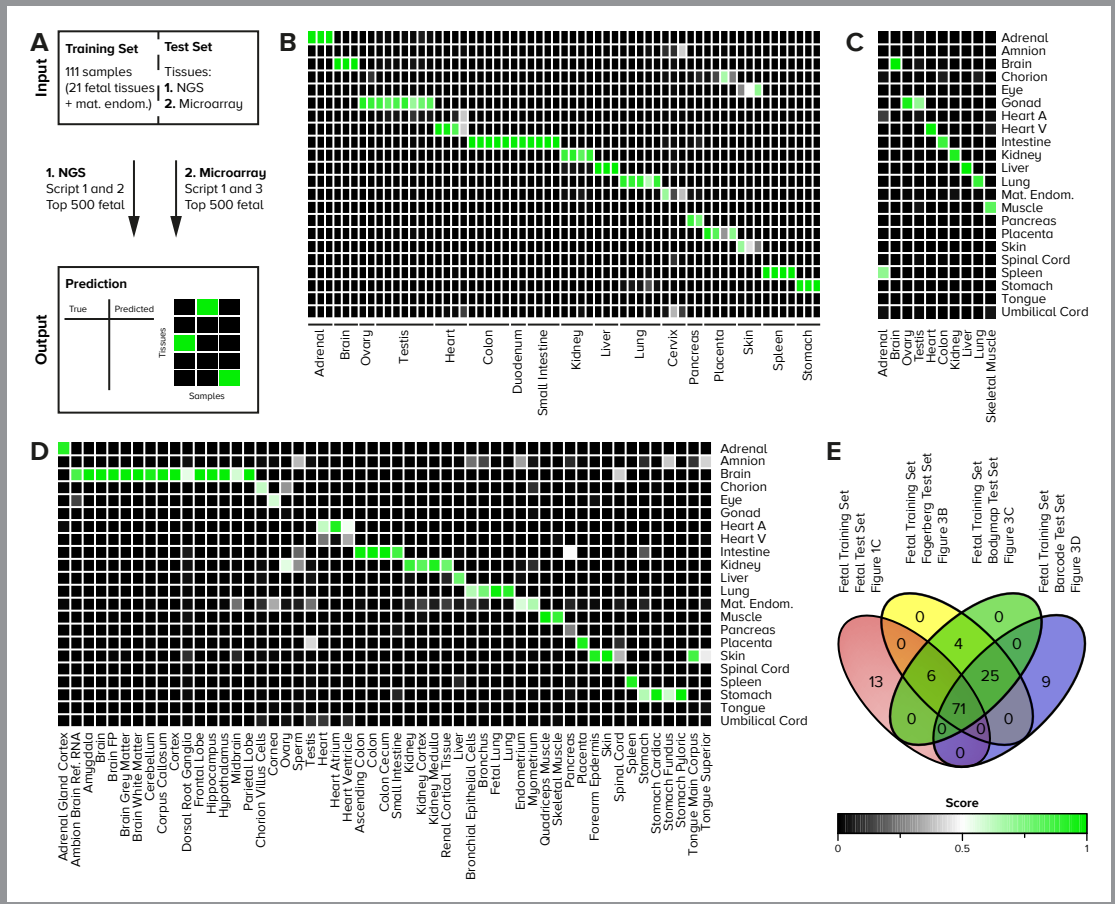
We next investigated whether the fetal transcriptional barcode could be used to predict the identity of adult human tissues. To this end, we applied KeyGenes to two different NGS datasets of human adult organs as test sets, using our complete human fetal dataset (111 samples representing 21 fetal tissues and the maternal endometrium) as the training set [Fig. 3A]


A
Adrenal
Amnion
Brain
Chorion
Eye
Gonad
Heart A
Heart V
Intestine
Kidney
Liver
Lung
Mat. Endom.
Muscle
Pancreas
Placenta
Skin
Sp. Cord
Spleen
Stomach
Tongue
Umb. Cord
1T2T
2T
1T
SULT2A1 Adrenal
PHOX2B Adrenal
NR5A1 Adrenal
ARHGAP36 Adrenal
CYP17A1 Adrenal
CPB1 Adrenal
TNMD Amnion
GABRP Amnion
ITGB6 Amnion
HOXD9 Amnion
VTCN1 Amnion
KRT24 Amnion
MUC16 Amnion
C1orf61 Brain
NCAN Brain
EMX2 Brain
MYT1L Brain
HOXC6 Brain
SERPINE1 Chorion
NOTUM Chorion
HLA-DRB1 Chorion
SERPINA3 Chorion
DCT Eye
FOXF2 Eye
RSPO2 Eye
MAB21L1 Eye
SHOX Eye
STMN2 Gonad
WNT6 Gonad
LHX9 Gonad
REG3G Gonad
RP13-491I5.5 Gonad
WT1 Gonad
NPPA-AS1 Heart A
MYL3 Heart V
MEP1A Intestine
MUC13 Intestine
MAB21L2 Intestine
SLC12A1 Kidney
PAX2 Kidney
NPHS2 Kidney
ITIH1 Liver
TF Liver
ANGPTL3 Liver
CYP4A22 Liver
SCGB3A2 Lung
LINC00514 Lung
MEOX2 Mat. Endom.
CADM3 Mat. Endom.
DEFB1 Mat. Endom.
CRABP1 Mat. Endom.
PRL Mat. Endom.
APOD Mat. Endom.
TNMD Muscle
MYOG Muscle
HOXC10 Muscle
CXorf64 Muscle
RP11-59J5.1 Muscle
SERPINI2 Pancreas
MNX1 Pancreas
PDX1 Pancreas
FXYD6-FXYD2 Pancreas
LGALS14 Placenta
ALDH1A2 Placenta
SLC13A4 Placenta
KISS1 Placenta
TBX15 Skin
GJB6 Skin
RSPO2 Skin
KRT1 Skin
KRTDAP Skin
LINC00514 Skin
PRSS8 Sp. Cord
HOXB8 Sp. Cord
NHLH2 Sp. Cord
CD5L Spleen
RNF128 Spleen
NR5A1 Spleen
CLDN18 Stomach
BARX1 Stomach
GIF Stomach
MUC6 Stomach
MUC5AC Stomach
C7 Tongue
MYBPH Tongue
SPRR1B Tongue
ALX4 Umb. Cord
DCX Umb. Cord
CHGB Umb. Cord
WIF1 Umb. Cord
PITX2 Umb. Cord
CRABP1 Umb. Cord
GPR22 Umb. Cord
DNER Umb. Cord
HOXA10 Umb. Cord
B
Number of Classifier Genes
0
2
4
6
8
Adrenal
Amnion
Brain
Chorion
Eye
Gonad
Heart A
Heart V
Intestine
Kidney
Liver
Lung
Mat. Endom.
Muscle
Pancreas
Placenta
Skin
Sp. Cord
Spleen
Stomach
Tongue
Umb. Cord
C
Process
FDR
GO_BP
Anterior/posterior pattern formation 8.8 x 10-4
Regionalization 1.1 x 10-3
Pattern specification process 1.58 x 10-3
Embryonic morphogenesis 4.07 x 10-2
Embryonic organ development 4.32. x 10-2
GO_CC
Extracellular region 8.39 x 10-5
GO_MF
Sequence–specific DNA binding 5.03 x 10-5
Transcription factor activity 2.58 x 10-4
Transcription regulator activity 4.61 x 10-2
D
Nuclear – transcription factors and associated proteins
Extracellular – ECM and associated proteins
Cytoplasm – cytoskeletal and enzymes
Extracellular – circulating enzymes, transporters and proteins
Membrane bound – cell adhesion, receptors, ion channels
lncRNAs / asRNA
Other
E
Correlation Coefficient
0
0.5
1
1T + 2T
All
Top 500 fetal
Classifier Genes

[Fig. 2]

KeyGenes provides a transcriptional barcode for first and second trimester of human fetal development.

[A] Heatmap of the expression levels of the fetal transcriptional barcode (90 classifier genes) and the organ they represent during first (1T) and second trimester (2T). The expression levels (in counts per million, CPM) of the classifier genes in the corresponding organ/tissue are depicted in green (50-100 CPM, light green; 100-1000 CPM, green; > 1000 CPM, dark green), whereas the expression levels in other than the classifying tissues are displayed in blue (50-100 CPM, light blue; 100-1000 CPM, blue; > 1000 CPM, dark blue). All organs/tissues per trimester show a unique expression pattern of the classifier genes. **[B]** Number of fetal classifier genes per organ/tissue used by KeyGenes to predict identity of the samples in the fetal test set. **[C]** Enrichment of gene ontology (GO) terms for biological processes (BP), cellular compartment (CC) and molecular function (MF) with a false discovery rate (FDR) < 0.05 of the 90 fetal classifier genes. All GO terms are provided in Supplementary Table 2. **[D]** Categorisation of the 90 fetal classifier genes by cellular compartment and molecular function. **[E]** Mean Pearson correlation coefficient (± standard deviation) between samples of the same organ based on the expression levels of the 500 most variably expressed genes (Top 500 fetal) or the 90 fetal classifier genes. The mean correlation coefficient was higher using the 90 fetal classifier genes, both when the trimester was taken into account (1T + 2T) and when samples were considered regardless of the trimester (All). Abbreviations: 1T, first trimester; 2T, second trimester; asRNA, antisense RNA; CPM, counts per million; ECM, extracellular matrix; heart A, heart atrium; heart V, heart ventricle; lncRNA, long non-coding RNA; mat. endom., maternal endometrium; sp. cord, spinal cord; umb. cord, umbilical cord.


A
Input
Training Set
111 samples
(21 fetal tissues
+ mat. endom.)
Test Set
Tissues:
1. NGS
2. Microarray
1. NGS
Script 1 and 2
Top 500 fetal
2. Microarray
Script 1 and 3
Top 500 fetal
Output
Prediction
True
Predicted
Tissues
Samples
B
Adrenal
Brain
Ovary
Testis
Heart
Colon
Duodenum
Small Intestine
Kidney
Liver
Lung
Cervix
Pancreas
Placenta
Skin
Spleen
Stomach
C
Adrenal
Amnion
Brain
Chorion
Eye
Gonad
Heart A
Heart V
Intestine
Kidney
Liver
Lung
Mat. Endom.
Muscle
Pancreas
Placenta
Skin
Spinal Cord
Spleen
Stomach
Tongue
Umbilical Cord
Adrenal
Brain
Ovary
Testis
Heart
Colon
Kidney
Liver
Lung
Skeletal Muscle
D
Adrenal
Amnion
Brain
Chorion
Eye
Gonad
Heart A
Heart V
Intestine
Kidney
Liver
Lung
Mat. Endom.
Muscle
Pancreas
Placenta
Skin
Spinal Cord
Spleen
Stomach
Tongue
Umbilical Cord
Adrenal Gland Cortex
Ambion Brain Ref RNA
Amygdala
Brain FP
Brain Grey Matter
Brain White Matter
Cerebellum
Corpus Callosum
Cortex
Dorsal Root Ganglia
Frontal Lobe
Hippocampus
Hypothalamus
Midbrain
Parietal Lobe
Chorion Villus Cells
Cornea
Ovary
Sperm
Testis
Heart
Heart Atrium
Heart Ventricle
Ascending Colon
Colon
Colon Cecum
Small Intestine
Kidney
Kidney Cortex
Kidney Medulla
Renal Cortical Tissue
Liver
Bronchial Epithelial Cells
Bronchus
Fetal Lung
Lung
Endometrium
Myometrium
Quadriceps Muscle
Skeletal Muscle
Pancreas
Placenta
Forearm Epidermis
Skin
Spinal Cord
Spleen
Stomach
Stomach Cardiac
Stomach Fundus
Stomach Pyloric
Tongue Main Corpus
Tongue Superior
E
Fetal Training Set
Fetal Test Set
Figure 1C
Fetal Training Set
Fagerberg Test Set
Figure 3B
Fetal Training Set
Bodymap Test Set
Figure 3C
Fetal Training Set
Barcode Test Set
Figure 3D
0
0
0
4
0
13
6
25
9
0
71
0
0
0
0
Score
0
0.5
1

[Fig. 3]

KeyGenes classifies human adult organs and tissues based on the fetal transcriptome.
[A] KeyGenes was applied to human adult next-generation sequencing (NGS) and microarray datasets (test sets) using the fetal dataset as training set. (1) For NGS-derived count data, KeyGenes searches for the best classifier genes out of the 500 most variable genes of the fetal dataset (Top 500 fetal). (2) For microarray-derived data, KeyGenes searches for the best classifier genes out of the Top 500 fetal and applies a scaling step to a broad panel of housekeeper genes [26]. **[B–D]** Identity scores of 17 human adult organ/tissue samples from a NGS dataset [22] **[B]**; of ten human adult organ/tissue samples from the NGS dataset from the Illumina Body Map 2.0 **[C]**; and of 53 human adult organ/tissue samples from the microarray dataset in Gene Expression Barcode 3.0 [23] **[D]**. The rows represent the 22 organs/tissues from the fetal training set and the columns depict the samples in the test set. The identity scores range from zero (black) to one (green). The values of all identity scores are given in Supplementary Table 3. **[E]** Venn diagram showing the intersection of the classifier fetal genes used by KeyGenes to identify/predict the different test sets (fetal NGS test set in Fig. 1C; adult NGS test set in Fig. 3B [22]; adult NGS test set from the Illumina Body Map 2.0 in Fig. 3C; adult microarray test set from the Gene Expression Barcode 3.0 in Fig. 3D [23]). The 71 common fetal classifier genes are depicted in Supplementary Table 2. Abbreviations: heart A, heart atrium; heart V, heart ventricle; mat. endom., maternal endometrium; sp. cord, spinal cord; umb. cord, umbilical cord.

and the top 500 most variable fetal genes determined previously [Supplementary Table 3]**.**

KeyGenes was first applied to a NGS dataset consisting of 61 samples representing adult counterparts of 17 organs present in our fetal training set [22]**. Surprisingly, 56 samples (92% of all adult samples) were predicted correctly with a high mean identity score of 0.90** [Fig. 3B; Supplementary Table 3]**. This suggested that the degree of inter-individual variation between biopsies of the same organ is remarkably low. More importantly, adult tissue biopsies of 17 different organs retained significant similarities to fetal organs with respect to the basic transcriptional wiring, independently of cellular heterogeneity in the adult organ** [Fig. 3B]**. Even though five samples (two of skin, one term placenta, two of cervix) were not predicted correctly, their correctly predicted biological replicates showed high identity scores (one skin, three term placenta, one cervix), and the second best score of those misclassified samples pointed to the correct organ/tissue** [Fig. 3B; Supplementary Table 3]**. Further illustrating the predictive power of KeyGenes, three of the four heart samples were classified as «heart ventricles» (mean identity score = 0.93) and were probably ventricular biopsies; whereas one heart sample was assigned a score of 0.39 for «heart atrium» and 0.4 for «heart ventricle», suggesting that this sample may contain both ventricular and atrial heart tissue** [Fig. 3B; Supplementary Table 3]**.**

Additionally, we used a second resource of NGS data from biopsies of ten human adult organs (Human Body Map 2.0, Illumina), which were represented in our fetal training set. From this, KeyGenes predicted nine out of ten samples (90%) correctly with a mean identity score of 0.90 [Fig. 3C; Supplementary Table 3]**. The one misclassified sample was the adrenal gland, which was classified as «spleen» (0.72). However, the previous adult dataset contained three biological replicates of adrenal samples, all correctly classified with a mean identity score of 0.99** [Fig. 3B; Supplementary Table 3]**, suggesting that the adrenal**

gland sample present in the Human Body Map 2.0 by Illumina may in fact not be an adrenal.

We showed that KeyGenes together with the provided fetal dataset is a platform capable of predicting the identity of adult counterparts of fetal tissues with high confidence. Moreover, our results suggested that the minimal transcriptional organ barcode identified during human fetal development is in fact maintained into adulthood, regardless of the increasing cellular complexity of each organ. In addition, we also presented the differentially expressed genes between the first trimester, second trimester and adult for each organ analysed, showing the transcriptional trajectory of maturation of the tissue/organs analysed [Supplementary Fig. 2; Supplementary Table 4]**.**

USING THE FETAL DATA, KEYGENES PREDICTS HUMAN ADULT TISSUE IDENTITY FROM MICROARRAY DATA

Although the use of NGS is of increasing importance, DNA microarray technology is still widely used as it provides fast and relatively inexpensive gene expression data. Because of the fundamentally different nature of data from NGS and the relative (fluorescence) intensity from microarray, we adapted our algorithm to process microarray data [Fig. 3A]**. KeyGenes still used the fetal dataset as training set and the 500 most variably expressed fetal genes determined before (Top 500 fetal)** [Supplementary Table 3]**, but it now contained a scaling step using a broad panel of housekeeping genes** [26]**.**

We applied KeyGenes to 53 samples, mostly human adult organs, tissues and cell subpopulations, available online from the Gene Expression Barcode 3.0 (Affymetrix) [23]**, that were represented in our fetal training set. KeyGenes was able to predict 45 out of 53 samples (85%) correctly with a high mean identity score of 0.86** [Fig. 3D; Supplementary Table 3]**, suggesting a successful adaptation of the algorithm.**

Two of the mismatched organs were the adult ovary and testis, but both tissues were predicted with high confidence as «gonad» before using both NGS datasets [Fig. 3B and 3C]**, suggesting that the probes of important classifier genes for «gonad» may not be well represented on the microarray. Consequently, the sample «sperm» was also not well predicted. It would be interesting to see whether NGS data from this very specialized population of cells would still be identified as «gonad». Another misclassified organ was the pancreas, which was predicted as «intestine», but was predicted correctly using NGS data** [Fig. 3B]**, suggesting that the probes for pancreatic classifier genes were probably not well represented on the microarray. The sample «heart ventricle» had a very similar mixed «atrium and ventricle» prediction to one of the NGS heart biopsies** [Fig. 3B]**. Finally, the two tongue muscles included in the microarray datasets were identified as «skin» instead of «tongue» or «muscle»; and the only spinal cord sample was identified as «brain» instead of «spinal cord». As we were unable to find available NGS data from both adult tongue and spinal cord, we could not conclude whether the mismatch is due to a lack of representation of classifier genes on the microarray or a genuine mismatch between the expression profile between the fetal and adult organ.**

It is noteworthy that microarray data from specific subpopulations of cells or tissues of the adult organs were also assigned correctly to the tissue of origin. The 15 samples from different anatomical regions related to «brain» were all classified correctly and sub-regions of the intestine (four sub-regions), kidney (four sub-regions) and stomach (four sub-regions) were all also assigned to the correct main organ [Fig. 3D]**. Moreover, the sample adult lung and fetal lung were predicted as «lung» with an identity score of 0.91 and 0.98, respectively, whereas the bronchus and the bronchial epithelial cells had a «lung» prediction but a lower identity score (0.77 and 0.65, respectively), suggesting a better match with the full tissue than**

when using sub-populations [Fig. 3D; Supplementary Table 3]**.**

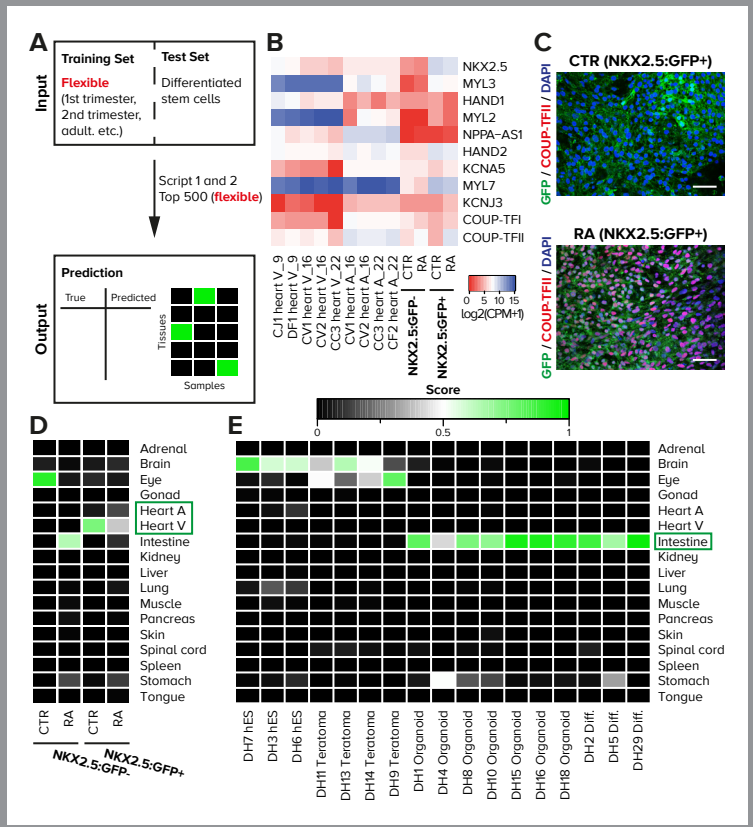
Furthermore, we analysed predictions in a second microarray dataset from a different platform (Illumina) [24]**. KeyGenes was capable of predicting 18 of 21 relevant samples (86%) correctly with a relatively high mean identity score of 0.6** [Supplementary Fig. 3; Supplementary Table 3]**. The mean identity score of the correctly predicted fetal tissues (n = 12) was 0.67, whereas it was 0.52 for the adult tissues (n = 7)** [Supplementary Fig. 3; Supplementary Table 3]**.**

In summary, our algorithm is not only robust enough for NGS or microarray data, but it is also capable of assigning correct organ predictions to datasets from specific parts of adult human organs, suggesting that each organ's transcriptional basic wiring or barcode identified by KeyGenes from the fetal set remains stable until adulthood, at least with respect to the 18 main adult organs analysed. The intersection of the fetal classifier genes obtained for each of the four predictions («fetal from fetal» and «adult from fetal») resulted in 71 common genes, suggesting that these may be the most relevant genes (represented as well on the microarray platform) to characterize the basic wiring of both human fetal and adult organs [Fig. 3E; Supplementary Table 2]**.**

MONITORING DIFFERENTIATION EFFICIENCY FROM HPSCS WITH KEYGENES

One challenging issue in stem cell research is still how to reliably determine the identity and extent of differentiation of stem cells towards a specific cell or tissue type. To determine whether KeyGenes was suitable for this purpose, we used it on multiple NGS datasets that represented hPSCs differentiated to four different tissue/cell types. We adapted both the fetal training set and calculated the top 500 most variably expressed genes now excluding the extraembryonic fetal tissues and the maternal endometrium [Fig. 4A; Supplementary Table 3]**.**

First, we generated atrial-like and ventricular-like


A
Input
Training Set
Flexible
(1st trimester, 2nd trimester, adult. etc.)
Test Set
Differentiated stem cells
Script 1 and 2
Top 500 (flexible)
Output
Prediction
True
Predicted
Tissues
Samples
B
NKX2.5
MYL3
HAND1
MYL2
NPPA-AS1
HAND2
KCNA5
MYL7
KCNJ3
COUP-TFI
COUP-TFII
CJ1 heart V_9
DF1 heart V_9
CV1 heart V_16
CV2 heart V_16
CC3 heart V_22
CV1 heart A_16
CV2 heart A_16
CC3 heart A_22
CF2 heart A_22
CTR
RA
CTR
RA
NKX2.5:GFP-
NKX2.5:GFP+
0 5 10 15
log2(CPM+1)
C
CTR (NKX2.5:GFP+)
RA (NKX2.5:GFP+)
GFP / COUP-TFII / DAPI
D
Adrenal
Brain
Eye
Gonad
Heart A
Heart V
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Skin
Spinal cord
Spleen
Stomach
Tongue
CTR
RA
CTR
RA
NKX2.5:GFP-
NKX2.5:GFP+
Score
0
0.5
1
E
DH7 hES
DH3 hES
DH6 hES
DH11 Teratoma
DH13 Teratoma
DH14 Teratoma
DH9 Teratoma
DH1 Organoid
DH4 Organoid
DH8 Organoid
DH10 Organoid
DH15 Organoid
DH16 Organoid
DH18 Organoid
DH2 Diff.
DH5 Diff.
DH29 Diff.

[Fig. 4]

KeyGenes to monitor human pluripotent stem cell differentiation.

[A] KeyGenes was applied to next-generation sequencing (NGS) datasets from differentiated derivatives of human pluripotent stem cells (hPSCs) (test set), using flexible training sets. The selection of the classifier genes is based on the 500 most variable genes across a given dataset and can be modified. **[B]** Heatmap visualizing the binary logarithm of the expression levels (counts per million, CPM) of selected atrial and ventricular marker genes in the human fetal heart atrium (A) and ventricle (V) samples and in the control (CTR) and retinoic acid (RA) differentiated NKX2.5:GFP-positive and -negative cells. **[C]** Immunofluorescence for COUP-TFII in the control (CTR) and retinoic acid (RA) differentiated NKX2.5:GFP-positive subpopulation. Scale bar: 40 µm. **[D–E]** Identity scores of the control (CTR) and retinoic acid (RA)-treated NKX2.5:GFP-positive and -negative cells **[D]**; of human embryonic stem cell (hESC) samples, teratomas, intestinal organoids derived from hESCs, and organoids that have been differentiated further towards intestine (Diff.) [20] **[E]**. The fetal dataset was used as training set excluding the extraembryonic tissues and the maternal endometrium samples. The selection of the classifier genes is based on the 500 most variable genes of the embryonic training set without extraembryonic and maternal endometrium samples (Top 500 fetal w/o). The rows represent the 17 organs/tissues from the fetal training set and the columns depict the samples in the test set. The identity scores range from zero (black) to one (green). The values of all identity scores are given in Supplementary Table 3. Abbreviations: CTR, control; heart A, heart atrium; heart V, heart ventricle; RA, retinoic acid; sp. cord, spinal cord.

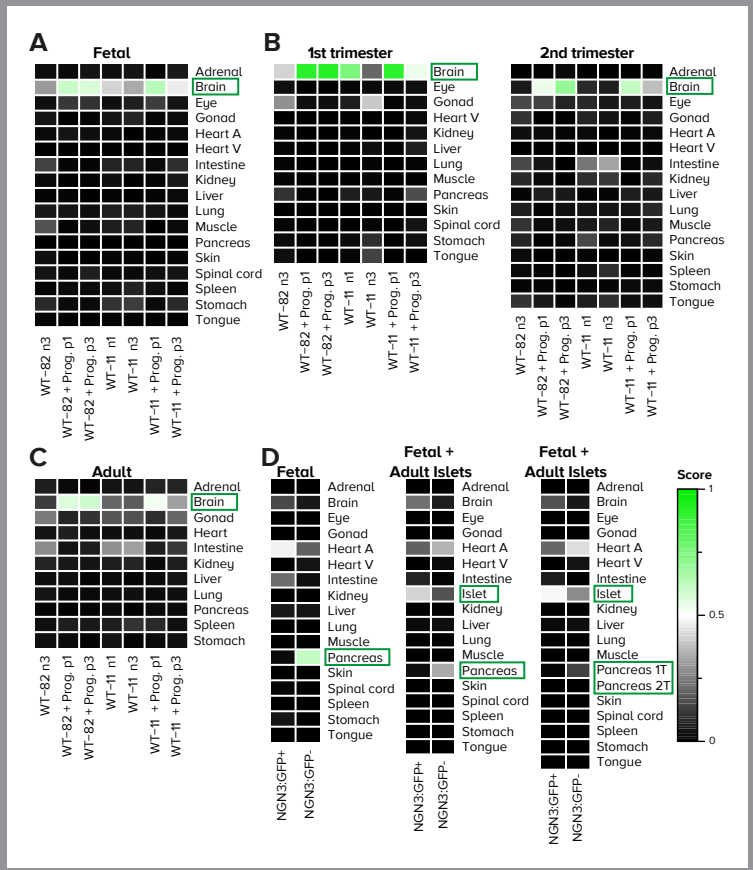
cardiomyocytes [Supplementary Fig. 4A] **from a NKX2.5:GFP human embryonic stem cell (hESC) reporter line** [35] **and assessed their signature with KeyGenes. Cardiac differentiation of hESCs resulted predominantly in GFP-positive cells with ventricular-like identity, while exogenous treatment with retinoic acid (RA) during differentiation resulted in GFP-positive cells with an atrial-like identity as shown by the upregulation of atrial markers *(COUP-TFII)* and downregulation of ventricular markers *(MYL3)*, much like the atria and ventricles of the fetal heart** [Fig. 4B] [33]**. Furthermore, atrial identity of RA-treated cardiomyocytes was also demonstrated at the protein level by co-expression of the atrial-marker COUP-TFII and GFP** [Fig. 4C]**. Using KeyGenes, the control NKX2.5:GFP-positive hESC cardiomyocytes showed a high identity score with heart ventricle** [Fig. 4D]**, whilst in the RA-treated group, NKX2.5:GFP-positive hESC cardiomyocytes increased their identity score for «heart atrium» (from 0.06 to 0.15), although these cells were still predicted to be «heart ventricle» but with a lower identity score (reduced from 0.75 to 0.40)** [Supplementary Table 3]**. Interestingly, in the fetal barcode matrix, RA-treated NKX2.5:GFP-positive hESC cardiomyocytes show similar *MYL3* expression levels as atrial cardiomyocytes, but lack expression of *NPPA-AS1*** [Fig. 4B; Supplementary Fig. 5]**.**

Next, we used a dataset from LGR5:GFP reporter hESCs that were first induced to form teratomas in mice [20] [Supplementary Fig. 4B]**; from the teratomas, «adult intestinal stem cell»-like cells were FACS sorted on the basis of GFP fluorescence intensity and then induced to form organoids resembling intestinal tissue before being differentiated further. We analysed the NGS data from different time points during the differentiation. Using our fetal dataset as training set, nine out of ten organoids (organoids and differentiated organoids) were classified as «intestine» with a mean identity score 0.85** [Fig. 4E; Supplementary Table 3]**. We also included the NGS data of undifferentiated hESCs in the analysis and, interest-**

ingly, the highest identity score was for «brain», suggesting that this ectodermal tissue during fetal development remains relatively immature. Furthermore, KeyGenes detected a large difference between different organoids, but not a large difference in maturation levels between the organoids and their differentiated counterparts underlining the ability of KeyGenes to monitor differentiation and indicate developmental stage equivalents.

ASSIGNING DEVELOPMENTAL STAGES TO DIFFERENTIATED HPSC DERIVATIVES WITH KEYGENES

In order to examine the broader applicability of KeyGenes in assigning developmental stages to differentiated hPSC derivatives, we analysed NGS data from dopaminergic neurons derived by the differentiation of human induced pluripotent stem cells (hiPSCs) from old (82 years) and young (11 years) donors [19]**. These dopaminergic neurons were then either transfected with GFP-progerin, a protein involved in premature aging, or nuclear GFP as control** [Supplementary Fig. 4C]**. These aged neurons are regarded as being possibly useful as a platform to investigate late-onset diseases such as Parkinson's disease. All seven samples analysed were classified as «brain» with a mean score of 0.47** [Fig. 5A; Supplementary Table 3]**. KeyGenes confirmed that the aging of the neurons by progerin had a clear impact on differentiation, always assigning a higher identity score for «brain» to progerin-transfected neurons (mean score of 0.57) than to controls (mean score of 0.35)** [Fig. 5A; Supplementary Table 3]**. When the fetal training set was divided into first and second trimester** [Fig. 5B] **with the corresponding top 500 most variably expressed genes, we observed that the samples had higher identity scores with first trimester «brain» tissue than with second trimester «brain» tissue. Moreover, using adult samples in the training set resulted in lower identity scores** [Fig. 5C] **than when either the fetal training set or the first and second trimester samples were used separately.**


A
Fetal
Adrenal
Brain
Eye
Gonad
Heart A
Heart V
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Skin
Spinal cord
Spleen
Stomach
Tongue
WT−82 n3
WT−82 + Prog. p1
WT−82 + Prog. p3
WT−11 n1
WT−11 n3
WT−11 + Prog. p1
WT−11 + Prog. p3
B
1st trimester
2nd trimester
C
Adult
D
Fetal
Fetal + Adult Islets
Fetal + Adult Islets
Score
1
0.5
0
NGN3:GFP+
NGN3:GFP-
Islet
Pancreas 1T
Pancreas 2T

[Fig. 5]

KeyGenes can be modified to assign developmental stages.

[A–C] Identity scores of dopaminergic neurons derived from two different human induced pluripotent stem cell (hiPSC) lines WT-82 and WT-11 [19]. The dopaminergic neurons were further differentiated by transfection with progerin-GFP (Prog.) or nuclear GFP as a control. As training set, KeyGenes used either the whole fetal dataset excluding the extraembryonic tissues and the maternal endometrium samples **[A]**, only the first trimester or second trimester samples excluding the extraembryonic tissues and the maternal endometrium samples **[B]**, or the adult NGS datasets ([22] and Illumina Bodymap) **[C]** with the corresponding 500 most variable genes (Top 500 fetal w/o, Top 500 fetal w/o 1T, Top 500 fetal w/o 2T, Top 500 adult). **[D]** Identity scores of NGN3:GFP-positive and NGN3:GFP-negative cells derived from human embryonic stem cells (hESCs) [18]. The fetal training set excluding the extraembryonic tissues and the maternal endometrium samples was expanded by a NGS dataset consisting of five human adult islet of Langerhans samples [21]. The 500 most variable genes of the embryonic training set without extraembryonic and maternal endometrium samples (Top 500 fetal w/o) were used. The rows represent the organs/tissues from the different training sets and the columns depict the samples in the test set. The identity scores range from zero (black) to one (green). The values of all identity scores are given in Supplementary Table 3. Abbreviations: heart A, heart atrium; heart V, heart ventricle; RA, retinoic acid; sp. cord, spinal cord.

This suggested that, even though the progerin-samples are predicted as being «brain», the identity score compared with adult brain is even lower. Using each training set, we noted that there were evident differences in identity scores between cells that had undergone different numbers of passage in culture [Fig. 5A–5C]**. This showed how KeyGenes can be used to monitor differentiation conditions as well as to compare replicates to improve protocol outcomes.**

Finally, we applied KeyGenes to NGS data from pancreatic endocrine progenitor cells differentiated from NGN3:eGFP hESCs [18] [Supplementary Fig. 4D]**. *NGN3* is an essential transcription factor during pancreas development, specifying the fate of its endocrine cells** [36]**. A multi-step differentiation protocol was used to differentiate NGN3:GFP hESCs towards pancreatic β-cell lineage; cells were then FACS sorted for GFP and analysed at the endocrine progenitor stage** [Supplementary Fig. 4D]**. The NGN3:GFP-negative differentiated population was classified as «pancreas» using the fetal training set, confirming the presence of pancreatic exocrine (progenitor) cells** [Fig. 5D; Supplementary Table 3]**. Interestingly, the NGN3:GFP-positive differentiated population, presumably containing the endocrine progenitors, was misclassified as «heart atrium»** [Fig. 5D; Supplementary Table 3]**. However, large proportions of the pancreas are in fact exocrine cells and the endocrine cells contribute only to 1–2% to the pancreatic cell mass** [37, 38]**. This also explains why the classifier genes for the (fetal) pancreas are mainly genes involved in exocrine function of the pancreas** [Fig. 2A; Supplementary Table 2]**.**

Therefore, to have meaningful predictions with respect to the endocrine lineage (NGN3:GFP-positive differentiated population), the fetal training data used by KeyGenes was enriched by an available NGS dataset consisting of five human adult pancreatic islets [21]**. Using the 500 most variably expressed genes across our fetal training set excluding the extraembryonic fetal tissues and the maternal endometrium** [Supple-

mentary Table 3]**, the misclassification of the NGN3:GFP-positive cell population disappeared and was predicted as «islet» with a score of 0.41** [Fig. 5D; Supplementary Table 3]**. Additionally, by separating first and second trimester pancreas (1T and 2T) in the training set, the identity score for the NGN3:GFP-positive cell population increased for «islet» to 0.49** [Fig. 5D; Supplementary Table 3]**. This highlights the importance of being able to implement the appropriate training set, depending on the purpose of each differentiation experiment, to be able to assign meaningful identity scores and determine the equivalent developmental stage (either by using a different training set for each stage or by separating the developmental stages of the tissue/cells of interest in the same training set) to improve the outcomes of the differentiation protocols.**

DISCUSSION

We have shown here through detailed genomic analysis that human organs and tissues retain a transcriptional signature from W9 until adulthood even though each organ is composed of multiple progenitor cell types that mature over time. It was remarkable that the transcriptional expression profile of a set of less than 100 genes was sufficient to identify 21 different human fetal organs/tissues (plus maternal endometrium) and 18 adult human organs. These classifier genes that we identified not only included genes involved in transcription regulation but genes that define cellular shape and metabolism. Moreover, some of the classifier genes were long non-coding and antisense RNAs, highlighting the regulatory importance of this class of genes [34]**. This was notably illustrated by *NPPA-AS1*, which is thought to regulate *NPPA* expression, a gene that encodes atrial natriuretic factor and is involved in heart development and chamber specification** [39, 40]**. This also underscores one of the advantages of KeyGenes, developed to compare data to a NGS (fetal) training set, in contrast to the existent algorithms** [9–11] **that compare data to networks deduced from microarray datasets, which contain a fixed set of probes with low representation of non-coding RNAs such as lncRNAs and antisense RNAs. In our case, 3 out of the 4 lncRNAs identified as fetal classifier genes *(RP13-49I15.5, NPPA-AS1, LINC00514)* were not present in the microarray adult dataset, but present in the NGS adult dataset and identified there as well as classifier genes for predictions.**

The transcriptional human fetal atlas dataset presented here, even though limited in number of samples and organs analysed, is a unique resource that will provide deeper understanding of the signalling cascades and molecular dynamics during human development that lead into the matura-

tion of progenitor cells within each human organ. The human fetal NGS dataset was paramount to the development and validation of our prediction algorithm KeyGenes and has proven sufficient as training set to identify both human adult organs (using NGS and microarray data) as well as several differentiated derivatives of hPSCs. All hPSC derivatives expressed genes identified as fetal classifier genes (and many helper classifier genes) of the specific tissue to which they were claimed to have differentiated [Supplementary Fig. 5]**. Comparing differentiated derivatives of hPSCs to human adult gene expression data is important but can be misleading, as it compares immature cells with far later stages of development when entirely different physiological parameters have affected cell behaviour. Our fetal dataset, used alone or in combination with, for example, adult data as a training set, provided an efficient way to assess progression of hPSC differentiation.**

One important feature of the algorithm is that the ever growing number of NGS datasets available online, for example of different organs, (FACS-sorted) cell types or simply more biological replicates, can easily be incorporated in the training set as we have demonstrated. The predictive possibilities can therefore be extended not only to organs but, as exemplified here, also to tissues (intestinal organoids) and specific cell types (cardiomyocytes, dopaminergic neurons and pancreatic endocrine cells). These will become even more refined as more datasets become available and incorporated.

We have described and validated a valuable resource to the community that will help to determine genes important for identifying cell types and their stages of development, so that protocols for enhancing lineage differentiation efficiency or cell maturation will have accurate benchmarks to monitor the process. In addition, by incorporating similar data from human fetuses with congenital defects or derivatives of diseased patient-derived hiPSCs, it may be easier to identify the underlying molecular mechanism for the pathology.

SUPPLEMENTARY INFORMATION

The Supplementary Tables in this chapter can be found online at http://dx.doi.org/10.1016/j.stemcr.2015.05.002.

ACCESSION NUMBERS

The accession numbers for the data from the fetal organs and the cardiac differentiation reported in this paper are GEO: GSE66302 and GSE67866, respectively.

AUTHOR CONTRIBUTIONS

M.S.R., C.L.M., R.P., S.C.d.S.L. conceived the study. M.S.R., L.v.I., Y.A., H.B., W.A., J.J.G., H.D.D., E.W.v.Z., S.C.d.S.L. conducted experiments and/or performed bioinformatic analysis. All authors were involved in the analysis of the data. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We would like to thank H. Locher for help with the fetal dissection, M. Gomes Fernandes for the sex genotyping, G.M. de Roo for FACS sorting the NKX2.5:GFP cells, S.A.J. de Zeeuw for bioinformatics support, and the Centre for Contraception, Abortion and Sexuality (CASA) in Leiden and The Hague for the collection of the human fetal material. S.C.d.S.L. is supported by the Netherlands Organisation for Scientific Research (NWO, ASPASIA 015.007.037) and the Interuniversity Attraction Poles (IAP, P7/07); M.S.R., C.L.M., F.C. by the Bontius Stichting [PANCREAS]; and E.J.P.d.K. by Stichting DON.

REFERENCES

[1] Ju W, Greene CS, Eichinger F, Nair V, Hodgin JB, Bitzer M, Lee YS, Zhu Q, Kehata M, Li M *et al*: **Defining cell-type specificity at the transcriptional level in human disease**. *Genome research* 2013, **23**(11):1862–1873.

[2] Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K *et al*: **A public database for gene expression in human cancers**. *Cancer research* 1999, **59**(21):5403–5407.

[3] Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D *et al*: **Epigenomic analysis of multilineage differentiation of human embryonic stem cells**. *Cell* 2013, **153**(5):1134–1148.

[4] Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R *et al*: **Transcriptional and epigenetic dynamics during specification of human embryonic stem cells**. *Cell* 2013, **153**(5):1149–1163.

[5] Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH *et al*: **Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines**. *Cell* 2011, **144**(3):439–452.

[6] Fung FK, Chan DW, Liu VW, Leung TH, Cheung AN, Ngan HY: **Increased expression of PITX2 transcription factor contributes to ovarian cancer progression**. *PloS one* 2012, **7**(5):e37076.

[7] Lin Q, Aihara A, Chung W, Li Y, Huang Z, Chen X, Weng S, Carlson RI, Wands JR, Dong X: **LRH1 as a driving factor in pancreatic cancer growth**. *Cancer letters* 2014, **345**(1):85–90.

[8] Hoshijima M, Chien KR: **Mixed signals in heart failure: cancer rules**. *The Journal of clinical investigation* 2002, **109**(7):849–855.

[9] Hwang PI, Wu HB, Wang CD, Lin BL, Chen CT, Yuan S, Wu G, Li KC: **Tissue-specific gene expression templates for accurate molecular characterization of the normal physiological states of multiple**

human tissues with implication in development and cancer studies. *BMC genomics* 2011, **12**:439.

[10] Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ: **CellNet: Network Biology Applied to Stem Cell Engineering**. *Cell* 2014, **158**(4):903–915.

[11] Morris SA, Cahan P, Li H, Zhao AM, San Roman AK, Shivdasani RA, Collins JJ, Daley GQ: **Dissecting Engineered Cell Types and Enhancing Cell Fate Conversion via CellNet**. *Cell* 2014, **158**(4):889–902.

[12] Patterson M, Chan DN, Ha I, Case D, Cui Y, Van Handel B, Mikkola HK, Lowry WE: **Defining the nature of human pluripotent stem cell progeny**. *Cell research* 2012, **22**(1):178–193.

[13] Hrvatin S, O'Donnell CW, Deng F, Millman JR, Pagliuca FW, DiIorio P, Rezania A, Gifford DK, Melton DA: **Differentiated human stem cells resemble fetal, not adult, beta cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**(8):3038–3043.

[14] Nakahori Y, Hamano K, Iwaya M, Nakagome Y: **Sex identification by polymerase chain reaction using X-Y homologous primer**. *American journal of medical genetics* 1991, **39**(4):472–473.

[15] Heeren A, van Iperen L, Klootwijk DB, de Melo Bernardo A, Roost MS, Gomes Fernandes MM, Louwe LA, Hilders CG, Helmerhorst FM, van der Westerlaken LA *et al*: **Development of the follicular basement membrane during human gametogenesis and early folliculogenesis**. *BMC developmental biology* 2015, **15**(1):4.

[16] Mastrokolias A, den Dunnen JT, van Ommen GB, t Hoen PA, van Roon-Mom WM: **Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA**. *BMC genomics* 2012, **13**:28.

[17] Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A: **Transcriptome analysis by strand-specific sequencing of complementary DNA**. *Nucleic acids research* 2009, **37**(18):e123.

[18] Liu H, Yang H, Zhu D, Sui X, Li J, Liang Z, Xu L, Chen Z, Yao A, Zhang L *et al*: **Systematically labeling developmental stage-specific**

genes for the study of pancreatic beta-cell differentiation from human embryonic stem cells. *Cell research* 2014, **24**(10):1181–1200.

[19] Miller JD, Ganat YM, Kishinevsky S, Bowman RL, Liu B, Tu EY, Mandal PK, Vera E, Shim JW, Kriks S *et al*: **Human iPSC-based modeling of late-onset disease via progerin-induced aging**. *Cell stem cell* 2013, **13**(6):691–705.

[20] Forster R, Chiba K, Schaeffer L, Regalado SG, Lai CS, Gao Q, Kiani S, Farin HF, Clevers H, Cost GJ *et al*: **Human intestinal tissue with adult stem cell properties derived from pluripotent stem cells**. *Stem cell reports* 2014, **2**(6):838–852.

[21] Cnop M, Abdulkarim B, Bottu G, Cunha DA, Igoillo-Esteve M, Masini M, Turatsinze JV, Griebel T, Villate O, Santin I *et al*: **RNA sequencing identifies dysregulation of the human pancreatic islet transcriptome by the saturated fatty acid palmitate**. *Diabetes* 2014, **63**(6):1978–1993.

[22] Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K *et al*: **Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics**. *Molecular & cellular proteomics : MCP* 2014, **13**(2):397–406.

[23] McCall MN, Jaffee HA, Zelisko SJ, Sinha N, Hooiveld G, Irizarry RA, Zilliox MJ: **The Gene Expression Barcode 3.0: improved data processing and mining tools**. *Nucleic acids research* 2014, **42**(Database issue):D938–943.

[24] Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Muller FJ, Wang YC, Boscolo FS *et al*: **Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives**. *Cell stem cell* 2012, **10**(5):620–634.

[25] Friedman J, Hastie T, Tibshirani R: **Regularization Paths for Generalized Linear Models via Coordinate Descent**. *Journal of statistical software* 2010, **33**(1):1–22.

[26] Eisenberg E, Levanon EY: **Human housekeeping genes, revisited**. *Trends in genetics : TIG* 2013, **29**(10):569–574.

[27] Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor**

package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010, **26**(1):139–140.

[28] Robinson MD, Smyth GK: **Small-sample estimation of negative binomial dispersion, with applications to SAGE data**. *Biostatistics* 2008, **9**(2):321–332.

[29] Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nature protocols* 2009, **4**(1):44–57.

[30] Bardou P, Mariette J, Escudie F, Djemiel C, Klopp C: **jvenn: an interactive Venn diagram viewer**. *BMC Bioinformatics* 2014, **15**:293.

[31] Wickham H: **ggplot2: Elegant Graphics for Data Analysis**: Springer New York; 2009.

[32] Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S *et al*: **gplots: Various R programming tools for plotting data**. *R package version* 2014, **2.14.2.**

[33] Devalla HD, Schwach V, Ford JW, Milnes JT, El-Haou S, Jackson C, Gkatzis K, Elliott DA, Chuva de Sousa Lopes SM, Mummery CL *et al*: **Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology**. *EMBO Mol Med* 2015, **7**(4):394–410.

[34] Washietl S, Kellis M, Garber M: **Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals**. *Genome research* 2014, **24**(4):616–628.

[35] Elliott DA, Braam SR, Koutsis K, Ng ES, Jenny R, Lagerqvist EL, Biben C, Hatzistavrou T, Hirst CE, Yu QC *et al*: **NKX2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes**.
Nature methods 2011, **8**(12):1037–1040.

[36] Rukstalis JM, Habener JF: **Neurogenin3: a master regulator of pancreatic islet differentiation and regeneration**. *Islets* 2009, **1**(3):177–184.

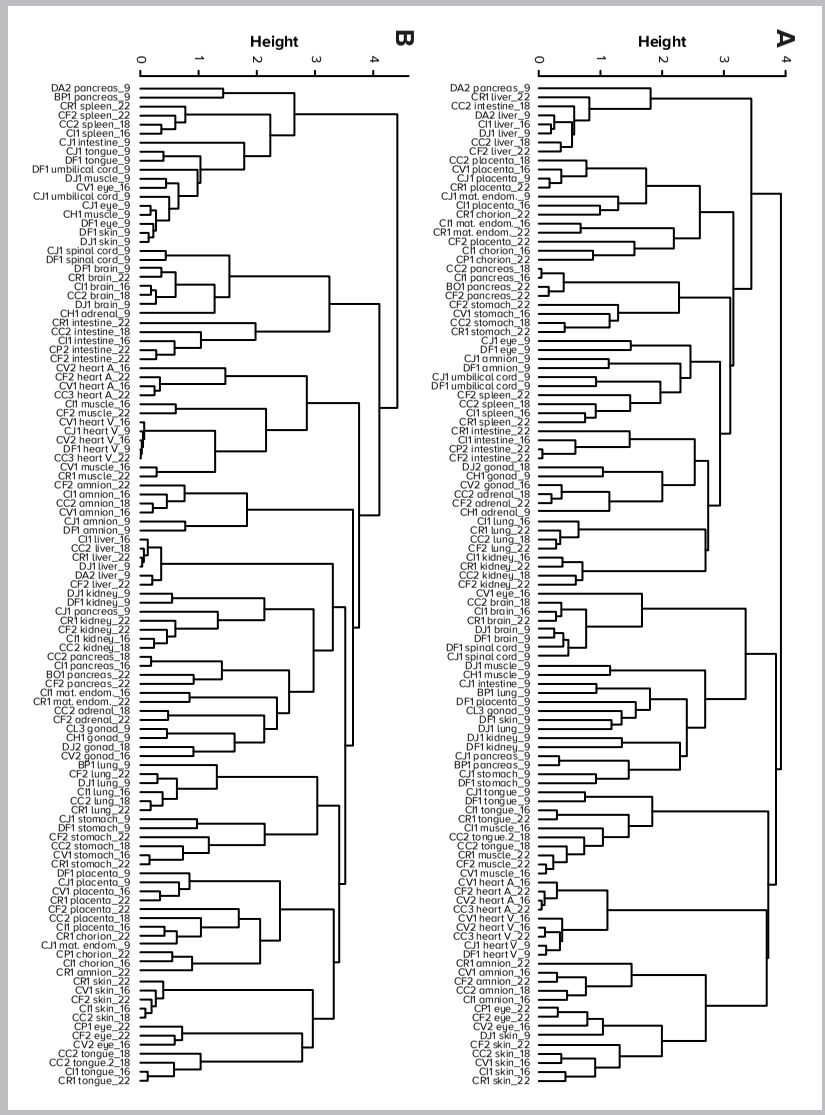
[37] Chu K, Nemoz-Gaillard E, Tsai MJ: **BETA2 and pancreatic islet development**. *Recent progress in hormone research* 2001, **56**:23–46.

[38] Rahier J, Wallon J, Henquin JC: **Cell populations in the endo-**

crine pancreas of human neonates and infants. *Diabetologia* 1981, **20**(5):540–546.

[39] Houweling AC, van Borren MM, Moorman AF, Christoffels VM: **Expression and regulation of the atrial natriuretic factor encoding gene Nppa during development and disease**. *Cardiovascular research* 2005, **67**(4):583–593.

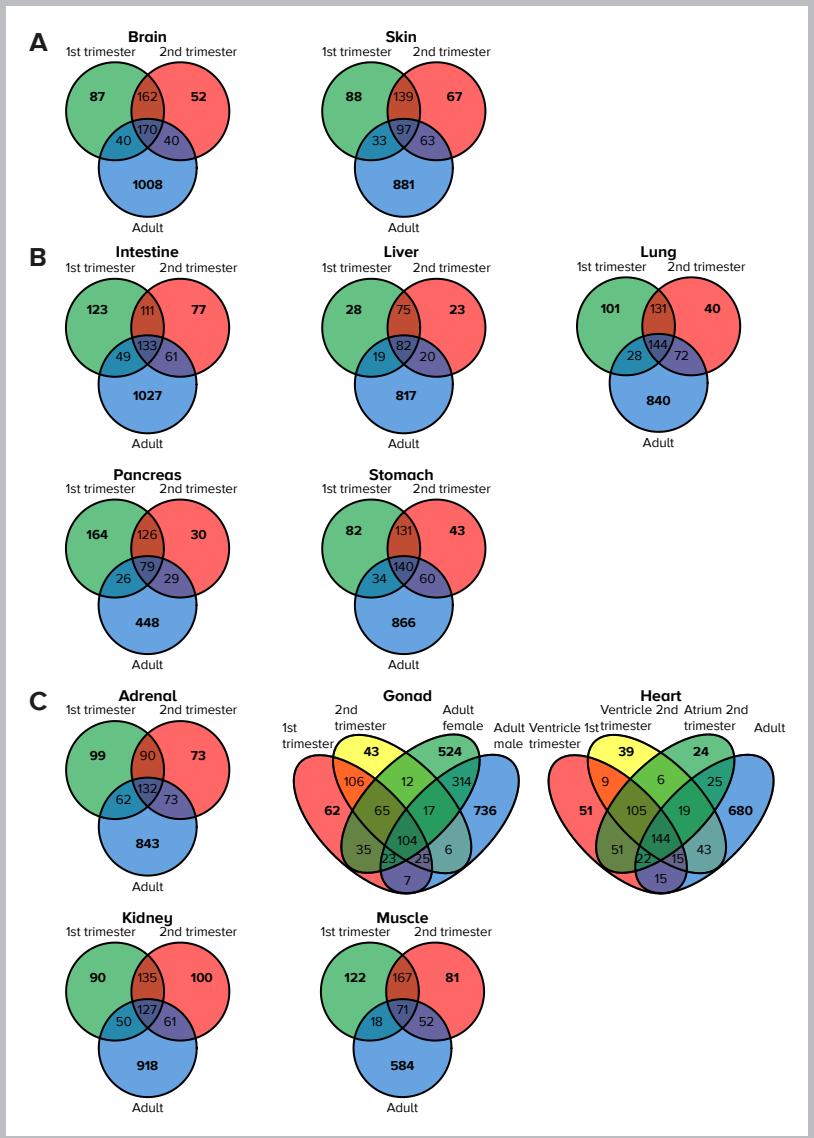
[40] Annilo T, Kepp K, Laan M: **Natural antisense transcript of natriuretic peptide precursor A (NPPA): structural organization and modulation of NPPA expression**. *BMC molecular biology* 2009, **10**:81.


A
Height
0
1
2
3
4
DA2 pancreas_9
CR1 liver_22
CC2 intestine_18
DA2 liver_9
CI1 liver_16
DJ1 liver_9
CC2 liver_18
CF2 liver_22
CC2 placenta_18
CV1 placenta_16
CJ1 placenta_9
CR1 placenta_22
CJ1 mat. endom._9
CI1 placenta_16
CR1 chorion_22
CI1 mat. endom._16
CR1 mat. endom._22
CF2 placenta_22
CI1 chorion_16
CP1 chorion_22
CC2 pancreas_18
CI1 pancreas_16
BO1 pancreas_22
CF2 pancreas_22
CF2 stomach_22
CV1 stomach_16
CC2 stomach_18
CR1 stomach_22
CJ1 eye_9
DF1 eye_9
CJ1 amnion_9
DF1 amnion_9
CJ1 umbilical cord_9
DF1 umbilical cord_9
CF2 spleen_22
CC2 spleen_18
CI1 spleen_16
CR1 spleen_22
CR1 intestine_22
CI1 intestine_16
CF2 intestine_22
CF2 intestine_22
DJ2 gonad_18
CH1 gonad_9
CV2 gonad_16
CC2 adrenal_18
CF2 adrenal_22
CH1 adrenal_9
CI1 lung_16
CR1 lung_22
CC2 lung_18
CF2 lung_22
CI1 kidney_16
CR1 kidney_22
CC2 kidney_18
CF2 kidney_22
CV1 eye_16
CC2 brain_18
CI1 brain_16
CR1 brain_22
DJ1 brain_9
DF1 brain_9
DF1 spinal cord_9
CJ1 spinal cord_9
DJ1 muscle_9
CH1 muscle_9
CJ1 intestine_9
BP1 lung_9
DF1 placenta_9
CL3 gonad_9
DF1 skin_9
DJ1 lung_9
DJ1 kidney_9
DF1 kidney_9
CJ1 pancreas_9
BP1 pancreas_9
CJ1 stomach_9
DF1 stomach_9
CJ1 tongue_9
DF1 tongue_9
CI1 tongue_16
CR1 tongue_22
CI1 muscle_16
CC2 tongue_2_18
CC2 tongue_18
CR1 muscle_22
CF2 muscle_22
CV1 muscle_16
CV1 heart A_16
CF2 heart A_22
CV2 heart A_16
CC3 heart A_22
CV1 heart V_16
CV2 heart V_16
CC3 heart V_22
CJ1 heart V_9
DF1 heart V_9
CR1 amnion_22
CV1 amnion_16
CF2 amnion_22
CC2 amnion_18
CI1 amnion_16
CP1 eye_22
CF2 eye_22
CV2 eye_16
DJ1 skin_9
CF2 skin_22
CC2 skin_18
CV1 skin_16
CI1 skin_16
CR1 skin_22
B
Height
0
1
2
3
4
DA2 pancreas_9
BP1 pancreas_9
CR1 spleen_22
CF2 spleen_22
CC2 spleen_18
CI1 spleen_16
CJ1 intestine_9
CJ1 tongue_9
DF1 tongue_9
DF1 umbilical cord_9
DJ1 muscle_9
CV1 eye_16
CJ1 umbilical cord_9
CJ1 eye_9
CH1 muscle_9
DF1 eye_9
DF1 skin_9
DJ1 skin_9
CJ1 spinal cord_9
DF1 spinal cord_9
DF1 brain_9
CR1 brain_22
CI1 brain_16
CC2 brain_18
DJ1 brain_9
CH1 adrenal_9
CR1 intestine_22
CC2 intestine_18
CI1 intestine_16
CF2 intestine_22
CF2 intestine_22
CV2 heart A_16
CF2 heart A_22
CV1 heart A_16
CC3 heart A_22
CI1 muscle_16
CF2 muscle_22
CV1 heart V_16
CJ1 heart V_9
CV2 heart V_16
DF1 heart V_9
CC3 heart V_22
CV1 muscle_16
CR1 muscle_22
CF2 amnion_22
CI1 amnion_16
CC2 amnion_18
CV1 amnion_16
CJ1 amnion_9
DF1 amnion_9
CI1 liver_16
CC2 liver_18
CR1 liver_22
DJ1 liver_9
DA2 liver_9
CF2 liver_22
DJ1 kidney_9
DF1 kidney_9
CJ1 pancreas_9
CR1 kidney_22
CF2 kidney_22
CI1 kidney_16
CC2 kidney_18
CC2 pancreas_18
CI1 pancreas_16
BO1 pancreas_22
CF2 pancreas_22
CI1 mat. endom._16
CR1 mat. endom._22
CC2 adrenal_18
CF2 adrenal_22
CL3 gonad_9
CH1 gonad_9
DJ2 gonad_18
CV2 gonad_16
BP1 lung_9
CF2 lung_22
DJ1 lung_9
CI1 lung_16
CC2 lung_18
CR1 lung_22
CJ1 stomach_9
DF1 stomach_9
CF2 stomach_22
CC2 stomach_18
CV1 stomach_16
CR1 stomach_22
DF1 placenta_9
CJ1 placenta_9
CV1 placenta_16
CR1 placenta_22
CF2 placenta_22
CC2 placenta_18
CI1 placenta_16
CR1 chorion_22
CJ1 mat. endom._9
CP1 chorion_22
CI1 chorion_16
CR1 amnion_22
CR1 skin_22
CV1 skin_16
CF2 skin_22
CI1 skin_16
CC2 skin_18
CP1 eye_22
CF2 eye_22
CV2 eye_16
CC2 tongue_18
CC2 tongue_2_18
CI1 tongue_16
CR1 tongue_22

[Supplementary Fig. 1]

Hierarchical clustering of the fetal samples.

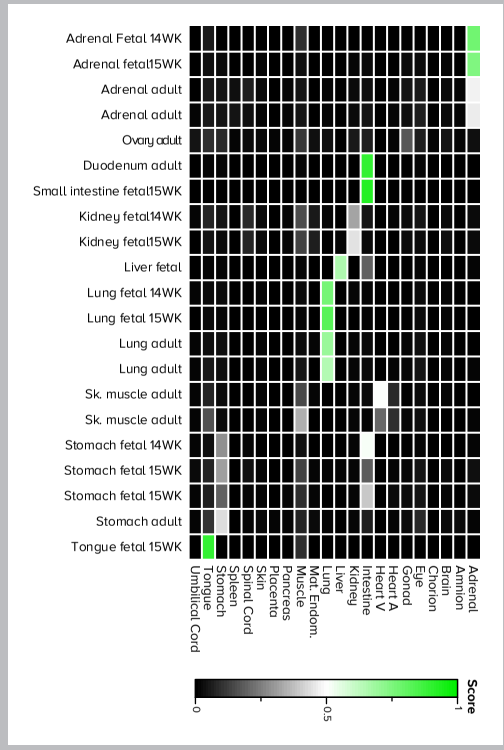
[A] Dendrogram showing the hierarchical clustering of all fetal and maternal endometrium samples, based on the Pearson correlation of the expression levels of the 90 fetal classifier genes. **[B]** Dendrogram showing the hierarchical clustering of all fetal and maternal endometrium samples, based on the Pearson correlation of the expression levels of the 500 most variable genes of the entire fetal data set (Top 500 fetal). Abbreviations: mat. endom., maternal endometrium.


A
Brain
1st trimester
2nd trimester
87
162
52
170
40
40
1008
Adult
Skin
1st trimester
2nd trimester
88
139
67
97
33
63
881
Adult
B
Intestine
1st trimester
2nd trimester
123
111
77
133
49
61
1027
Adult
Liver
1st trimester
2nd trimester
28
75
23
82
19
20
817
Adult
Lung
1st trimester
2nd trimester
101
131
40
144
28
72
840
Adult
Pancreas
1st trimester
2nd trimester
164
126
30
79
26
29
448
Adult
Stomach
1st trimester
2nd trimester
82
131
43
140
34
60
866
Adult
C
Adrenal
1st trimester
2nd trimester
99
90
73
132
62
73
843
Adult
Gonad
1st trimester
2nd trimester
Adult female
Adult male
43
524
106
12
314
62
65
17
736
104
35
23
25
6
7
Heart
Ventricle 1st trimester
Ventricle 2nd trimester
Atrium 2nd trimester
Adult
39
24
9
6
25
51
105
19
680
144
51
22
15
43
15
Kidney
1st trimester
2nd trimester
90
135
100
127
50
61
918
Adult
Muscle
1st trimester
2nd trimester
122
167
81
71
18
52
584
Adult

[Supplementary Fig. 2]

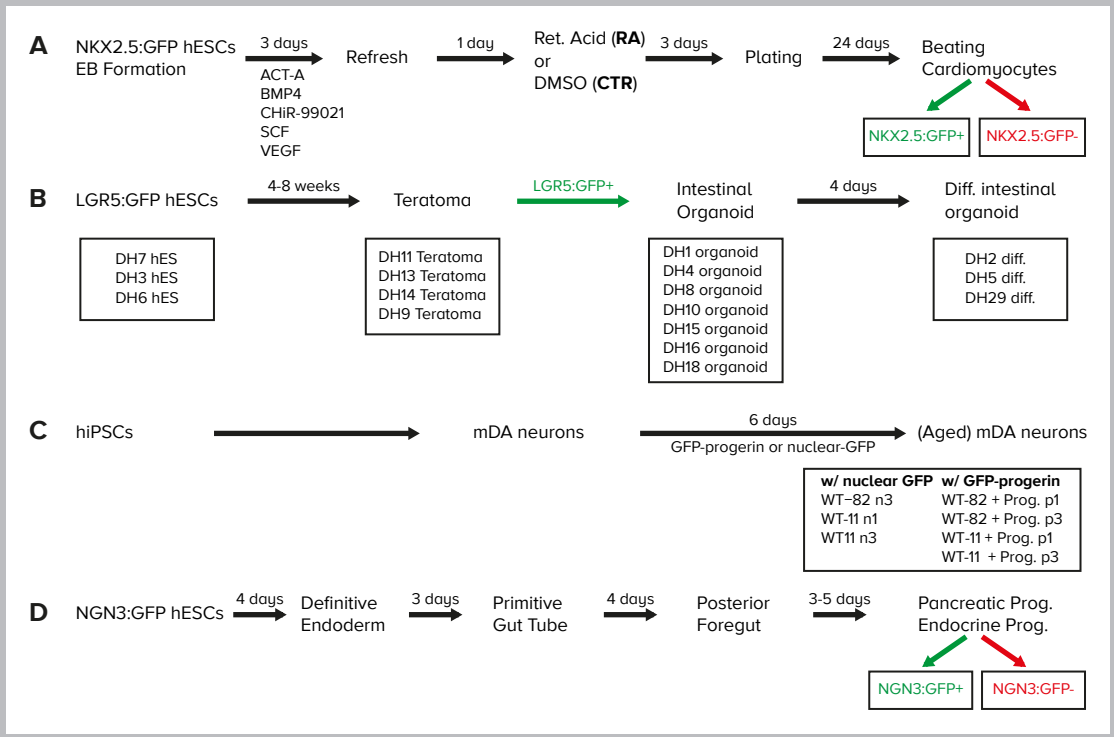
Stage-specific gene expression signatures per organ.

Venn diagrams of commonly and uniquely highly expressed genes of the first trimester, second trimester and adult samples per organ derived mainly from **[A]** ectoderm, **[B]** endoderm, and **[C]** mesoderm. The genes shown here have 10 x higher gene expression levels in counts per million (CPM) than the average expression of the corresponding dataset. The stage-specific genes per organ and the intersections genes are given in Supplementary Table 4.


Adrenal Fetal 14WK
Adrenal fetal15WK
Adrenal adult
Adrenal adult
Ovary adult
Duodenum adult
Small intestine fetal15WK
Kidney fetal14WK
Kidney fetal15WK
Liver fetal
Lung fetal 14WK
Lung fetal 15WK
Lung adult
Lung adult
Sk. muscle adult
Sk. muscle adult
Stomach fetal 14WK
Stomach fetal 15WK
Stomach fetal 15WK
Stomach adult
Tongue fetal 15WK
Umbilical Cord
Tongue
Stomach
Spleen
Spinal Cord
Skin
Placenta
Pancreas
Muscle
Mat. Endom.
Lung
Liver
Kidney
Intestine
Heart V
Heart A
Gonad
Eye
Chorion
Brain
Amnion
Adrenal
Score
0
0.5
1

[Supplementary Fig. 3]

Prediction of the microarray human adult and fetal dataset from Nazor *et al*. (2012).
Identity scores of 21 human adult and fetal organ/tissue samples from the microarray dataset generated by Nazor *et al*. [24]. The rows represent the 22 organs/tissues from the fetal training set and the columns depict the samples in the test set. The identity scores range from zero (black) to one (green). The values of all identity scores are given in Supplementary Table 3.


A
NKX2.5:GFP hESCs
EB Formation
3 days
ACT-A
BMP4
CHiR-99021
SCF
VEGF
Refresh
1 day
Ret. Acid (RA)
or
DMSO (CTR)
3 days
Plating
24 days
Beating
Cardiomyocytes
NKX2.5:GFP+
NKX2.5:GFP-
B
LGR5:GFP hESCs
4-8 weeks
Teratoma
LGR5:GFP+
Intestinal
Organoid
4 days
Diff. intestinal
organoid
DH7 hES
DH3 hES
DH6 hES
DH11 Teratoma
DH13 Teratoma
DH14 Teratoma
DH9 Teratoma
DH1 organoid
DH4 organoid
DH8 organoid
DH10 organoid
DH15 organoid
DH16 organoid
DH18 organoid
DH2 diff.
DH5 diff.
DH29 diff.
C
hiPSCs
mDA neurons
6 days
GFP-progerin or nuclear-GFP
(Aged) mDA neurons
w/ nuclear GFP
WT–82 n3
WT-11 n1
WT11 n3
w/ GFP-progerin
WT-82 + Prog. p1
WT-82 + Prog. p3
WT-11 + Prog. p1
WT-11 + Prog. p3
D
NGN3:GFP hESCs
4 days
Definitive
Endoderm
3 days
Primitive
Gut Tube
4 days
Posterior
Foregut
3-5 days
Pancreatic Prog.
Endocrine Prog.
NGN3:GFP+
NGN3:GFP-

[Supplementary Fig. 4]

Schematics of the differentiation protocols used to generate hPSC-derived cardiomyocytes, intestinal organoids, dopaminergic neurons and pancreatic progenitor cells.

[A] Generation of human cardiomyocytes from NKX2.5:GFP human embryonic stem cells (hESCs) [33]. **[B]** Generation of intestinal organoids from LGR5:GFP human embryonic stem cells (hESCs) [20]. **[C]** Generation of (aged) dopaminergic neurons (mDA neurons) from human induced pluripotent stem cells (hiPSCs) [19]. **[D]** Generation of pancreatic progenitors from NGN3:GFP human embryonic stem cells (hESCs) [18].

Adrenal
Amnion
Brain
Chorion
Eye
Gonad
Heart A
Heart V
Intestine
Kidney
Liver
Lung
Mat. Endom.
Muscle
Pancreas
Placenta
Skin
Sp. Cord
Spleen
Stomach
Tongue
Umb. Cord
NKX2.5:GFP- CTR
NKX2.5:GFP- RA
NKX2.5:GFP+ CTR
NKX2.5:GFP+ RA
Devalla et al.
DH1 Organoid
DH4 Organoid
DH8 Organoid
DH10 Organoid
DH15 Organoid
DH16 Organoid
DH18 Organoid
DH2 Diff.
DH5 Diff.
DH29 Diff.
Forster et al.
WT- 82 n3
WT- 82 + Prog. p1
WT- 82 + Prog. p3
WT- 11 n1
WT- 11 n3
WT- 11 + Prog. p1
WT- 11 + Prog. p3
Miller et al.
NGN3:GFP+
NGN3:GFP-
Liu et al.
SULT2A1 Adrenal
PHOX2B Adrenal
NR5A1 Adrenal
ARHGAP36 Adrenal
CYP17A1 Adrenal
CPB1 Adrenal
TNMD Amnion
GABRP Amnion
ITGB6 Amnion
HOXD9 Amnion
VTCN1 Amnion
KRT24 Amnion
MUC16 Amnion
C1orf61 Brain
NCAN Brain
EMX2 Brain
MYT1L Brain
HOXC6 Brain
SERPINE1 Chorion
NOTUM Chorion
HLA-DRB1 Chorion
SERPINA3 Chorion
DCT Eye
FOXF2 Eye
RSPO2 Eye
MAB21L1 Eye
SHOX Eye
STMN2 Gonad
WNT6 Gonad
LHX9 Gonad
REG3G Gonad
RP13-491I15.5 Gonad
WT1 Gonad
NPPA-AS1 Heart A
MYL3 Heart V
MEP1A Intestine
MUC13 Intestine
MAB21L2 Intestine
SLC12A1 Kidney
PAX2 Kidney
NPHS2 Kidney
ITIH1 Liver
TF Liver
ANGPTL3 Liver
CYP4A22 Liver
SCGB3A2 Lung
LINC00514 Lung
MEOX2 Mat. Endom.
CADM3 Mat. Endom.
DEFB1 Mat. Endom.
CRABP1 Mat. Endom.
PRL Mat. Endom.
APOD Mat. Endom.
TNMD Muscle
MYOG Muscle
HOXC10 Muscle
CXorf64 Muscle
RP11-59J5.1 Muscle
SERPINI2 Pancreas
MNX1 Pancreas
PDX1 Pancreas
FXYD6-FXYD2 Pancreas
LGALS14 Placenta
ALDH1A2 Placenta
SLC13A4 Placenta
KISS1 Placenta
TBX15 Skin
GJB6 Skin
RSPO2 Skin
KRT1 Skin
KRTDAP Skin
LINC00514 Skin
PRSS8 Sp. Cord
HOXB8 Sp. Cord
NHLH2 Sp. Cord
CD5L Spleen
RNF128 Spleen
NR5A1 Spleen
CLDN18 Stomach
BARX1 Stomach
GIF Stomach
MUC6 Stomach
MUC5AC Stomach
C7 Tongue
MYBPH Tongue
SPRR1B Tongue
ALX4 Umb. Cord
DCX Umb. Cord
CHGB Umb. Cord
WIF1 Umb. Cord
PITX2 Umb. Cord
CRABP1 Umb. Cord
GPR22 Umb. Cord
DNER Umb. Cord
HOXA10 Umb. Cord

[Supplementary Fig. 5]

Expression levels of the 90 classifier genes in the hPSC-derived cardiomyocytes, intestinal organoids, dopaminergic neurons and pancreatic progenitor cells.

Heatmap of the expression levels of the fetal transcriptional barcode (90 classifier genes, Fig. 2) and the organ they represent during first and second trimester compared to the differentiation experiments from Fig. 4 and 5 [18-20, 33]. The expression levels (in counts per million, CPM) of the classifier genes in the corresponding tissue are depicted in green (50-100 CPM, light green; 100-1000 CPM, green; > 1000 CPM, dark green), whereas the expression levels in other than the classifying tissues are displayed in blue (50-100 CPM, light blue; 100-1000 CPM, blue; > 1000 CPM, dark blue).

CHAPTER 4

DNA METHYLATION LANDSCAPES OF HUMAN FETAL DEVELOPMENT

Roderick C. Slieker[1,*], Matthias S. Roost[2,*], Liesbeth van Iperen[2], H. Eka D. Suchiman[1], Elmar W. Tobi[1], Françoise Carlotti[3], Eelco J.P. de Koning[3,4], P. Eline Slagboom[1], Bastiaan T. Heijmans[1,#], Susana M. Chuva de Sousa Lopes[2,5,#]

[1] Molecular Epidemiology Section,

[2] Department of Anatomy and Embryology,

[3] Department of Nephrology, Leiden University Medical Center, Leiden, the Netherlands

[4] Hubrecht Institute for Developmental Biology and Stem Cell Research, University Medical Center, Utrecht, the Netherlands

[5] Department for Reproductive Medicine, Ghent University Hospital, Ghent, Belgium

[*] Equal first authorship contribution

[#] Equal senior authorship contribution

PLoS Genetics 2015, **11**(10):e1005583

ABSTRACT

Remodelling the methylome is a hallmark of mammalian development and cell differentiation. However, current knowledge of DNA methylation dynamics in human tissue specification and organ development largely stems from the extrapolation of studies *in vitro* and animal models. Here, we report on the DNA methylation landscape using the 450k array of four human tissues (amnion, muscle, adrenal and pancreas) during the first and second trimester of gestation (9, 18 and 22 weeks). We show that a tissue-specific signature, constituted by tissue-specific hypomethylated CpG sites, was already present at 9 weeks of gestation (W9). Furthermore, we report large-scale remodelling of DNA methylation from W9 to W22. Gain of DNA methylation preferentially occurred near genes involved in general developmental processes, whereas loss of DNA methylation mapped to genes with tissue-specific functions. Dynamic DNA methylation was associated with enhancers, but not promoters. Comparison of our data with external fetal adrenal, brain and liver revealed striking similarities in the trajectory of DNA methylation during fetal development. The analysis of gene expression data indicated that dynamic DNA methylation was associated with the progressive repression of developmental programs and the activation of genes involved in tissue-specific processes. The DNA methylation landscape of human fetal development provides insight into regulatory elements that guide tissue specification and lead to organ functionality.

INTRODUCTION

Methylation of CpG dinucleotides in the mammalian genome is a key epigenetic mark. Adult tissues have highly distinct genome-wide DNA methylation signatures consistent with the regulation of cell differentiation by epigenetic mechanisms [1–3]**. Differences in DNA methylation between tissues have been shown to mark differences between germ layers** [4]**, preferentially at regions with low CpG content** [2, 5, 6]**, at enhancers** [4] **and alternative promoters** [7, 8]**.**

Multiple studies have reported on the reprogramming of the human methylome during preimplantation embryo development [9–11]**. In line with previous data on mice** [12]**, in humans DNA methylation is largely erased after conception, the paternal genome being actively and the maternal genome passively demethylated, to become remethylated with the implantation of the embryo** [9, 10, 13, 14]**. However, systematic and detailed reports on DNA methylation dynamics during human fetal development remain scarce** [15]**, while such data is key to understand how epigenetic mechanisms drive tissue specification and organ functionality. Current views of fetal DNA methylation dynamics are largely extrapolated from studies on the differentiation of human and mouse cells *in vitro*** [7, 15–21]**, and the comparison of differentiated tissues to human induced pluripotent stem cells and human embryonic stem cell lines** [15]**. An exception is fetal brain development in humans, for which recently reported *in vivo* data showed significant DNA methylation remodelling** [15, 22]**.**

Recent developments in technology for interrogating genome-wide DNA methylation at single-nucleotide resolution [23] **and detailed functional annotation of the human genome** [24, 25] **provide an opportunity to chart DNA methylation during development and assign biological roles to the**

regions involved. Taking advantage of these developments, we report on DNA methylation dynamics during human fetal development of one extraembryonic tissue and three organs relevant for complex human diseases. This organ-specific catalogue of DNA methylation during development provides fundamental insights into processes guiding human development, but also into the biological function of non-coding regions, which are emerging as important from genome-wide association studies (GWASs) of complex diseases [26]**. In addition, this catalogue may serve as a reference for studies on the role of epigenetic mechanisms in the association between an adverse prenatal environment and adulthood disease** [27] **since DNA methylation marks may have an heightened sensitivity for environmental perturbations during remodelling** [28]**.**

MATERIALS AND METHODS

ETHICAL STATEMENT

The Medical Ethical Committee of the Leiden University Medical Center approved this study (P08.087). Informed consent was obtained on the basis of the Declaration of Helsinki (World Medical Association).

FETAL TISSUE

Human fetal tissues (amnion, skeletal muscle, adrenal glands, pancreas) at gestational age W9, W18, W22 [Supplementary Fig. 2A] **were collected from elective abortion material (vacuum aspiration) without medical indication. In this study, «weeks of gestation» was used as determined by the last menstrual period (LMP). After collection, the material was washed with 0.9% NaCl (Fresenius Kabi, France) and the identified organs were immediately snap-frozen using dry ice and stored at -80°C until further processing. Histology was performed as previously described** [29]**. The images were taken with an Olympus AX70 microscope (Olympus, Japan) provided with a XC50 digital colour camera (Olympus, Japan).**

DNA EXTRACTION

Tissues were homogenized with a pestle and lysed overnight at 56°C with proteinase K (600 mAU/ml, Qiagen, Germany) in ATL buffer (Qiagen, Germany). After lysis, residual RNA in the samples was degraded using RNase A (10 mg/µl, Invitrogen, USA). Subsequently, genomic DNA (gDNA) was extracted on the basis of phenol/chloroform. Briefly, lysates were transferred to Phase Lock Heavy Gel 2ml Eppendorf tubes (5PRIME, Germany) and 700 µl of 25:24:1 Phenol/Chloroform/Isoamyl alcohol was added and spun down for 5 minutes. The aqueous phase was transferred to a Phase Lock tube and

the latter step was repeated. The aqueous phase was transferred to a new Phase Lock tube and 700 µl 24:1 Chloroform/Isoamyl alcohol was added and spun down for 5 minutes. The aqueous phase was transferred to a Phase Lock tube and the latter step was repeated. The aqueous phase was transferred to a new 2 µl Eppendorf tube (Eppendorf AG, Germany), 70 µl 3M sodium acetate (Ambion, USA) and 1400 µl ice cold 100% ethanol were added. gDNA was precipitated over night at -20°C. Eppendorf tubes were spun down at 4°C for 15 minutes and washed twice with 70% ethanol. After the pellet was dry, gDNA was solubilized in AE buffer (Qiagen, Germany) and stored at 4°C. DNA concentration was determined using the Qubit dsDNA BR Assay Kit on a Qubit 2.0 Fluorometer (Invitrogen, USA). gDNA was bisulfite converted using the EZ-96 DNA methylation kit (Zymo Research, Orange County, USA) with an average input of 600 ng gDNA. Following bisulfite conversion, DNA methylation data was generated using Illumina HumanMethylation450 BeadChip according to the manufacturer's protocol.

(PRE-) PROCESSING OF THE ILLUMINA 450K BEADCHIP DATA

All analyses were performed using R statistics, version 3.0.1. The 65 polymorphic SNP probes featured on the 450k array were used to exclude potential sample mix ups. Data was imported in R using minfi [30] **and processed and normalized using a custom pipeline: Arrays were removed if they had a low median intensity, a high background signal or with incomplete bisulfite conversion, but none were excluded** [Supplementary Fig. 1B–1D]**. The CpGs on chromosome X and Y were used to confirm sex** [Supplementary Fig. 1C]**. Next, probes with a low bead count (< 3), high detection P-value (> 0.01), and with a low success rate (< 95%), and ambiguously mapped probes** [31] **were removed. After probe filtering, all arrays contained > 95% of the original number of probes. Background correction and colour correction were applied and the data was quantile**

normalized (lumi [32]**). To adjust for the type I/II bias BMIQ was applied** [33]**. In our analyses, CpG sites in the sex chromosomes (Y and X) were excluded. High correlation was found between samples from the same time point and tissue but also between time points of the same tissue [Supplementary Fig. 1D]. To exclude chromosomal abnormalities, we calculated the copy number aberration based on the signal intensities using the method published by Feber *et al.*** [34] **as implemented in the R package ChAMP** [35]**. From these results, no abnormalities were found in the samples used [Supplementary Fig. 1E].**

BIOINFORMATICS ANALYSES

Multidimensional scaling and clustering was performed based on Euclidean distance. For the DNA methylation over time within features, a genic annotation was combined with a CGI-centric annotation as presented before [6] **to determine the median methylation per combined feature. P were calculated using quantile regression based on the median (R package quantreg** [36]**). Figures were made using the R-packages ggplot2** [37] **and GenomeGraphs** [38]**.**

Tissue-specific hypomethylation: CpGs with a standard deviation ≥ 0.1 within the tissue of interest were discarded. Relative tissue-specific hypomethylation was defined as hypomethylation of the tissue of interest compared to the other tissues, with a difference of ≥ 0.20 in beta value. CpGs sites were selected if a difference was consistent in each of the time points.

Dynamic methylation: CpGs with a high standard deviation ≥ 0.10 within time point/tissues indicative of an instable estimate of DNA methylation were discarded from this analysis. Gain and loss of methylation was defined a gain/loss of ≥ 0.20 between W9 and W22 and W18 in between the two time points (W18 was allowed to be 0.05 lower/higher in beta value than W9/W22 respectively). The CpGs with a gain or a loss of methylation were used for the combined genic/CGI-centric

annotation and expressed as an odds ratio. Chromatin state segmentation data [25] were downloaded from the Epigenomics Roadmap Project for fetal muscle (W15 female), fetal adrenal (W13 male), amnion (W16 male) and pancreatic islets (adult) and enrichment of dynamically methylated CpGs was calculated.

DMRs: In both, the relative hypomethylated CpGs (tHRs) and the CpGs with a gain or a loss of methylation (dDMRs), DMRs were called using an algorithm described before [6]. Briefly, DMRs (tHRs and dDMRs) were defined by three consecutive CpGs that matched a criterion (that is, relative hypomethylation or gain/loss of methylation) with a maximum of 1 kb between CpGs and with at highest three CpGs that did not match the criterion.

Gene ontology: Tissue-specific hypomethylated and dynamically methylated CpGs were mapped to their nearest gene (that is to the nearest TSS or TES of a gene) and tested for enrichment of gene ontology terms using DAVID [39]. For the tissue-specific hypomethylation we used a P cut-off of 0.05 on the raw P as the number of CpGs was relatively low. For the dynamically methylated CpGs a FDR cut-off of 0.05 was set as cut-off for enrichment in GO terms. A background set was used containing nearest genes of all CpGs covered on the array.

Gene expression data: Transcriptional data of the four tissues at W9, W18, W22 (amnion: n = 2, 3, 2; muscle: n = 2, 2, 2; adrenal: n = 1, 1, 1; pancreas: n = 3, 2, 2) were used. The counts per million (CPM) expression levels were calculated using the R package edgeR 3.2.4 [40, 41]. For the plots, the arithmetic mean of the biological replicates was used and the median of all genes plotted. To access enrichment of up- and down-regulation, a probability test was used.

MYOD, DNAse I, histone marks and WGBS: Overlaps between *MYOD* binding peaks and muscle dDMRs were calculated. To test for significance, we calculated an empirical distribution by performing 20,000 permutations with 482 (gain of methylation) and 530 (loss of methylation) DMR-like regions

each and determined the overlap with the *MYOD* binding sites. DMR-like regions were defined as regions with equal characteristics as dDMRs identified: an inter-CpG distance smaller than 1 kb and an average length of five CpGs per DMR-like region (n ~8×10^4 regions). The two-sided P was determined using the empirical distribution.

DNAse I and histone mark data of human skeletal muscle myoblasts (HSMMs) and human skeletal muscle myotubes (HSMMtubes) were downloaded from the ENCODE website [24]. DNAse I hypersensitivity was expressed as the count of DNAse-seq tags. The enrichment of histone marks was expressed as the log2 of the ChIP/input. The total number of reads within the myotubes was different from the total number of reads in the myoblast data and, therefore, the data was normalized. dDMRs were classified as island (CGIs and their shores) or non-island dDMRs, and DNAse-seq tags and histone marks around dDMRs were mapped up to 5 kb up- and downstream.

CpG sites of WGBS data were mapped to hypomethylated and dynamic regions and their 5 kb flanking regions. Using a smooth spline, the methylation around the regions was smoothed for the adult and fetal data.

ACCESSION NUMBERS

Methylation data has been deposited in the NCBI's Gene Expression Omnibus [42] under accession number GSE56515. External datasets that have been used in this manuscript include: fetal and adult DNA methylation data of various tissues from Nazor *et al.* (Gene Expression Omnibus (GEO) accession number: GSE31848) [15], fetal brain DNA methylation data from Spiers *et al.* (GEO accession number: GSE58885) [22], fetal liver DNA methylation data from Bonder *et al.* (GEO accession number: GSE61279) [43], adult DNA methylation data of various tissues from Slieker *et al.* (GEO accession number: GSE48472) [6], fetal Deep SAGE expression data of the four

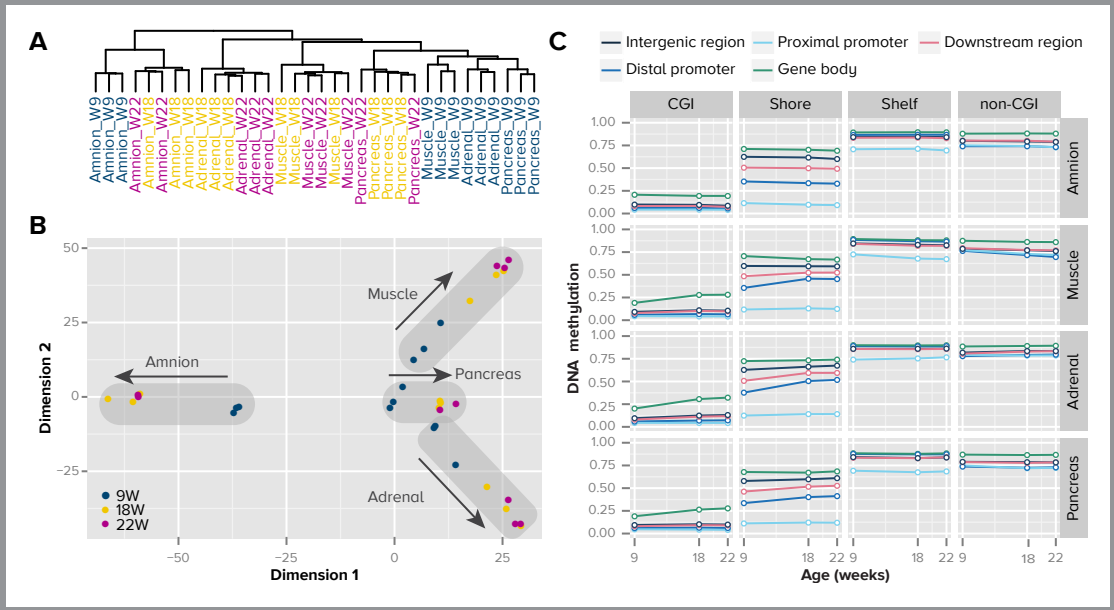
tissues studied here from Roost *et al.* (GEO accession number: GSE66302) [29]**, adult DNA methylation brain data from Pidsley *et al.* (GEO accession number: GSE61431)** [44]**, WGBS data of fetal and adult muscle generated by the Epigenomics Roadmap consortium (GEO accession numbers: GSM1172596 and GSM1010986), *MYOD* binding peaks from MacQuarrie *et al.* (GEO accession numbers: GSM1218849 and GSM1218850)** [45]**.**

RESULTS

FETAL DNA METHYLATION REFLECTS TISSUE ORIGIN AND DEVELOPMENTAL AGE

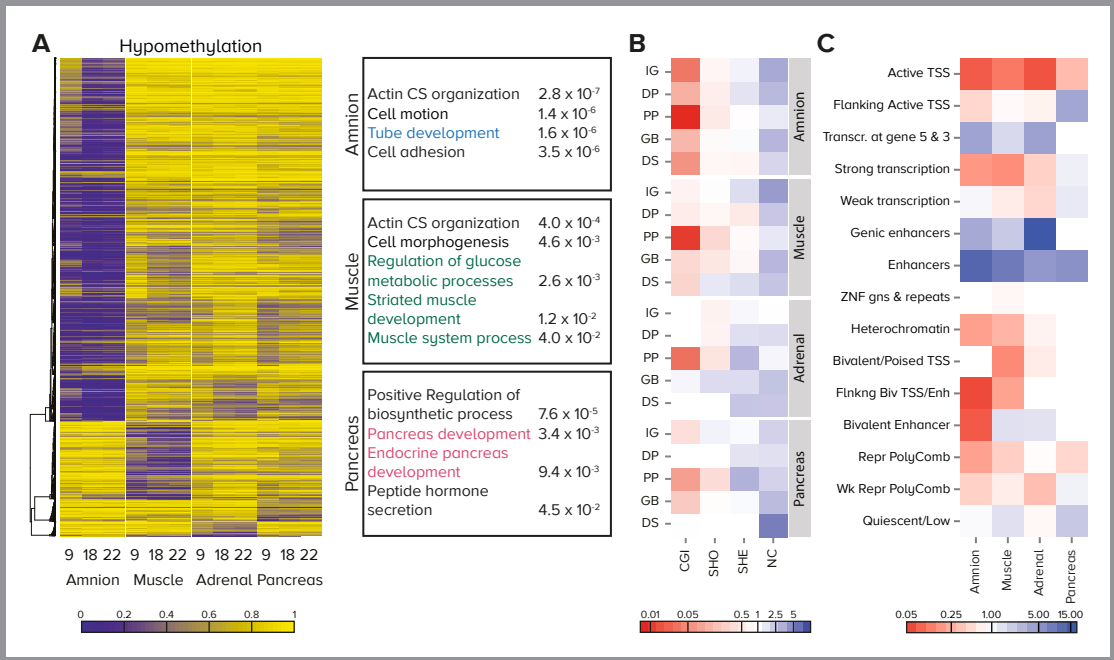
To study DNA methylation dynamics in human fetal development, amnion, muscle, adrenal and pancreas samples of 11 fetuses were obtained at 9, 18 and 22 weeks of gestation (W9, W18 and W22 [Supplementary Fig. 1A]**). Genome-wide DNA methylation was investigated with the Illumina 450k array resulting in data on 452,490 CpG sites after quality control** [46] [Supplementary Fig. 1B–1E]**. The study included three biological replicates per tissue and time point, except for W22 amnion (n = 2) and W22 pancreas (n = 2)** [Supplementary Fig. 2A]**.**

We first assessed differences in overall DNA methylation patterns between time points and tissues using hierarchical clustering based on Euclidean distance [Fig. 1A] **and multidimensional scaling (MDS)** [Fig. 1B; Supplementary Fig. 2B]**. DNA methylation patterns clearly differentiated the four tissue types studied** [Fig 1A and 1B; Supplementary Fig. 2B]**. The amnion, representing an extraembryonic tissue, clustered separately from the three embryonic tissues** [Fig. 1A]**. Within the embryonic cluster, all W9 tissues (representative of the first trimester) clustered together, whereas W18–W22 tissues (representative of the second trimester) were present towards the edges of the MDS plot** [Fig. 1B; Supplementary Fig. 2B]**. Despite the distinct differences in DNA methylation patterns from the first to second trimester, the total number of hypo-, intermediately and hypermethylated CpGs remained constant across time points and tissues (including the extraembryonic amnion) on both autosomes and, in females, the X chromosome** [Supplementary Fig. 2C and 2D]**. This suggests that the observed differences in the MDS plot were not driven by changes in average levels**


A
Amnion_W9
Amnion_W9
Amnion_W9
Amnion_W22
Amnion_W18
Amnion_W22
Amnion_W18
Amnion_W18
Adrenal_W18
Adrenal_W18
Adrenal_W18
Adrenal_W22
Adrenal_W22
Adrenal_W22
Muscle_W18
Muscle_W18
Muscle_W22
Muscle_W22
Muscle_W18
Muscle_W22
Pancreas_W22
Pancreas_W18
Pancreas_W18
Pancreas_W18
Pancreas_W22
Muscle_W9
Muscle_W9
Muscle_W9
Adrenal_W9
Adrenal_W9
Adrenal_W9
Pancreas_W9
Pancreas_W9
Pancreas_W9
B
Muscle
Amnion
Pancreas
Adrenal
9W
18W
22W
Dimension 2
Dimension 1
50
25
0
−25
−50
−25
0
25
C
Intergenic region
Proximal promoter
Downstream region
Distal promoter
Gene body
CGI
Shore
Shelf
non-CGI
Amnion
Muscle
Adrenal
Pancreas
DNA methylation
1.00
0.75
0.50
0.25
0.00
9
18
22
Age (weeks)

[Fig. 1]

Tissue- and time-specific DNA methylation patterns during human fetal development.
[A] Clustering based on Euclidean distance. **[B]** Multidimensional scaling based on Euclidean distance of the four tissues. **[C]** Median DNA methylation for each of the four tissues over time with a combined genic and CGI-centric annotation. CGI, CpG island.


A
Hypomethylation
9 18 22 9 18 22 9 18 22 9 18 22
Amnion Muscle Adrenal Pancreas
0 0.2 0.4 0.6 0.8 1
Amnion
Actin CS organization 2.8 x 10-7
Cell motion 1.4 x 10-6
Tube development 1.6 x 10-6
Cell adhesion 3.5 x 10-6
Muscle
Actin CS organization 4.0 x 10-4
Cell morphogenesis 4.6 x 10-3
Regulation of glucose metabolic processes 2.6 x 10-3
Striated muscle development 1.2 x 10-2
Muscle system process 4.0 x 10-2
Pancreas
Positive Regulation of biosynthetic process 7.6 x 10-5
Pancreas development 3.4 x 10-3
Endocrine pancreas development 9.4 x 10-3
Peptide hormone secretion 4.5 x 10-2
B
IG DP PP GB DS
Amnion
Muscle
Adrenal
Pancreas
CGI SHO SHE NC
0.01 0.05 0.5 1 2.5 5
C
Active TSS
Flanking Active TSS
Transcr. at gene 5 & 3
Strong transcription
Weak transcription
Genic enhancers
Enhancers
ZNF gns & repeats
Heterochromatin
Bivalent/Poised TSS
Flnkng Biv TSS/Enh
Bivalent Enhancer
Repr PolyComb
Wk Repr PolyComb
Quiescent/Low
Amnion Muscle Adrenal Pancreas
0.05 0.25 1.00 5.00 15.00

[Fig. 2]

Sets of hypomethylated CpG sites are tissue-specific.

[A] Heatmap representing hypomethylated CpGs per tissue, defined by a beta difference of ≥ 0.2 of the studied tissue compared to the other tissues. **[B]** Combined genic and CGI-centric annotation for the hypomethylated CpGs per tissue represented as the odds ratio (see Supplementary Fig. 3A for odds ratios). CGI, CpG island; DP, distal promoter; DS, downstream; GB, gene body; IG, intergenic; NC, non-CGI; PP, proximal promoter; SHE, shelves; SHO, shores. **[C]** Enrichment of hypomethylated CpGs in the chromatin state segmentation states for the matching tissues (amnion, fetal muscle, fetal adrenal and adult pancreatic islets; see Supplementary Fig. 3B for odds ratios).

of DNA methylation, but rather due to tissue- and time-specific changes in DNA methylation.

To validate our findings, we integrated our data with three previously published Illumina 450k datasets on 10 human fetal tissues [15, 22, 43]**. Hierarchical clustering of all data together (n = 117) confirmed the presence of distinct tissue- and time-specific DNA methylation patterns in fetal tissues** [Supplementary Fig. 2E]**.**

The characteristics and biological function of DNA methylation depend on the local CpG content and position relative to genes [47]**. We mapped CpG sites (CpGs) to CpG islands (CGIs as defined in the UCSC genome browser; 138,919 CpGs), their shores (±2 kb of CGIs; 103,453 CpGs) and shelves (±2 kb of shores; 42,227 CpGs), and remaining CpG-poor non-CGI regions (157,560 CpGs), and to regions relative to gene locations including distal promoters (-10 kb – -1.5 kb; 21,101 CpGs), proximal promoters (-1.5 kb – +0.5 kb; 171,077 CpGs), gene bodies (+0.5 kb – 3' untranslated region (UTR); 175,062 CpGs), downstream regions (3' UTR – +5 kb; 8,563 CpGs) and remaining intergenic regions (66,356 CpGs** [Fig. 1C]**). CpGs were commonly hypomethylated in CGIs, intermediately methylated in shores and hypermethylated in both shelves and non-CGI regions** [Fig. 1C]**. These patterns differed by genic position, e.g. CGI methylation was lowest in proximal promoters and highest in gene bodies. Annotation-specific methylation differences were found between W9 and W22, as CpGs in CGIs and shores tended to increase (e.g. gene body CGIs and distal promoter shores), whereas CpGs in non-CGI regions decreased in methylation (e.g. non-CGI promoters). For a subset of annotations, the amnion showed slightly different DNA methylation patterns than for embryonic tissues, e.g. for gene body CGIs** [Fig. 1C]**. Taken together, our data imply that DNA methylation is highly dynamic during fetal development without affecting the average level of DNA methylation.**

HYPOMETHYLATION DISCRIMINATES TISSUES INDEPENDENT OF DEVELOPMENTAL AGE

It has been shown that each adult tissue is defined by tissue-specific DNA hypomethylation [15, 48, 49]**. Since the four fetal tissues analysed showed a clear DNA methylation signature that corresponded to separated clusters** [Fig. 1A and 1B]**, we investigated whether combinations of tissue-specific DNA hypomethylated CpGs were present irrespective of its developmental stage. To do this, we identified CpGs that were relatively hypomethylated (defined as a DNA methylation difference of > 20%) in each tissue compared to all others throughout the three time points of fetal development investigated. The analysis showed indeed that, independently of the developmental age, each tissue showed a cluster of tissue-specific hypomethylated CpGs** [Fig. 2A]**. The early lineage segregation of the amnion was further confirmed by the comparatively large number of CpGs (3,536 CpGs) that were exclusively hypomethylated across amniotic samples. In contrast, the embryonic tissues contained much fewer tissue-specific hypomethylated CpGs (muscle 756 CpGs; adrenal 140 CpGs; pancreas 220 CpGs) reflecting their common origin of the epiblast, that gives rise to all embryonic tissues. Genes mapping (i.e. the nearest gene locus) to the specific hypomethylated CpGs per tissue regardless of the time point (amnion 2372, muscle 548, adrenal 120, pancreas 175) were enriched for biological processes that included GO terms characteristic of amnion, muscle and pancreas development and function** [Supplementary Table 1]**.**

When annotated to genic and CGI-related location, it became evident that tissue-specific hypomethylation was enriched for non-CGI regions (P < 0.0001) and highly depleted at CGIs, in particular when mapping to proximal promoters (P < 0.0001; [Fig. 2B; Supplementary Fig. 3A]**). To gain further insight in the biological role of genomic regions displaying tissue-specific hypomethylation, we used chromatin state segmentations for fetal muscle, fetal adrenal, amnion and adult**

pancreatic islets generated by the Epigenomics Roadmap [25]**. Tissue-specific hypomethylation was strongly enriched at enhancers (P < 0.001;** [Fig. 2C; Supplementary Fig. 3B]**). The functional relevance of those tissue-specific hypomethylated CpGs was further validated by comparison to additional fetal samples** [15] **and adult somatic tissues** [6] **from available external datasets** [Supplementary Fig. 3C]**. Intriguingly, a high degree of similarity between tissues sharing the same origin, even into adulthood, was observed** [Supplementary Fig. 3C]**.**

We next investigated whether tissue-specific hypomethylated CpGs clustered into Hypomethylated Regions (tHRs, defined as 3 consecutive hypomethylated CpGs within 1 kb of each other) [6]**. This was the case for amnion, muscle and pancreas** [Table 1, Supplementary Table 2]**. tHRs comprise robust development-independent epigenetic markers as exemplified by the mapping of pancreatic tHRs to proximal promoters of the nearest genes *ACY3*, *HNF1A*, and *HNF4A*** [Table 1, Supplementary Fig. 3D]**, genes with a key role in pancreas development** [50, 51]**. Muscle tHRs** [Supplementary Table 2] **mapped to distal elements of transcription factors involved in muscle development (*NFATC1*) and somitogenesis (*UNCX*)** [52, 53] [Supplementary Fig. 3D; Table 1; Supplementary Table 2]**. Importantly, we could confirm the tHRs identified with the relatively sparse Illumina 450k array with fetal and adult muscle whole-genome bisulfite sequencing (WGBS) data** [Supplementary Fig. 3E] [25]**. These data indicate that it is feasible to use combinations of tHRs as tissue-specific and development-independent barcodes.**

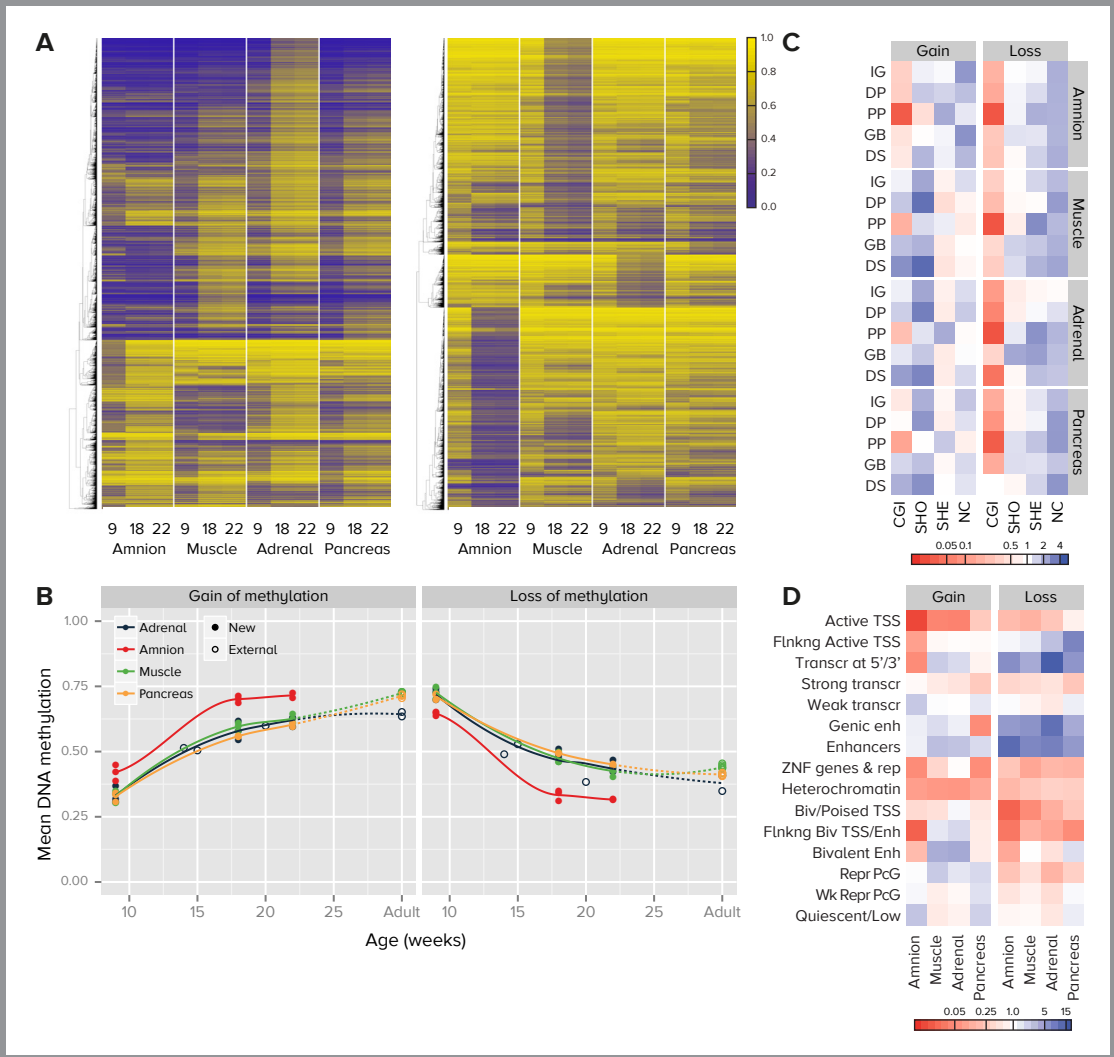
Tissue	Gene	Feature	Function
Amnion (67 tHRs)	*SLC22A2*	PP	Tubular uptake of organic compounds from circulation
	VTCN1	PP	B7 costimulatory protein family
	SLC39A2	PP	Zinc, iron, and calcium homeostasis
Muscle (15 tHRs)	*UNCX*	IG	TF involved in somitogenesis and neurogenesis
	NFATC1	IG	Involved in SM development/differentiation
	DPT	PP	Extracellular structure organization
Pancreas (3 tHRs)	*ACY3*	PP	Aminoacylase activity
	HNF1A	PP	Transcriptional activator
	HNF4A	PP	Transcriptional activator

[Table 1]

Table with the numbers of tHRs and three representative genes per tissue associated with tHRs. SM, skeletal muscle; TF, transcription factor.

DISTINCT ROLES FOR GAIN AND LOSS OF DNA METHYLATION IN FETAL DEVELOPMENT

We provide evidence for large-scale DNA methylation dynamics between W9 and W22 (DNA methylation difference > 20%) that affected 11.5% of evaluated CpGs (52,134/452,490). Approximately equal numbers of CpGs showed a gain of methylation (GOM) (26,555 CpGs; amnion 5,988; muscle 7,631; adrenal 13,997; pancreas 8,620) and a loss of methylation (LOM) (25,579 CpGs; amnion 10,811; muscle 11,925; adrenal 4,476; pancreas 3,286) [Fig. 3A]. **DNA methylation remodelling occurred predominantly between W9 and W18 and not between W18 and adulthood** [Fig. 3B; Supplementary Fig. 4A]**. Intriguingly, the integration and re-analysis of external DNA methylation data of fetal adrenal, brain and liver** [6, 15, 22, 43, 44] **revealed a striking confirmation of DNA methylation dynamics during fetal development. Furthermore, for all embryonic tissues, the DNA methylation levels at W22 were similar to those**


A
9 18 22 9 18 22 9 18 22 9 18 22
Amnion Muscle Adrenal Pancreas
1.0
0.8
0.6
0.4
0.2
0.0
C
Gain
Loss
IG
DP
PP
GB
DS
Amnion
Muscle
Adrenal
Pancreas
CGI SHO SHE NC CGI SHO SHE NC
0.05 0.1 0.5 1 2 4
B
Gain of methylation
Loss of methylation
Adrenal
Amnion
Muscle
Pancreas
New
External
Mean DNA methylation
1.00
0.75
0.50
0.25
0.00
10 15 20 25 Adult
Age (weeks)
D
Gain
Loss
Active TSS
Flnkng Active TSS
Transcr at 5'/3'
Strong transcr
Weak transcr
Genic enh
Enhancers
ZNF genes & rep
Heterochromatin
Biv/Poised TSS
Flnkng Biv TSS/Enh
Bivalent Enh
Repr PcG
Wk Repr PcG
Quiescent/Low
Amnion Muscle Adrenal Pancreas Amnion Muscle Adrenal Pancreas
0.05 0.25 1.0 5 15

[Fig. 3]

Gain and loss of DNA methylation during human fetal development.

[A] Heatmap of CpGs with a gain and a loss, respectively, of methylation over time. Gain and loss of methylation was defined as a difference of beta ≥ 0.2 between W9 and W22, and W18 in between. **[B]** Mean methylation of CpGs with a gain or loss of DNA methylation for fetal tissues and their adult counterpart. **[C]** Combined genic and CGI-centric annotation for CpGs with a gain or a loss of methylation represented as the odds ratio (see Supplementary Fig. 4B for odds ratios). CGI, CpG island; DP, distal promoter; DS, downstream; GB, gene body; IG, intergenic; NC, non-CGI; PP, proximal promoter; SHE, shelves; SHO, shores. **[D]** Enrichment of dynamically methylated CpGs in the chromatin segmentation states for the matching tissues (fetal muscle, fetal adrenal, amnion and adult pancreatic islets; see Supplementary Fig. 4C for odds ratios).

found in the adult counterpart [Fig. 3B; Supplementary Fig. 4A]**, suggesting that the extent of changes after W22 are limited for these CpGs.**

GOM CpGs did not show tissue-specific patterns [Fig. 3A] **and, in line with this observation, often corresponded to genes involved in generic developmental and cellular processes, including embryonic morphogenesis and regulation of transcription** [Supplementary Table 3]**. In contrast, the LOM CpGs were highly tissue-specific** [Fig. 3A] **and mapped to genes involved in tissue-specific processes that matched the organ in which the LOM CpGs were identified** [Supplementary Table 3]**. CpGs that lost methylation in the amnion mapped, amongst others, to genes that were associated with the regulation of apoptosis and cytoskeleton organization; in the muscle to genes associated with cytoskeleton organization and muscle system processes; in the adrenal to genes associated with regulation of macromolecule metabolism** [Supplementary Table 3]**. In the pancreas no significant enrichments were found.**

GOM and LOM CpGs differed in their genomic annotation. While GOM CpGs were generally enriched in CGIs and CGI-shores, LOM CpGs were enriched for CGI-shelves and non-CGI regions [Fig. 3C; Supplementary Fig. 4B]**. LOM- and GOM-specific enrichments were also observed for Epigenomics Roadmap chromatin state segmentations. LOM CpGs were strongly enriched for (genic) enhancers and transcribed regions, whereas GOM CpGs were enriched for bivalent and repressed regions and only modestly at enhancers** [Fig. 3D; Supplementary Fig. 4C]**. The results underscore the relevance of DNA methylation in enhancer activity, in addition to the well-studied relationship between DNA methylation and promoter activity** [54]**.**

We previously reported on transcriptional data of amnion (n = 7), muscle (n = 6), adrenal (n = 3) and pancreas (n = 7) at W9, W18 and W22 [29] **and used this data to test the hypothesis that GOM is associated with the epigenetic down-regulation of developmental programs and LOM with upregula-**

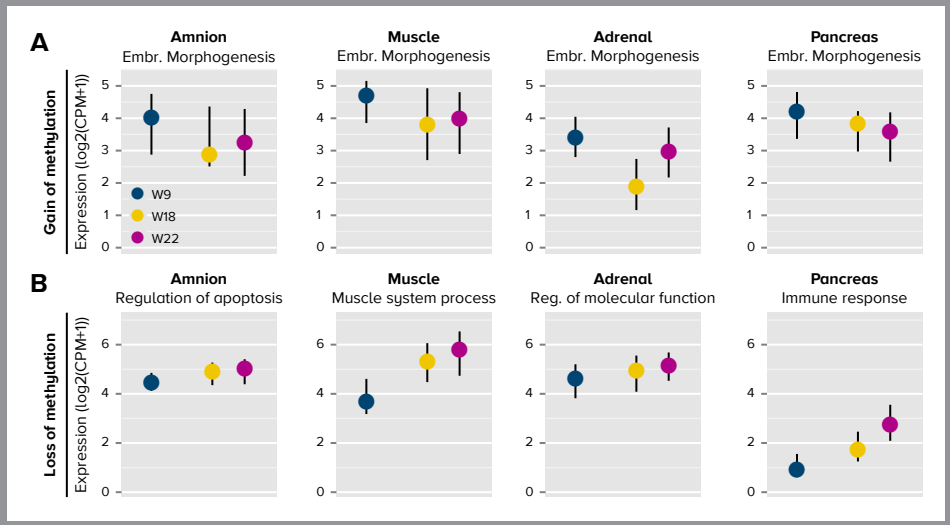
tion of tissue-specific processes. Genes associated with GOM and involved in embryonic morphogenesis (a process enriched for GOM in all tissues) showed a decrease in transcriptional activity from W9 to W22 in all tissues (amnion, muscle, pancreas: P < 0.05; adrenal: P = 0.87; [Fig. 4A]**). In contrast, genes involved in tissue-specific processes found to be enriched for LOM** [Supplementary Table 3] **increased in transcription from W9 to W22 (P < 0.05;** [Fig. 4B; Supplementary Fig. 4D]**).**

Altogether, these findings emphasize that DNA methylation dynamics during human fetal development is associated with the availability to transcription of both general embryonic programs (shutting those down for transcription) as well as tissue-specific developmental programs (making those available for transcription).

DYNAMICALLY METHYLATED REGIONS CORRELATE WITH DEVELOPMENTAL AND TISSUE-SPECIFIC GENES

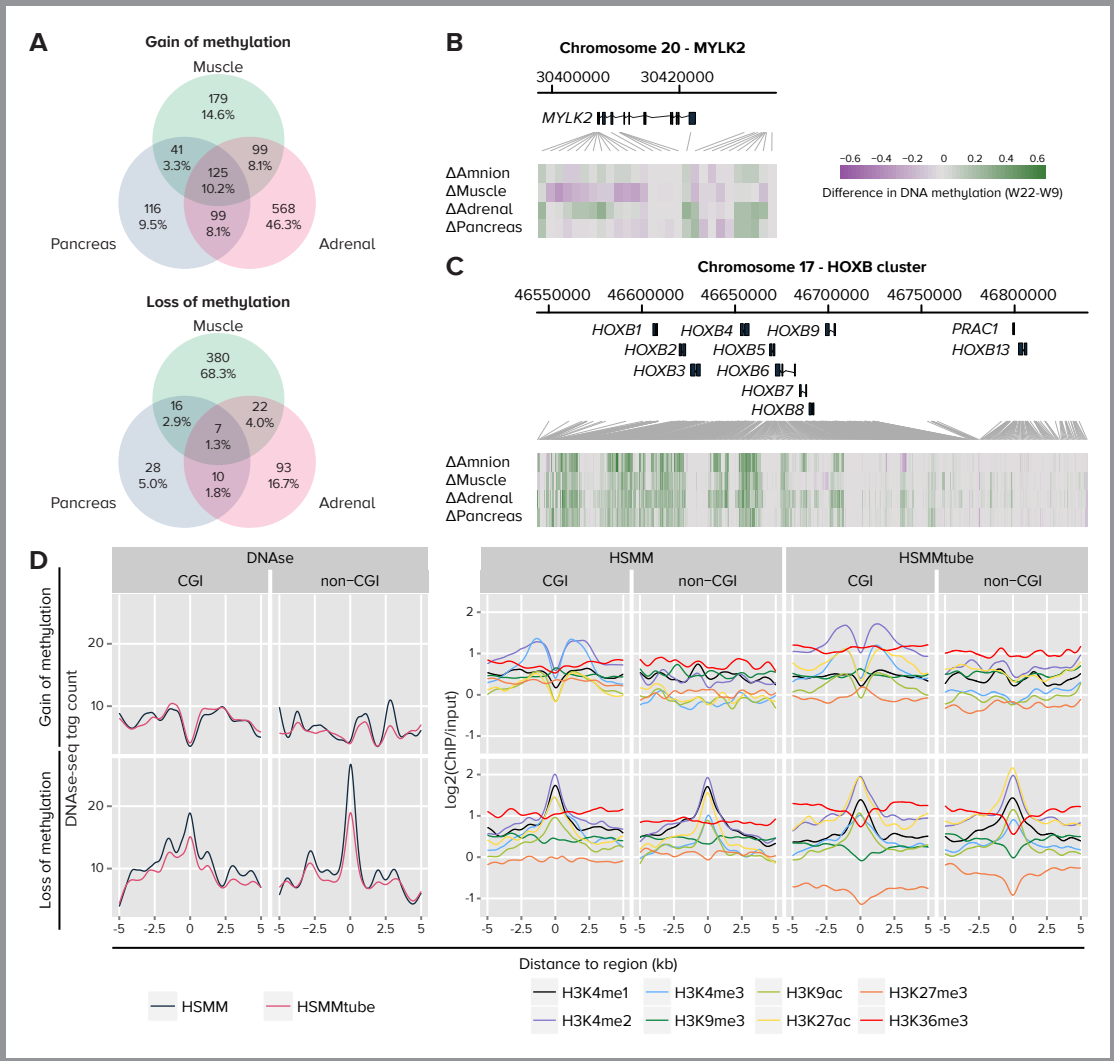
From the dynamically methylated CpGs, we identified 2,229 development-related differentially methylated regions (dDMRs, defined as 3 consecutive differentially methylated CpGs within 1 kb of each other) undergoing GOM (amnion 185; muscle 530; adrenal 1,065; pancreas 449) and 1,017 undergoing LOM (amnion 388; muscle 482; adrenal 136; pancreas 61; [Supplementary Table 4]**). After mapping the dDMRs to the nearest gene locus we observed that the percentage of common genes in the embryonic tissues showing LOM dDMRs was 1.3%, whereas those showing GOM was 10.2%** [Fig. 5A]**.**

LOM dDMRs were associated with genes involved in tissue-specific functions such as *MYH3* in muscle (muscle contractile protein), *MC2R* (adrenocorticotropic hormone receptor) in adrenal and *PFKFB3* (involved in insulin secretion) in the pancreas [Table 2; Supplementary Fig. 5A]**. As an example of tissue specificity encountered in the dDMRs showing a loss of DNA methylation, we zoomed in on the *MYLK2* locus, a mus-**


A
Amnion
Embr. Morphogenesis
Muscle
Embr. Morphogenesis
Adrenal
Embr. Morphogenesis
Pancreas
Embr. Morphogenesis
Gain of methylation
Expression (log2(CPM+1))
W9
W18
W22
B
Amnion
Regulation of apoptosis
Muscle
Muscle system process
Adrenal
Reg. of molecular function
Pancreas
Immune response
Loss of methylation
Expression (log2(CPM+1))

[Fig. 4]

DNA methylation dynamics are accompanied by changes in gene expression.
[A] Expression profiles of genes in embryonic morphogenesis near dynamic regions with gain of methylation represented as median with the interquartile range (IQR) [29]. **[B]** Expression profiles of genes near dynamic regions with loss of methylation grouped by significant, tissue-specific Gene Ontology terms for each of the four tissues from Supplementary Table 3 represented as median with IQR [29]. Reg., regulation.


A
Gain of methylation
Muscle
179
14.6%
41
3.3%
99
8.1%
125
10.2%
116
9.5%
99
8.1%
568
46.3%
Pancreas
Adrenal
Loss of methylation
Muscle
380
68.3%
16
2.9%
22
4.0%
7
1.3%
28
5.0%
10
1.8%
93
16.7%
Pancreas
Adrenal
B
Chromosome 20 - MYLK2
30400000
30420000
MYLK2
ΔAmnion
ΔMuscle
ΔAdrenal
ΔPancreas
−0.6 −0.4 −0.2 0 0.2 0.4 0.6
Difference in DNA methylation (W22-W9)
C
Chromosome 17 - HOXB cluster
46550000
46600000
46650000
46700000
46750000
46800000
HOXB1
HOXB4
HOXB9
PRAC1
HOXB2
HOXB5
HOXB13
HOXB3
HOXB6
HOXB7
HOXB8
ΔAmnion
ΔMuscle
ΔAdrenal
ΔPancreas
D
DNAse
CGI
non−CGI
HSMM
HSMMtube
Gain of methylation
Loss of methylation
DNAse-seq tag count
log2(ChIP/input)
Distance to region (kb)
HSMM
HSMMtube
H3K4me1
H3K4me2
H3K4me3
H3K9me3
H3K9ac
H3K27ac
H3K27me3
H3K36me3

[Fig. 5]

Association of gain and loss of DNA methylation, DNAse I hypersensitive sites and histone modifications.

[A] Venn diagram visualizing the overlaps between genes with a gain and a loss of methylation of the three embryonic tissues. **[B]** Methylation difference between W9 and W22 of *MYLK2* in the four tissues. **[C]** Methylation difference between W9 and W22 of the *HOXB* cluster. **[D]** Mean DNAse I hypersensitive (DHS) and histone modifications signals in a 5 kb flanking region of the muscle dDMRs in HSMMs and HSMMtubes. HSMM, human skeletal muscle myoblasts; HSMMtube, human skeletal muscle myotubes.

cle-specific gene [55]**. The methylation of the *MYLK2* promoter and first exon in the muscle decreased during development, but increased (or remained constant) in the other organs studied** [Fig. 5B]**. Interestingly, GOM dDMRs were associated with (tissue-specific) developmental genes, such as *PAX3* in muscle, a key gene in myogenesis** [56]**, and *NKX6.1* in pancreas, an important gene in beta-cell development** [57] [Table 2; Supplementary Fig. 5A]**, but also near well-known developmental genes including the *HOXB*** [Fig. 5C] **and other *HOX* clusters** [Supplementary Fig. 6A] **that play a key role in embryonic patterning and morphogenesis** [58]**.**

However, when comparing all identified dDMRs to previously identified adult (tissue-specific) tDMRs using the 450k array [6]**, about 50% of the GOM dDMRs were not identified as tDMRs in adult tissues** [Supplementary Fig. 5B]**, while 32% and 38%, of the LOM dDMRs in muscle and pancreas, respectively, were unique for those fetal tissues (adult data on adrenal was absent). The persistence into adulthood of subsets of GOM and LOM dDMRs was confirmed using WGBS data for adult muscle** [25] [Supplementary Fig. 5C]**. These results suggest that the study of DNA methylation dynamics in fetal development will identify regions that are remodelled during development and are missed when studying adult tissues only.**

We further explored the potential biological validity of the 1,012 muscle dDMRs using ENCODE data [24] **on human skeletal muscle myoblasts (HSMMs) and their differentiated derivatives, human skeletal muscle myotubes (HSMMtubes). In HSMMs and HSMMtubes, DNAse I hypersensitive sites (DHSs), which mark genomic regions of open chromatin associated with transcriptional activity, were abundant at LOM dDMRs, particularly in CpG-poor regions (CGI-shelves and non-CGI regions,** [Fig. 5D, left]**). DHSs were depleted at GOM dDMRs in CpG-rich regions (CGIs and CGI-shores,** [Fig. 5D, left]**). Consistent with an increased transcriptional activity, LOM dDMRs were also enriched in myotubes and myoblasts**

	Gain of methylation			Loss of methylation		
Tissue	Gene	Features	Function	Gene	Features	Function
Amnion	*TNXB*	GB	Cell adhesion	*TNXB*	PP/GB	Cell adhesion
	TFAP2A	IG/DP	Cell differentiation	*DIP2C*	GB	Ectoderm development
	TFAP2B	GB	Cell proliferation	*CD59*	PP	Lymphocyte signal transduction
Muscle	*SIX3*	IG/DP/GB/DS	Transcription factor in repression of WNT	*MYLK2*	PP	Myosin light chain kinase
	HLX	DP/GB/DS	Homeobox TF factor in muscle development	*MYOZ1*	PP	Calcineurin signaling in muscle
	PAX3	GB/PP	Muscle development	*MYH3*	PP	Muscle contractile protein
Adrenal	*TBX3*	IG/DP/DS	TF in developmental processes	*KCNQ1*	PP	Potassium channel protein
	NR2F2	IG/DS	Steroid thyroid hormone nuclear receptor	*MC2R*	PP	Adrenocorticotropin receptor
	KCNQ1	PP/GB	Potassium channel protein	*SEC14L1*	PP	Intracellular transport
Pancreas	*NKX6.1*	DP	Beta cell development	*SLC25A22*	PP	Glucose responsiveness
	PROX1	DP	Co-repressor of HNF4A	*FAIM3*	PP	Promotes β-cell proliferation
	PRDM16	GB	Transcription factor activity	*PFKFB3*	GB	Insulin secretion

[Table 2]

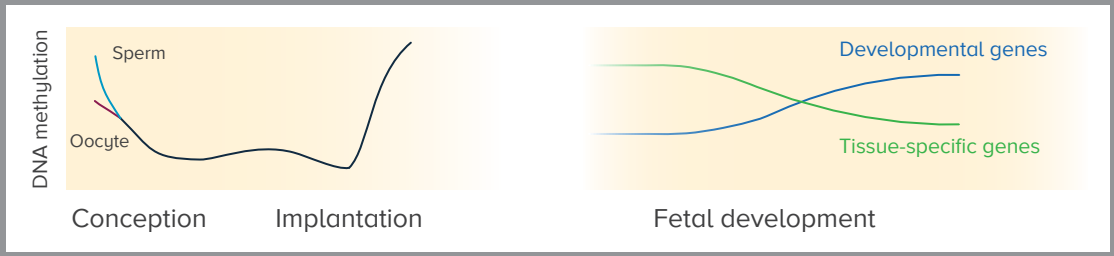
Table with six genes per tissue highlighting the tissue specificity of the genes found near dDMRs with a loss of methylation as well as the association of dDMRs with a gain of methylation with tissue-specific developmental genes.

(ENCODE [24]**) for histone H3 lysine 4 methylation ((H3K4me1, -me2, -me3)** [25, 59]**, and acetylation of histone H3 at lysine 9 and 27 (H3K9ac, H3K27ac), all marks associated with active regulatory regions** [25, 59] [Fig. 5D, right]**. In contrast, these active histone modifications were depleted for GOM dDMRs in CpG-rich regions** [Fig. 5D, right]**. LOM dDMRs were depleted of H3K9me3 (marking inactive DNA), H3K27me3 (marking Polycomb-repressed regions) and H3K36me3 in HSMMtubes but not in their precursor cells HSMMs** [Fig. 5D, right]**. A final indication for the functional relevance of the muscle dDMRs was that 124 out of 482 LOM dDMRs significantly overlapped (P < 0.0001) with binding sites of the muscle-specific transcription factor *MYOD* in HSMMs (188/482 in HSMMtubes), whereas only 7 out of the 530 GOM dDMRs mapped to *MYOD* binding sites** [45] **(8/530 in HSMMtubes,** [Supplementary Fig. 5D]**).**

DISCUSSION

Here, we show that human tissues already exhibit a specific DNA methylation signature as early as W9 of fetal development. In addition, the DNA methylation landscape is subjected to considerable changes from the first to second trimester of gestation as the developing organs gain complexity and functionality. Our study highlights that dynamic DNA methylation is not only an integral part of early preimplantation embryo development and implantation [9–11]**, but continues to be a key feature of epigenetic remodelling during human fetal development. While global changes in levels of DNA methylation characterize development until implantation** [Fig. 6]**, these are not observed during fetal development. Instead, distinct LOM occurs near tissue-specific genes and GOM occurs near developmental genes in a largely tissue-independent fashion** [Fig. 6]**. Our direct assessment of DNA methylation dynamics suggests that a larger proportion of the methylome is remodelled during development than previously thought** [3, 6, 60]**.**

Interestingly, the functional relevance of identified dynamic regions was further exemplified by the changes in expression of their nearest genes. While the nearest genes of regions gaining DNA methylation associated with embryonic morphogenesis showed loss of expression, the nearest genes of regions losing DNA methylation showed increased expression over time. In agreement with our observations, LOM of hematopoietic-specific genes has been observed during human hematopoietic differentiation *in vitro* [61] **and have been linked to transcriptional changes in human T-cell development** [62]**. Moreover, several mouse and human *in vitro* studies demonstrated that the methylation of developmental genes increases** [19, 63] **and tissue-specific functional genes lose methylation** [5, 64] **during stem and progenitor cell differentiation.**



[Fig. 6]

DNA methylation dynamics during human development.

This illustration depicts the current comprehensive knowledge of DNA methylation during human pre- and postimplantation development. The knowledge about DNA methylation during human preimplantation (left panel) is derived from [9-12], whereas our study sheds light on postimplantation development (right panel).

Lastly, DNA demethylating agents, such as 5-azacytidine, have been shown to promote stem cell differentiation and maturation of skeletal myotubes in mice [65, 66]**. Further experimental studies are required to evaluate the mechanistic role of DNA methylation in development.**

Although our study reveals general principles of DNA methylation dynamics during human fetal development, it should be noted that a limited number of tissues and individuals was investigated; and that we used a genome-wide method interrogating a relatively small proportion of all CpGs in the human genome. Expansion to more tissues and the application of whole-methylome technologies will lead to a more comprehensive catalogue of regulatory regions. However, by extensive inclusion of external fetal and adult 450k array datasets, we have consolidated our findings. Moreover, the use of external available WGBS data confirmed the results obtained by the 450k array data.

Since we studied organ biopsies similar to previous studies investigating biopsies of human adult tissues [3, 15]**, the methylation profiles we report reflect the average of multiple cell types. The cellular complexity of the organs investigated led to an underestimation of the actual DNA methylation dynamics in individual cell types. This is exemplified by the detection of a considerably larger number of CpGs displaying dynamic methylation in muscle which has an exclusive mesodermal origin in comparison with adrenal and pancreas which are composed by cells originating from two different germ layers (adrenal: mesoderm and ectoderm (neural crest); pancreas: endoderm and mesoderm). However, it is unlikely that the methylation dynamics observed is an epiphenomenon of this cellular complexity instead of being driven by cell differentiation and maturation. This is obvious for genes associated with GOM that appears to be shared across organs to repress general developmental programs during development. In contrast, genes associated with LOM displayed tissue-specific patterns. Their**

intricate involvement in organ-specific functions was emphasized by tight linkage to biological processes and chromatin states relevant to the organs investigated. Moreover, between W9 and W22, the organs analysed are mainly composed of progenitor cells; perfusion by blood and lymphatic vasculature, and innervation by neural crest cell derivatives still plays a minor role as compared with adult organs [Supplementary Fig. 1A]**. In the future, single-cell methodology** [67, 68] **will enable comparing single-cell DNA methylomes of the various adult cell types to their fetal progenitor counterparts.**

Studies of DNA methylation landscapes of human fetal development may serve as reference in the development of (organoid) differentiation models [69] **and, moreover, shed light on potential mechanisms underlying genetic associations and studies in the field of epigenetic epidemiology** [70] **focussing on the prenatal environment.**

SUPPLEMENTARY INFORMATION

The Supplementary Tables in this chapter can be found online at http://dx.plos.org/10.1371/journal.pgen.1005583.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: S.M.C.d.S.L., B.T.H., M.S.R., R.C.S., F.C., E.J.P.d.K., P.E.S.. Performed the experiments: M.S.R., L.v.I., E.W.T., H.E.D.S., R.C.S.. Analyzed the data: R.C.S., M.S.R., B.T.H., S.M.C.d.S.L.. Wrote the paper: R.C.S., M.S.R., B.T.H., S.M.C.d.S.L..

REFERENCES

[1] Byun HM, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, Yang AS: **Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns**. *Hum Mol Genet* 2009, **18**(24):4808–4817.

[2] Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M *et al*: **An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs)**. *Genome Res* 2008, **18**(9):1518–1529.

[3] Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE *et al*: **Charting a dynamic DNA methylation landscape of the human genome**. *Nature* 2013, **500**(7463):477–481.

[4] Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B: **Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues**. *Nat Genet* 2013, **45**(10):1198–1206.

[5] Nagae G, Isagawa T, Shiraki N, Fujita T, Yamamoto S, Tsutsumi S, Nonaka A, Yoshiba S, Matsusaka K, Midorikawa Y *et al*: **Tissue-specific demethylation in CpG-poor promoters during cellular differentiation**. *Hum Mol Genet* 2011, **20**(14):2710–2721.

[6] Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, Suchiman HE, Lameijer EW, Putter H, van den Akker EB *et al*: **Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array**. *Epigenetics Chromatin* 2013, **6**(1):26.

[7] Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R *et al*: **Transcriptional and epigenetic dynamics during specification of human embryonic stem cells**. *Cell* 2013, **153**(5):1149–1163.

[8] Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C,

Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y *et al*: **Conserved role of intragenic DNA methylation in regulating alternative promoters**. *Nature* 2010, **466**(7303):253–257.

[9] Smith ZD, Chan MM, Humm KC, Karnik R, Mekhoubad S, Regev A, Eggan K, Meissner A: **DNA methylation dynamics of the human preimplantation embryo**. *Nature* 2014, **511**(7511):611–615.

[10] Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, Yan J, Ren X, Lin S, Li J *et al*: **The DNA methylation landscape of human early embryos**. *Nature* 2014, **511**(7511):606–610.

[11] Okae H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, Kikuchi H, Yoshida H, Tanaka A, Suyama M *et al*: **Genome-wide analysis of DNA methylation dynamics during early human development**. *PLoS Genet* 2014, **10**(12):e1004868.

[12] Seisenberger S, Peat JR, Hore TA, Santos F, Dean W, Reik W: **Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers**. *Philos Trans R Soc Lond B Biol Sci* 2013, **368**(1609):20110330.

[13] Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A: **A unique regulatory phase of DNA methylation in the early mammalian embryo**. *Nature* 2012, **484**(7394):339–344.

[14] Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W: **The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells**. *Mol Cell* 2012, **48**(6):849–862.

[15] Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Muller FJ, Wang YC, Boscolo FS *et al*: **Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives**. *Cell Stem Cell* 2012, **10**(5):620–634.

[16] Bocker MT, Hellwig I, Breiling A, Eckstein V, Ho AD, Lyko F: **Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging**. *Blood* 2011, **117**(19):e182–189.

[17] Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D *et al*: **Epigenomic analysis of multilineage differentiation of human embryonic stem cells**. *Cell* 2013,

153(5):1134–1148.

[18] Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E *et al*: **Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver**. *Genome Res* 2009, **19**(6):1044–1056.

[19] Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J *et al*: **Dynamic changes in the human methylome during differentiation**. *Genome Res* 2010, **20**(3):320–331.

[20] Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A: **DNA methylation dynamics during in vivo differentiation of blood and skin stem cells**. *Mol Cell* 2012, **47**(4):633–647.

[21] Gilsbach R, Preissl S, Gruning BA, Schnick T, Burger L, Benes V, Wurch A, Bonisch U, Gunther S, Backofen R *et al*: **Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease**. *Nature communications* 2014, **5**:5288.

[22] Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CC, O'Donovan MC, Bray NJ, Mill J: **Methylomic trajectories across human fetal brain development**. *Genome Res* 2015, **25**(3):338–352.

[23] Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL *et al*: **High density DNA methylation array with single CpG site resolution**. *Genomics* 2011, **98**(4):288–295.

[24] The ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**(7414):57–74.

[25] Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al*: **Mapping and analysis of chromatin state dynamics in nine human cell types**. *Nature* 2011, **473**(7345):43–49.

[26] Hou L, Zhao H: **A review of post-GWAS prioritization approaches**. *Frontiers in genetics* 2013, **4**:280.

[27] Waterland RA, Michels KB: **Epigenetic epidemiology of the developmental origins hypothesis**. *Annual review of nutrition* 2007, **27**:363–388.

[28] Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, Susser ES, Slagboom PE, Lumey LH: **Persistent epigenetic differences associated with prenatal exposure to famine in humans**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(44):17046–17049.

[29] Roost MS, van Iperen L, Ariyurek Y, Buermans HP, Arindrarto W, Devalla HD, Passier R, Mummery CL, Carlotti F, de Koning EJ *et al*: **KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas**. *Stem Cell Reports* 2015, **4**(6):1112–1124.

[30] Hansen KD, Aryee M: **Minfi: Analyze Illumina's 450K methylation arrays.** *R package version* 2013, **1.16.0**.

[31] Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R: **Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray**. *Epigenetics* 2013, **8**(2):203–209.

[32] Du P, Kibbe WA, Lin SM: **lumi: a pipeline for processing Illumina microarray**. *Bioinformatics* 2008, **24**(13):1547–1548.

[33] Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S: **A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data**. *Bioinformatics* 2013, **29**(2):189–196.

[34] Feber A, Guilhamon P, Lechner M, Fenton T, Wilson GA, Thirlwell C, Morris TJ, Flanagan AM, Teschendorff AE, Kelly JD *et al*: **Using high-density DNA methylation arrays to profile copy number alterations**. *Genome Biol* 2014, **15**(2):R30.

[35] Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, Beck S: **ChAMP: 450k Chip Analysis Methylation Pipeline**. *Bioinformatics* 2014, **30**(3):428–430.

[36] Koenker R: **quantreg: Quantile Regression**; 2013.

[37] Wickham H: **ggplot2: Elegant Graphics for Data Analysis**: Springer New York; 2009.

[38] Durinck S, Bullard J: **GenomeGraphs: Plotting genomic information from Ensembl. R package version 1.22.0**.

[39] Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**.

Nat Protoc 2009, **4**(1):44–57.

[40] Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**(1):139–140.

[41] Robinson MD, Smyth GK: **Small-sample estimation of negative binomial dispersion, with applications to SAGE data**. *Biostatistics* 2008, **9**(2):321–332.

[42] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M *et al*: **NCBI GEO: archive for functional genomics data sets--update**. *Nucleic acids research* 2013, **41**(Database issue):D991–995.

[43] Bonder MJ, Kasela S, Kals M, Tamm R, Lokk K, Barragan I, Buurman WA, Deelen P, Greve JW, Ivanov M *et al*: **Genetic and epigenetic regulation of gene expression in fetal and adult human livers**. *BMC genomics* 2014, **15**:860.

[44] Pidsley R, Viana J, Hannon E, Spiers H, Troakes C, Al-Saraj S, Mechawar N, Turecki G, Schalkwyk LC, Bray NJ *et al*: **Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia**. *Genome Biol* 2014, **15**(10):483.

[45] MacQuarrie KL, Yao Z, Fong AP, Diede SJ, Rudzinski ER, Hawkins DS, Tapscott SJ: **Comparison of genome-wide binding of MyoD in normal human myogenic cells and rhabdomyosarcomas identifies regional and local suppression of promyogenic transcription factors**. *Molecular and cellular biology* 2013, **33**(4):773–784.

[46] Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M: **Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome**. *Epigenetics* 2011, **6**(6):692–702.

[47] Jones PA: **Functions of DNA methylation: islands, start sites, gene bodies and beyond**. *Nat Rev Genet* 2012, **13**(7):484–492.

[48] Colaneri A, Wang T, Pagadala V, Kittur J, Staffa NG, Jr., Peddada SD, Isganaitis E, Patti ME, Birnbaumer L: **A minimal set of tissue-specific hypomethylated CpGs constitute epigenetic signatures of developmental programming**. *PLoS One* 2013, **8**(9):e72670.

[49] Song F, Mahmood S, Ghosh S, Liang P, Smiraglia DJ, Nagase H, Held

WA: **Tissue specific differentially methylated regions (TDMR): Changes in DNA methylation during development**. *Genomics* 2009, **93**(2):130–139.

[50] Oliver-Krasinski JM, Stoffers DA: **On the origin of the beta cell**. *Genes Dev* 2008, **22**(15):1998–2021.

[51] Brink C: **Promoter elements in endocrine pancreas development and hormone regulation**. *Cellular and molecular life sciences : CMLS* 2003, **60**(6):1033–1048.

[52] Leitges M, Neidhardt L, Haenig B, Herrmann BG, Kispert A: **The paired homeobox gene Uncx4.1 specifies pedicles, transverse processes and proximal ribs of the vertebral column**. *Development* 2000, **127**(11):2259–2267.

[53] Schulz RA, Yutzey KE: **Calcineurin signaling and NFAT activation in cardiovascular and skeletal muscle development**. *Developmental biology* 2004, **266**(1):1–16.

[54] Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ *et al*: **DNA methylation status predicts cell type-specific enhancer activity**. *EMBO J* 2011, **30**(15):3028–3039.

[55] Stull JT, Kamm KE, Vandenboom R: **Myosin light chain kinase and the role of myosin light chain phosphorylation in skeletal muscle**. *Archives of biochemistry and biophysics* 2011, **510**(2):120–128.

[56] Bismuth K, Relaix F: **Genetic regulation of skeletal muscle development**. *Exp Cell Res* 2010, **316**(18):3081–3086.

[57] Nelson SB, Schaffer AE, Sander M: **The transcription factors Nkx6.1 and Nkx6.2 possess equivalent activities in promoting beta-cell fate specification in Pdx1+ pancreatic progenitor cells**. *Development* 2007, **134**(13):2491–2500.

[58] Heffer A, Pick L: **Conservation and variation in Hox genes: how insect models pioneered the evo-devo field**. *Annual review of entomology* 2013, **58**:161–179.

[59] Zhou VW, Goren A, Bernstein BE: **Charting histone modifications and the functional organization of mammalian genomes**. *Nat Rev Genet* 2011, **12**(1):7–18.

[60] Lokk K, Modhukur V, Rajashekar B, Martens K, Magi R, Kolde R,

Koltsina M, Nilsson TK, Vilo J, Salumets A *et al*: **DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns**. *Genome Biol* 2014, **15**(4):r54.

[61] Calvanese V, Fernandez AF, Urdinguio RG, Suarez-Alvarez B, Mangas C, Perez-Garcia V, Bueno C, Montes R, Ramos-Mejia V, Martinez-Camblor P *et al*: **A promoter DNA demethylation landscape of human hematopoietic differentiation**. *Nucleic acids research* 2012, **40**(1):116–131.

[62] Rodriguez RM, Suarez-Alvarez B, Mosen-Ansorena D, Garcia-Peydro M, Fuentes P, Garcia-Leon MJ, Gonzalez-Lahera A, Macias-Camara N, Toribio ML, Aransay AM *et al*: **Regulation of the transcriptional program by DNA methylation during human alphabeta T-cell development**. *Nucleic acids research* 2015, **43**(2):760–774.

[63] Kim M, Park YK, Kang TW, Lee SH, Rhee YH, Park JL, Kim HJ, Lee D, Lee D, Kim SY *et al*: **Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage**. *Hum Mol Genet* 2014, **23**(3):657–667.

[64] Miyata K, Miyata T, Nakabayashi K, Okamura K, Naito M, Kawai T, Takada S, Kato K, Miyamoto S, Hata K *et al*: **DNA methylation analysis of human myoblasts during in vitro myogenic differentiation: de novo methylation of promoters of muscle-related genes and its involvement in transcriptional down-regulation**. *Hum Mol Genet* 2015, **24**(2):410–423.

[65] Hupkes M, Jonsson MK, Scheenen WJ, van Rotterdam W, Sotoca AM, van Someren EP, van der Heyden MA, van Veen TA, van Ravestein-van Os RI, Bauerschmidt S *et al*: **Epigenetics: DNA demethylation promotes skeletal myotube maturation**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2011, **25**(11):3861–3872.

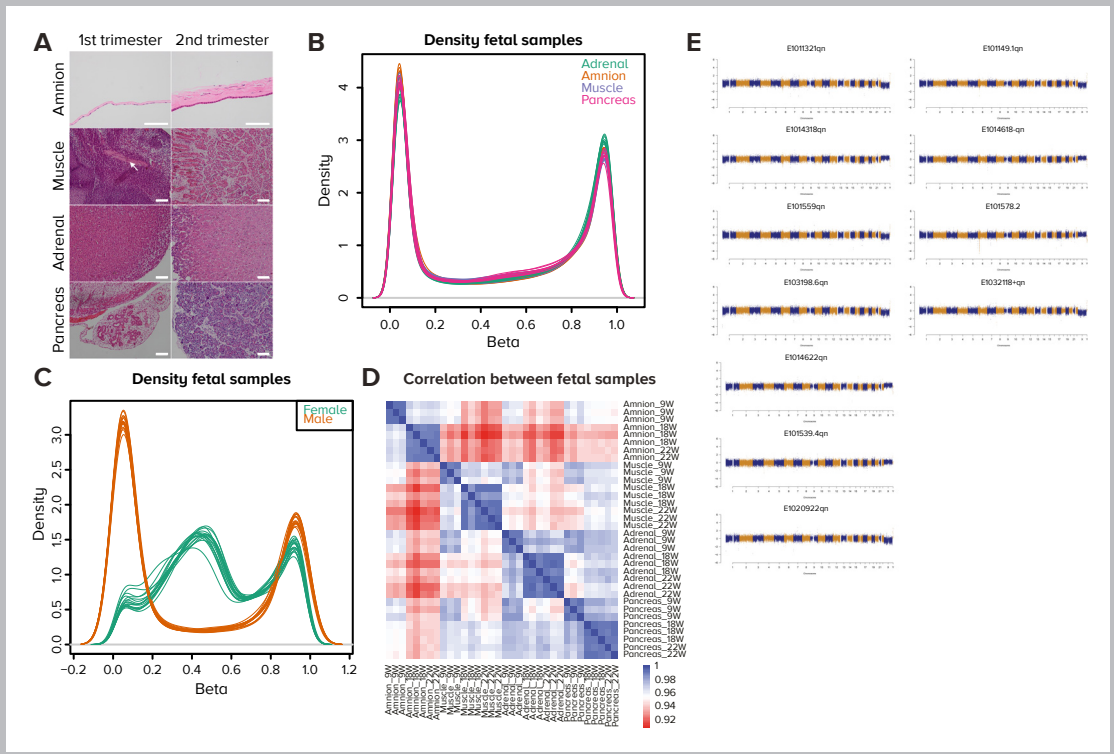
[66] Zhou Y, Hu Z: **Genome-wide demethylation by 5-aza-2'-deoxycytidine alters the cell fate of stem/progenitor cells**. *Stem cell reviews* 2015, **11**(1):87–95.

[67] Shapiro E, Biezuner T, Linnarsson S: **Single-cell sequencing-based technologies will revolutionize whole-organism science**. *Nat Rev Genet* 2013, **14**(9):618–630.

[68] Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G: **Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity**. *Nature methods* 2014, **11**(8):817–820.

[69] Lancaster MA, Knoblich JA: **Organogenesis in a dish: modeling development and disease using organoid technologies**. *Science* 2014, **345**(6194):1247125.

[70] Mill J, Heijmans BT: **From promises to practical strategies in epigenetic epidemiology**. *Nat Rev Genet* 2013, **14**(8):585–594.


A
1st trimester
2nd trimester
Amnion
Muscle
Adrenal
Pancreas
B
Density fetal samples
Adrenal
Amnion
Muscle
Pancreas
Density
Beta
C
Density fetal samples
Female
Male
Density
Beta
D
Correlation between fetal samples
E

[Supplementary Fig. 1]

Quality control.

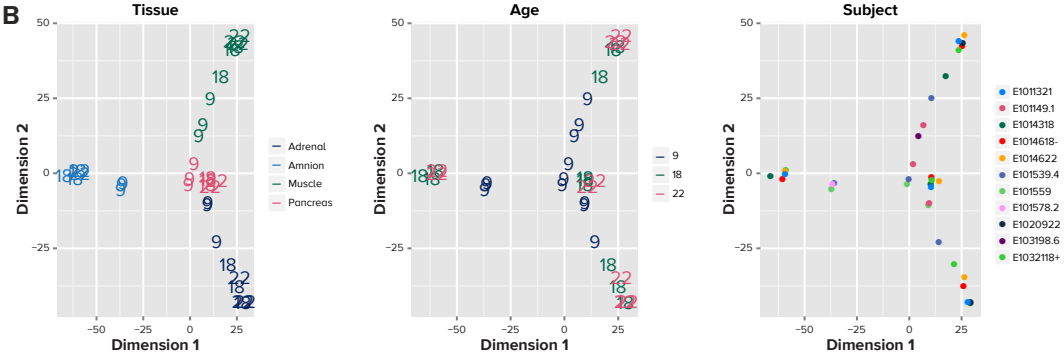
[A] Histology of the four tissues used in this study during development stained with Haematoxylin and Eosin (H&E). White arrow points to first trimester muscle. Scale bars: 100 µm. **[B]** Density plot of the data per sample coloured by tissue. **[C]** Density plot of the sex chromosomes. **[D]** Pearson correlation between the biological replicates. The highest correlation was found between tissues and time points. **[E]** Assessment of chromosomal abnormalities.

A

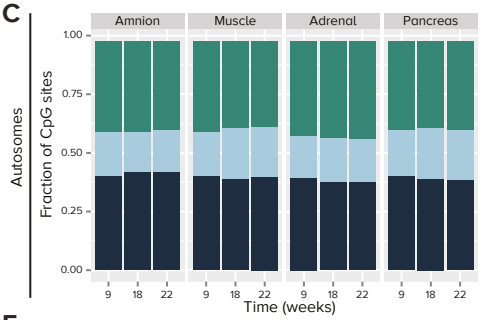
Tissue	ID	Age	Age (true)	Gender
Amnion	E101559	9	9	Male
	E101539.4	9	9.4	Female
	E101578.2	9	8.2	Female
	E1014618-	18	18	Female
	E1014318	18	18	Female
	E1032118+	18	18	Male
	E1014622	22	22	Female
	E1011321	22	21	Male
Muscle	E101149.1	9	9.1	Male
	E101539.4	9	9.4	Female
	E103198.6	9	8.6	Male
	E1014618-	18	18	Female
	E1014318	18	18	Female
	E1032118+	18	18	Male
	E1014622	22	22	Female
	E1020922	22	22	Male
	E1011321	22	21	Male

Tissue	ID	Age	Age (true)	Gender
Pancreas	E101559	9	9	Male
	E101149.1	9	9.1	Male
	E101539.4	9	9.4	Female
	E1014618-	18	18	Female
	E1014318	18	18	Female
	E1032118+	18	18	Male
	E1014622	22	22	Female
	E1011321	22	21	Male
Adrenal	E101559	9	9	Male
	E101539.4	9	9.4	Female
	E101149.1	9	9.1	Male
	E1014618-	18	18	Female
	E1014318	18	18	Female
	E1032118+	18	18	Male
	E1014622	22	22	Female
	E1020922	22	22	Male
	E1011321	22	21	Male

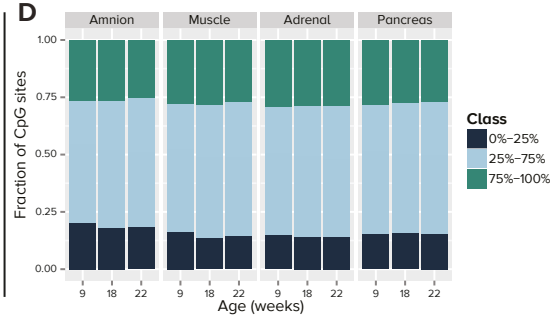
B



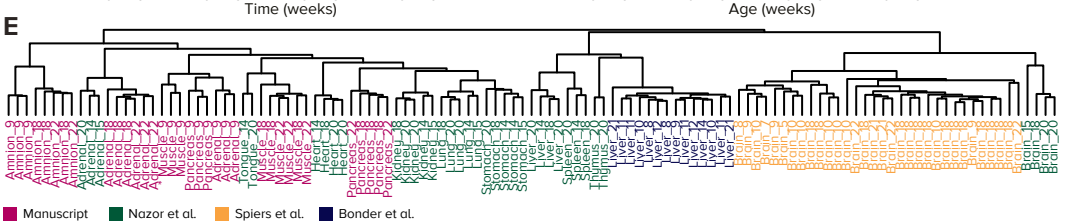
C



D



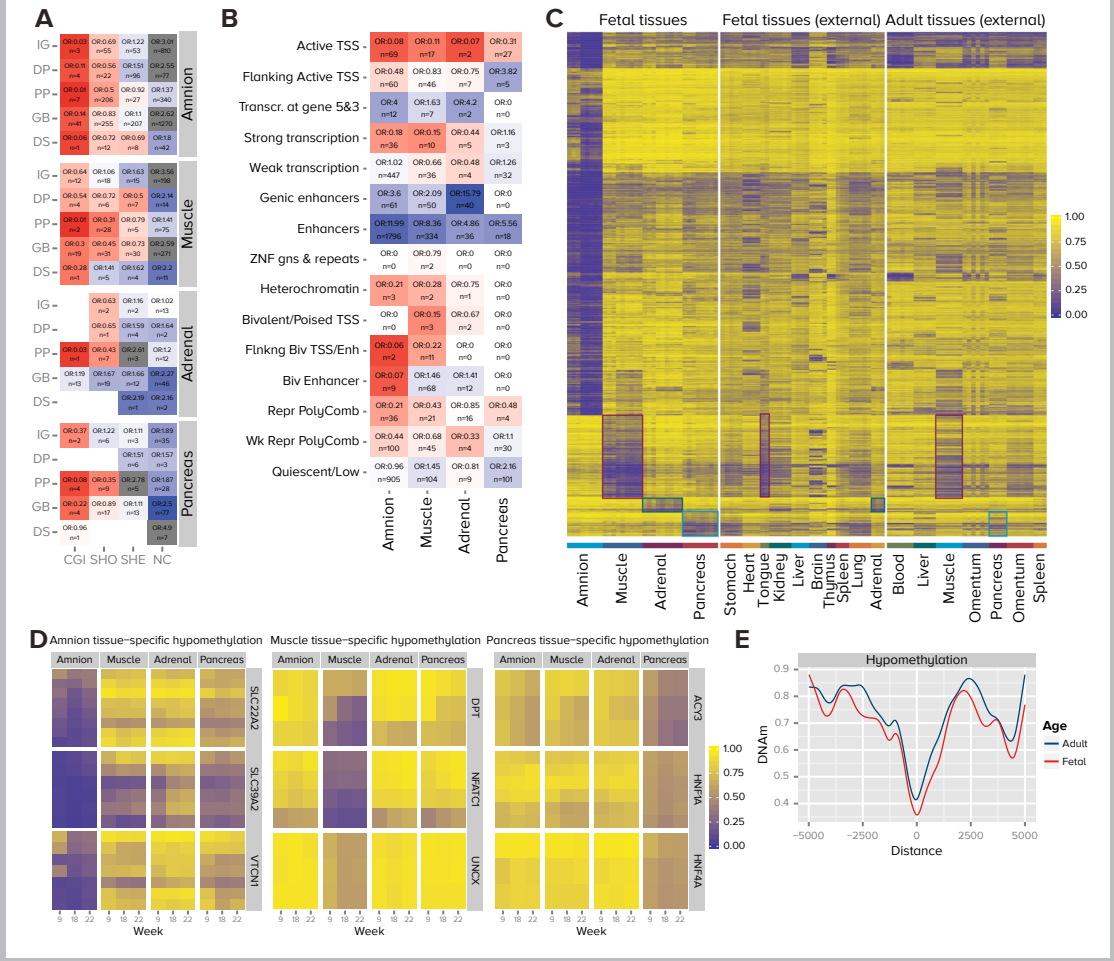
E



[Supplementary Fig. 2]

General description of tissue-specific DNA methylation signatures.

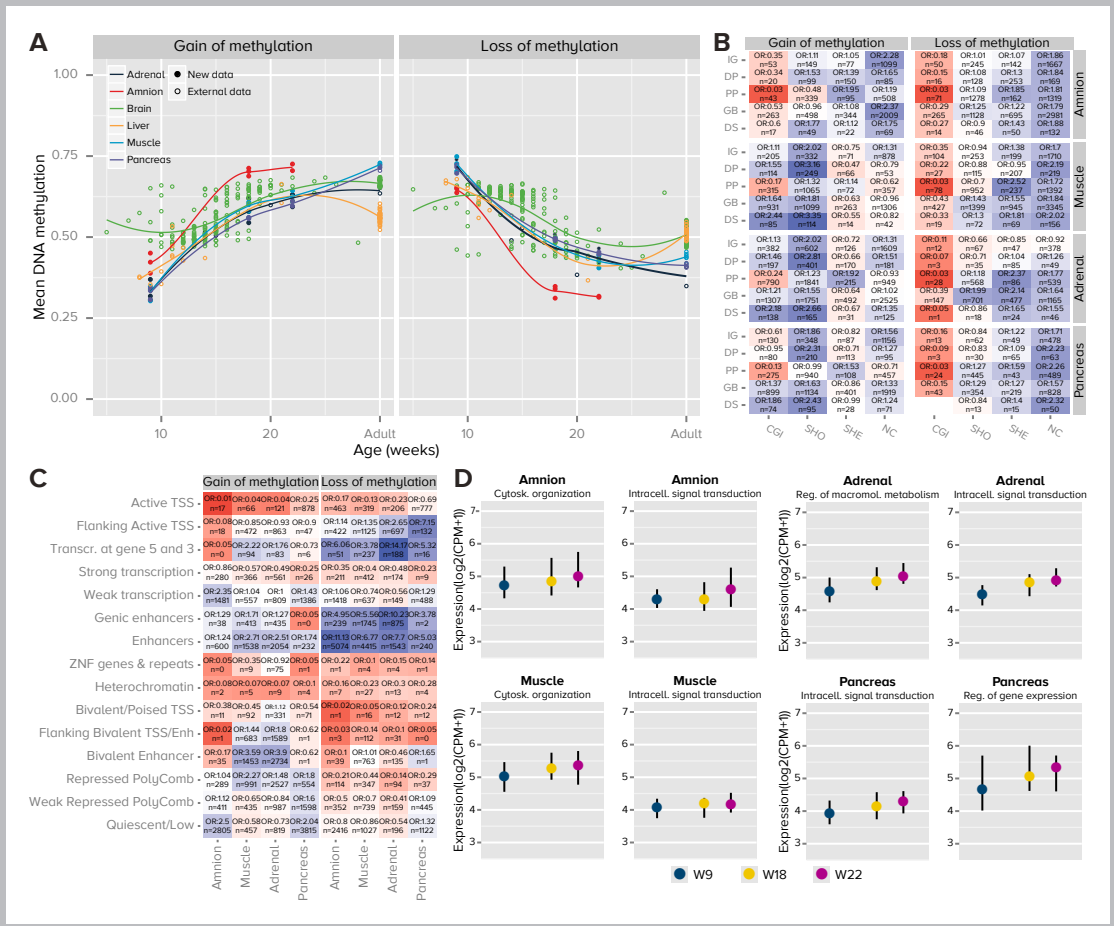
[A] Table of the fetal samples included in this study. **[B]** Multidimensional scaling based on Euclidean distance, from left to right: coloured by tissue, time point and individual. **[C]** Autosomal DNA methylation in three classes (0%–25%; 25%–75%, 75–100%). **[D]** DNA methylation of the X chromosome in female samples. **[E]** Hierarchical clustering of the current data with external fetal data of various tissues [15, 22, 43].


A
IG
DP
PP
GB
DS
Amnion
Muscle
Adrenal
Pancreas
CGI SHO SHE NC
B
Active TSS
Flanking Active TSS
Transcr. at gene 5&3
Strong transcription
Weak transcription
Genic enhancers
Enhancers
ZNF gns & repeats
Heterochromatin
Bivalent/Poised TSS
Flnkng Biv TSS/Enh
Biv Enhancer
Repr PolyComb
Wk Repr PolyComb
Quiescent/Low
Amnion
Muscle
Adrenal
Pancreas
C
Fetal tissues
Fetal tissues (external)
Adult tissues (external)
1.00
0.75
0.50
0.25
0.00
Amnion
Muscle
Adrenal
Pancreas
Stomach
Heart
Tongue
Kidney
Liver
Brain
Thymus
Spleen
Lung
Adrenal
Blood
Liver
Muscle
Omentum
Pancreas
Omentum
Spleen
D
Amnion tissue−specific hypomethylation
Muscle tissue−specific hypomethylation
Pancreas tissue−specific hypomethylation
Amnion
Muscle
Adrenal
Pancreas
SLC22A2
SLC39A2
VTCN1
DPT
NFATC1
UNCX
ACY3
HNF1A
HNF4A
1.00
0.75
0.50
0.25
0.00
9 18 22
Week
E
Hypomethylation
DNAm
0.9
0.8
0.7
0.6
0.5
0.4
−5000
−2500
0
2500
5000
Distance
Age
Adult
Fetal

[Supplementary Fig. 3]

Tissue-specific hypomethylated CpGs and tHRs.

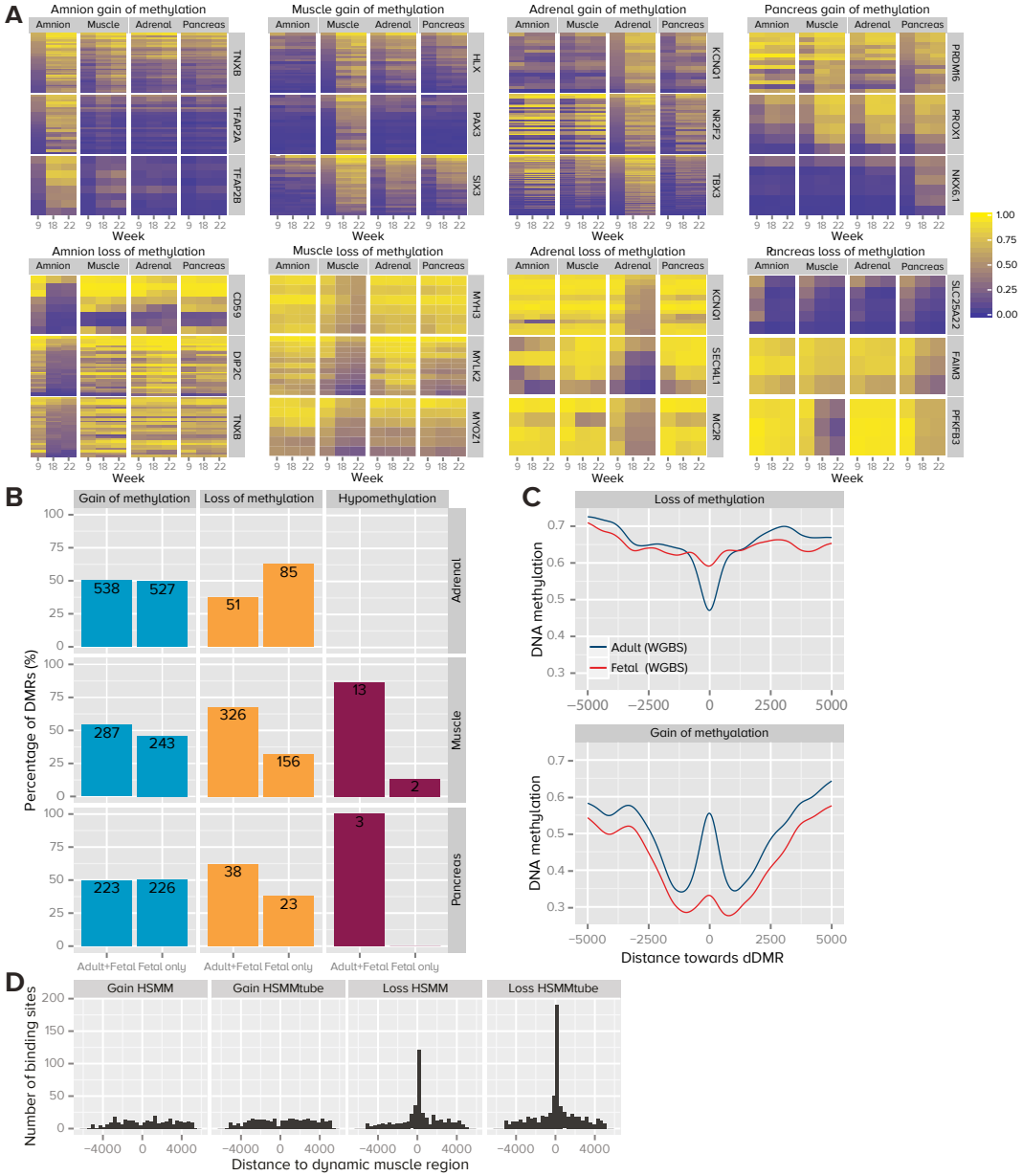
[A] The odds ratios of hypomethylated CpGs per tissue in a combined genic and CGI-centric annotation [Fig. 2B]. **[B]** The odds ratios of hypomethylated CpGs per tissue in the chromatin state segmentations of amnion, fetal muscle, fetal adrenal and adult pancreatic islets [Fig. 2C]. **[C]** Comparison of hypomethylated CpGs per tissue in fetal and adult external data [6, 15]. **[D]** Heatmap representing DNA methylation levels of identified tHRs in amnion, muscle and pancreas in Table 1. **[E]** WGBS DNA methylation profile near hypomethylated regions in muscle of fetal against adult muscle [25].


A
Gain of methylation
Loss of methylation
Mean DNA methylation
Age (weeks)
Adrenal
Amnion
Brain
Liver
Muscle
Pancreas
New data
External data
B
Gain of methylation
Loss of methylation
Amnion
Muscle
Adrenal
Pancreas
IG
DP
PP
GB
DS
CGI
SHO
SHE
NC
C
Gain of methylation
Loss of methylation
Active TSS
Flanking Active TSS
Transcr. at gene 5 and 3
Strong transcription
Weak transcription
Genic enhancers
Enhancers
ZNF genes & repeats
Heterochromatin
Bivalent/Poised TSS
Flanking Bivalent TSS/Enh
Bivalent Enhancer
Repressed PolyComb
Weak Repressed PolyComb
Quiescent/Low
D
Amnion
Cytosk. organization
Amnion
Intracell. signal transduction
Adrenal
Reg. of macromol. metabolism
Adrenal
Intracell. signal transduction
Muscle
Cytosk. organization
Muscle
Intracell. signal transduction
Pancreas
Intracell. signal transduction
Pancreas
Reg. of gene expression
Expression(log2(CPM+1))
W9
W18
W22

[Supplementary Fig. 4]

Dynamic DNA methylation during development.

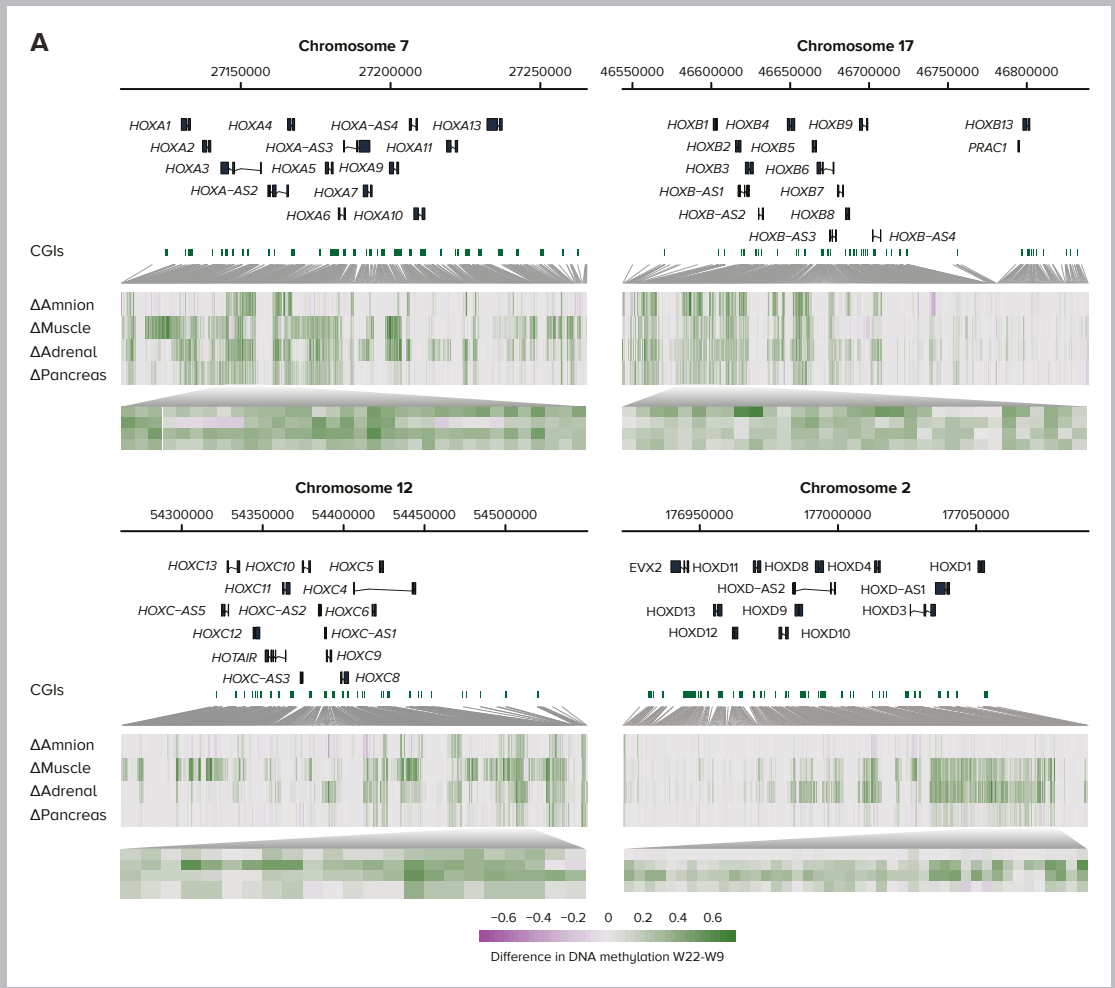
[A] Mean methylation of CpGs with a gain or loss of DNA methylation for fetal tissues and their adult counterpart, including fetal brain and fetal liver [6, 15, 22, 44]. **[B]** The enrichment of dynamically methylated CpGs in a combined genic and CGI-centric annotation [Fig. 3C], significant odds ratios (Chi-squared test $P < 0.05$) are depicted in black. **[C]** The enrichment of dynamically methylated CpGs in the chromatin state segmentations of amnion, fetal muscle, fetal adrenal and adult pancreatic islets [Fig. 3D]. **[D]** Expression profiles of genes near dynamic regions with loss of methylation grouped by the Gene Ontology terms for each of the four tissues from Supplementary Table 3 [29].


A
Amnion gain of methylation
Muscle gain of methylation
Adrenal gain of methylation
Pancreas gain of methylation
Amnion loss of methylation
Muscle loss of methylation
Adrenal loss of methylation
Pancreas loss of methylation
Amnion
Muscle
Adrenal
Pancreas
TNXB
TFAP2A
TFAP2B
HLX
PAX3
SIX3
KCNQ1
NR2F2
TBX3
PRDM16
PROX1
NKX6.1
CD59
DIP2C
TNXB
MYH3
MYLK2
MYOZ1
KCNQ1
SEC14L1
MC2R
SLC25A22
FAIM3
PFKFB3
9 18 22
Week
1.00
0.75
0.50
0.25
0.00
B
Gain of methylation
Loss of methylation
Hypomethylation
Adrenal
Muscle
Pancreas
Percentage of DMRs (%)
538
527
51
85
287
243
326
156
13
2
223
226
38
23
3
Adult+Fetal
Fetal only
C
Loss of methylation
Gain of methylation
DNA methylation
Adult (WGBS)
Fetal (WGBS)
−5000
−2500
0
2500
5000
Distance towards dDMR
D
Gain HSMM
Gain HSMMtube
Loss HSMM
Loss HSMMtube
Number of binding sites
−4000
0
4000
Distance to dynamic muscle region

[Supplementary Fig. 5]

Dynamic CpGs cluster into development-related DMRs.

[A] Average DNA methylation levels of the genes from Table 2. **[B]** Overlap between identified dynamic and hypomethylated regions per tissue and adult tDMRs expressed as percentage overlap. **[C]** WGBS DNA methylation profile near regions with gain and loss of methylation in muscle of fetal against adult muscle [25]. **[D]** Number of *MYOD* binding sites relative to the dynamic regions identified in HSMMs and HSMMtubes. HSMMtube, human skeletal muscle myotube.


A
Chromosome 7
27150000
27200000
27250000
HOXA1
HOXA4
HOXA-AS4
HOXA13
HOXA2
HOXA-AS3
HOXA11
HOXA3
HOXA5
HOXA9
HOXA-AS2
HOXA7
HOXA6
HOXA10
Chromosome 17
46550000
46600000
46650000
46700000
46750000
46800000
HOXB1
HOXB4
HOXB9
HOXB13
HOXB2
HOXB5
PRAC1
HOXB3
HOXB6
HOXB-AS1
HOXB7
HOXB-AS2
HOXB8
HOXB-AS3
HOXB-AS4
CGIs
ΔAmnion
ΔMuscle
ΔAdrenal
ΔPancreas
Chromosome 12
54300000
54350000
54400000
54450000
54500000
HOXC13
HOXC10
HOXC5
HOXC11
HOXC4
HOXC-AS5
HOXC-AS2
HOXC6
HOXC12
HOXC-AS1
HOTAIR
HOXC9
HOXC-AS3
HOXC8
Chromosome 2
176950000
177000000
177050000
EVX2
HOXD11
HOXD8
HOXD4
HOXD1
HOXD-AS2
HOXD-AS1
HOXD13
HOXD9
HOXD3
HOXD12
HOXD10
CGIs
ΔAmnion
ΔMuscle
ΔAdrenal
ΔPancreas
−0.6 −0.4 −0.2 0 0.2 0.4 0.6
Difference in DNA methylation W22-W9

[Supplementary Fig. 6]

Dynamic DNA methylation in the *HOX* clusters.

[A] DNA methylation patterns in the four developmental *HOX* clusters *HOXA*, *HOXB*, *HOXC* and *HOXD*. The bottom heatmap of each cluster zooms in on a smaller genomic region.

CHAPTER 5

DNA METHYLATION AND TRANSCRIPTIONAL TRAJECTORIES DURING HUMAN DEVELOPMENT AND REPROGRAMMING OF ISOGENIC INDUCED PLURIPOTENT STEM CELLS

Matthias S. Roost[1], Roderick C. Slieker[2], Monika Bialecka[1], Liesbeth van Iperen[1], Maria M. Gomes Fernandes[1], Nannan He[1], H. Eka D. Suchiman[2], Karoly Szuhai[3], Françoise Carlotti[4], Eelco J.P. de Koning[4,5], Christine L. Mummery[1], Bastiaan T. Heijmans[2], Susana M. Chuva de Sousa Lopes[1,6]

[1] Department of Anatomy and Embryology,

[2] Molecular Epidemiology Section,

[3] Department of Molecular Cell Biology,

[4] Department of Nephrology, Leiden University Medical Center, Leiden, the Netherlands

[5] Hubrecht Institute for Developmental Biology and Stem Cell Research, University Medical Center, Utrecht, the Netherlands

[6] Department for Reproductive Medicine, Ghent University Hospital, Ghent, Belgium

Submitted 2016

ABSTRACT

Determining cell identity and maturation status of differentiated human pluripotent stem cells (PSCs) requires detailed knowledge of the transcriptional and epigenetic trajectory of organs during development. Here, we generated a transcriptional and DNA methylation atlas covering 21 organs during human fetal development that serves this purpose. Analysis of multiple isogenic organ sets showed that organ-specific DNA methylation patterns during the same period of development were highly dynamic with strong organ-specific increases in both hypomethylated and hypermethylated regions of the genome from week (W)9 to W22 of gestation. We investigated the impact of reprogramming on the established organ-specific DNA methylation by generating human induced pluripotent stem cell (hiPSC) lines from six isogenic organs at W21. The isogenic hiPSCs acquired DNA methylation patterns comparable to existing hPSCs independent of their origin. However, hiPSCs derived from fetal brain retained some brain-specific DNA methylation marks that seemed sufficient to confer higher propensity to differentiate to neural derivatives. This systematic analysis of human fetal organs during development and associated isogenic hiPSC lines provides novel insights in the role of DNA methylation in two opposite processes, lineage commitment and epigenetic reprogramming in humans.

INTRODUCTION

Every organ in the body has a core, organ-specific transcriptional signature that ultimately determines the shape and functionality of each organ and ensures that this remains stable throughout the life of the organism. Whilst much has been published on organ-specific transcriptional and epigenetic landscapes in laboratory animals and stem cell models *in vitro*, equivalent comprehensive data using a large set of human organs from the same individual (isogenic analysis), that circumvents genetic differences confounding the outcome, has not been performed to date [1–12]**.**

We have determined the transcriptional profiles of human fetal organs from the first and second trimester of development and identified a set of core organ-specific genes or «key genes» (also referred to as classifier genes) that were highly expressed in the organ it identifies, often from as early as 9 weeks of gestation (W9) [7]**. In contrast to the organ-specific transcriptional identity, the core organ-specific pattern of DNA hypomethylation, that remains stable throughout adulthood, takes longer to be established. More precisely, between W9 and W22 the general development-related programs gain DNA methylation and are shutdown, whereas organ-specific genetic programs associated with organ functionality lose DNA methylation** [8]**. The DNA methylation pattern observed at W22 in some organs appears at least in part to be maintained during adulthood** [13–16]**, suggesting lineage commitment.**

Setting the correct patterns of DNA methylation is crucial during development, but removing those during the reverse process of reprogramming somatic cells from any human tissue to pluripotency as induced pluripotent stem cells (iPSCs) [17, 18] **is also important. Reprogramming is accompanied by extensive epigenetic remodeling, which results in a**

pluripotent state comparable to that of embryonic stem cells (ESCs) [19, 20]**. However, although the gene expression signatures of iPSCs and ESCs are similar, when large numbers of lines are compared, individual lines are not necessarily identical** [21–24]**. This led to the hypothesis that a residual epigenetic memory may be retained from the tissue of origin. Indeed, it was demonstrated that mouse and human iPSCs harbor some features of the tissue of origin, i.e. histone modifications, DNA methylation, and micro RNAs, which in turn can favor differentiation towards the lineage from which they were derived** [25–33]**. The main contribution to the variation between hiPSCs and hESCs has also been suggested to be the genetic background instead of epigenetic memory** [20, 34, 35]**. However, data to distinguish between these two possibilities is currently lacking. An intriguing difference between mouse and human iPSCs is that the epigenetic memory of mouse iPSCs is lost during continuous passage in culture, whereas human iPSCs appear to have more persistent epigenetic marks** [30, 32]**. Understanding how these factors influence the differentiation capacity of iPSCs would help determine a better framework for the use iPSCs in disease modeling, drug screening and regenerative medicine.**

We present here the most comprehensive analysis of human fetal DNA methylation and corresponding genome-wide transcription data to date: the analysis includes 21 human fetal organs (plus maternal endometrium) from different fetuses (n = 18) at W8–12, W16–18 and W20–22. Furthermore, we used this presently unique material to overcome one of the major challenges to assess epigenetic memory in human iPSCs with different origins by generating lines from six isogenic fetal organs (brain, skin, kidney, muscle, lung and pancreas). For all organs, we used the same primary cell isolation method, culture protocol and reprogramming conditions. We compared the DNA methylation profiles of hiPSCs to their organs of origin and exemplify how very small differences in DNA methylation may result in different propensity of hiPSCs to differentiate.

METHODS

ETHICAL STATEMENT

This study has been approved by the Medical Ethical Committee of the Leiden University Medical Center (P08.087). Informed consent was compliant with the Declaration of Helsinki (World Medical Association).

FETAL TISSUE PROCUREMENT AND PRIMARY CELL CULTURE

Human fetal organs used for DNA extraction were processed as previously described [8]**. For primary cell culture, pieces of human fetal organs were minced with a scalpel (Swann Morton, Sheffield, UK) and each transferred to gelatin-coated wells of 6-well plates with isolation medium (Dulbecco's Modified Eagle Media (DMEM)/F12 supplemented with Glutamax (Gibco, Bleiswijk, the Netherlands), 10mM NEAA, 2mM L-glutamine, 25 U/ml penicillin, 25 mg/ml streptomycin, 50 µg/ml gentamicin, 100mM b-mercaptoethanol (all Invitrogen, Breda, the Netherlands), 25 µg/ml normocin (Invivogen, San Diego, USA), 20% knock-out serum replacement (KOSR; Invitrogen, Breda, the Netherlands)). After two days, the medium was changed and one day later, the cells were washed with PBS (Gibco) and fresh isolation medium was added. After six to seven days, the cells were trypsinized (Gibco) and frozen in freezing medium (80% isolation medium, 10% KOSR, 10% DMSO (Sigma-Aldrich, St. Louis, USA)).**

GENERATION OF HIPSCS

To reprogram the fetal cells, the pRRL.PPT.SF.hOKSMidTomato-preFRT polycistronic lentiviral vector was used as previously published [36]**. Briefly, 2 x 10^4 cells/12-well plate well were plated and transduced the following day with the lentivirus at 1-2 MOI**

in isolation medium supplemented with 4 µg/ml polybrene (Sigma-Aldrich). After 24 hours, the medium was changed and three days later, all the cells were split 1:1 or 1:2 (depending on the fluorescence intensity) into 60 mm dishes coated with mouse embryonic fibroblasts (MEFs; 7.2 x 10^5 MEFs/dish). After culturing the transduced cells in isolation medium for one day, the medium was changed to hiPSC medium (Dulbecco's Modified Eagle Media (DMEM)/F12 supplemented with Glutamax, 10mM NEAA, 25 U/ml penicillin, 25 mg/ml streptomycin, 100µM b-mercaptoethanol, 20% knock-out serum replacement and 10 ng/ml basic FGF (PreproTech Neuilly-Sur-Seine, France)). After manual picking, hiPSC-like colonies were cultured for one to three passages and then either frozen in 90% fetal calf serum (Sigma-Aldrich) and 10% DMSO (Sigma-Aldrich) or further expanded on Matrigel (Corning, Wiesbaden, Germany) in mTESR1 (Stem Cell Technologies, Grenoble, France).

CHARACTERIZATION AND NEURAL INDUCTION OF HIPSCS

Immunocytochemistry was performed following standard procedures. Briefly, cells were fixed in 4% (w/v) paraformaldehyde (PFA, MERCK, Darmstadt, Germany) for 15 minutes at room temperature (RT). Subsequently, the cells were first permeabilized using 0.1% Triton X-100 (Sigma-Aldrich, St. Louis, USA) and then blocked with 1% bovine serum albumin, fraction V (BSA, Sigma-Aldrich) in 0.05% Tween-20 (Promega, Madison, USA) for 1 hour. The following primary antibodies were then applied overnight at 4°C: Goat anti-OCT4 (1:100, Santa Cruz Biotechnologies, Dallas, USA), mouse anti-TRA-1-81 (1:100, Millipore, Bedford, USA), mouse anti-SSEA4 (1:100, Santa Cruz Biotechnologies), rabbit anti-NANOG (1:250, Stemgent, San Diego, USA). The secondary antibodies Alexa Fluor 488 donkey anti-rabbit, Alexa Fluor 488 donkey anti-goat and Alexa Fluor 594 donkey anti-mouse (all 1:500, Life Technologies, Carlsbad, USA) were

added for 1 hour at RT and the nuclei were counterstained using 4',6-diamidino-2-phenylindole (DAPI, Life Technologies). Imaging was performed on an Eclipse Ti imaging system (Nikon) operated by NIS Elements software and compiled in Photoshop CS6 (Adobe).

RNA from the hiPSCs was extracted with the RNeasy Kit (Qiagen, Hilden, Germany) including on-column DNase digestion, followed by cDNA generation with the iScript™ cDNA Synthesis Kit (BioRad, Hercules, USA). Quantitative PCR was carried out on the CFX96TM Realtime system, C1000TM Thermal Cycler (Biorad) using the iQ SYBR Green Supermix (BioRad) and the following program: (1) 3 minutes. 95°C; (2). 40 cycles 15 seconds 95°C; 30 seconds 60°C, 45 seconds 72°C; and (3) 10 seconds 95°C; 5 seconds 65°C, 50 seconds 95°C,). The ΔΔCt method and normalization to *GAPDH* and *ACTB* was used to assess expression levels. The expression levels of hiPSCs were compared to those of the hESC-NKX2.5eGFP/w line, which was used as positive control [37]. The primer sequences are: *ACTB* - Fw CTG GAA CGG TGA AGG TGA CA and Rv AAG GGA CTT CCT GTA ACA ACG CA; *GAPDH* - Fw CTG CAC CAC CAA CTG CTT AG and Rv GTC TTC TGG GTG GCA GTG AT; *POU5F1* endogenous – Fw GAC AGG GGG AGG GGA GGA GCT AGG and Rv CTT CCC TCC AAC CAG TTG CCC CAA AC; *NANOG* - Fw TGC AAG AAC TCT CCA ACA TCC T and Rv ATT GCT ATT CTT CGG CCA GTT; *SOX2* endogenous – Fw GGG AAA TGG GAG GGG TGC AAA AGA GG and Rv TTG CGT GAG TGT GGA TGG GAT TGG TG [17, 38, 39].

The karyotypes of the 12 hiPSC lines were assessed using combined binary ratio labelling (COBRA) as previously described [40].

Neural induction of hiPSCs was performed using the Stemdiff Neural System (Stem cell Technologies, Catalog #05835) according to manufacturer's instructions. Briefly, embryoid bodies (4000 cells/embryoid body) were grown for 4 days and subsequently plated on matrigel-coated dishes.

Cells were analysed by dark-field microscopy and scored for the presence of neural rosettes at day 7 (% EBs composed of at least 50% rosettes, n = 3). After 7 and 12 days of neural differentiation, cells were fixed with 4% PFA (20 minutes, RT) and used for immunofluorescence as above, using as primary antibody rabbit anti-GFAP (1:200, DAKO), rabbit anti-SOX9 (1:200, Chemicon) and mouse anti-TUBB3 (1:200, Biosciences) and secondary antibody Alexa Fluor 488 donkey anti-rabbit (1:500) and Alexa Fluor 594 donkey anti-mouse (1:500).

DNA EXTRACTION AND 450K ARRAY DATA (PRE) PROCESSING

The genomic DNA (gDNA) of the different organs was extracted as previously described [8]**. For the hiPSCs (from passage 10–14), the Wizard Genomic DNA Purification Kit (Promega, Leiden, the Netherlands) was used as previously described** [41]**. The gDNA was quantified using the Qubit dsDNA BR Assay Kit on a Qubit 2.0 Fluorometer (Sigma-Aldrich). An average input of 600 ng was used for bisulfite conversion with the EZ-96 DNA methylation kit (Zymo Research, Orange County, USA). Subsequently, the DNA methylation profiles were determined with the Illumina HumanMethylation450 BeadChip according the manufacturer's protocol.**

The package minfi [42] **was used to import the data in R version 3.2.2. To normalize the data, a custom pipeline has been used as previously described** [8, 43, 44]**. For all the analyses, except for the DNA methylation profiles of isogenic samples, the probes in CG SNPs (with an allele frequency > 5%) as well as CG probes on the sex chromosomes were excluded** [45]**. The DNA methylation data of the fetal organs (n = 105 samples) and the stem cells (n = 12 hiPSC and n = 6 organs) were normalized separately.**

BIOINFORMATIC ANALYSES

The R package ggplot2 2.0.0 was used for plotting [46]**.**

Gene expression data: Gene expression data was downloaded from the Gene Expression Omnibus (GEO) database (GSE66302) [7] **and was normalized with the R package edgeR 3.2.4 as previously described** [7, 47, 48]**. The multi-dimensional scaling plots were also generated with R package edgeR 3.2.4.**

DNA methylation data: A genic and CGI-centric annotation was used for the 450k CpG probes as previously described [8, 16]**. For both, multidimensional scaling and clustering, Euclidian distance was used. For the clustering, average linkage was applied.**

Differential expression and correlation with methylation: The differentially expressed genes between W8–12 and W20–22 (with W16–18 in between), were identified with the R package edgeR 3.2.4 using a FDR < 0.01 [47, 48]**. For the organs that only had one replicate at one or more time points, the mean of all biological coefficients of variation (BCVs) of the organs with at least two replicates at all three time points was used as dispersion value. The CpGs in the nearest proximal promoters and gene bodies of all uniquely differentially expressed genes per organ were selected and the difference in beta value between W8–12 and W20–22 was calculated.**

Hypermethylation and hypomethylation: Relative hypermethylated and hypomethylated CpGs were defined as pairwise difference of > 0.2 or < 0.2 in beta values, respectively, in the sample of interest compared to the other samples.

Differentially methylated regions (DMRs): Organ-specific DMRs were identified as previously described [8, 16]**.**

EXTERNAL DATA

Gene expression and DNA methylation data were downloaded from GEO database (GSE66302 [7]**; GSE56515** [8]**; GSE30654** [5]**; GSE61461** [49]**).**

RESULTS

Organ-specific transcriptional and DNA methylation show different dynamics during human fetal development

We analyzed DNA methylation (n = 105 samples; Illumina 450k array) and, if available, the corresponding genome-wide transcriptional data (n = 111 samples; NGS Deep-SAGE) from 21 different organs from n = 18 human fetuses at three different time points during gestation W8–12, W16–18 and W20–22 [Fig. 1A; Supplementary Table 1].

After quality control [Supplementary Fig. 1], we applied multidimensional scaling (MDS) to the 111 NGS datasets and found that the organs primarily separated on the basis of their germ layer of origin (extraembryonic, ectoderm, mesoderm, endoderm) and that four distinct clusters formed (liver, brain/spinal cord, placenta/chorion/amnion/umbilical cord, muscle/heart/tongue) [Fig. 1B; Supplementary Fig. 2A]. Moreover, similar organs clustered together, forming their own spatial domains and showing limited degree of intermingling with other organs [Supplementary Fig. 2A].

Next, we performed similar MDS analysis and hierarchical clustering based on Euclidian distance using the sample-matched DNA methylation datasets [Fig. 1C and 1D]. The extraembryonic tissues, chorion and placenta, exhibited a fundamentally different DNA methylation pattern from the other (embryonic) organs. They clustered separately [Fig. 1C and 1D] and showed distinctive high levels of intermediate DNA methylation on autosomes [Supplementary Fig. 1D], previously also described in human term placenta [50]. On the dendrogram, some organs clustered in an organ-specific manner, independently of their developmental age (brain/eye, liver, heart ventricle/atrium, gonad, kidney and lung) and this was similar to the organ-specific clustering observed on the basis of the transcriptome;

while other organs including tongue/muscle, stomach/intestine/pancreas, adrenal and skin, clustered at W8–12, separating only at W16–22 in distinct organ-specific clusters. To increase resolution, we then performed a MDS analysis on the DNA methylation dataset excluding the extraembryonic tissues and confirmed the strong dynamics of DNA methylation between W8–22, in contrast to the more stable tissue-specific transcriptome during the same developmental period [Fig. 1E; Supplementary Fig. 2B]**.**

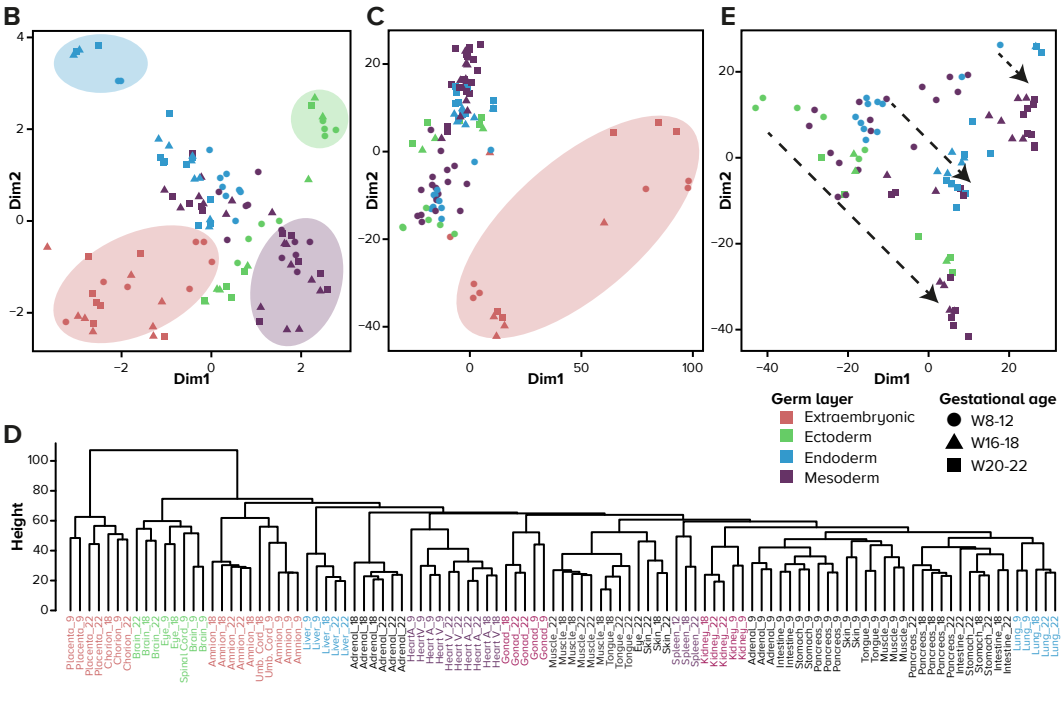
ORGAN-SPECIFIC TRANSCRIPTIONAL CHANGES CORRELATE WITH CHANGES IN DNA METHYLATION DURING HUMAN FETAL DEVELOPMENT

Next, we investigated the presence of differentially expressed genes or DEGs (upregulated and downregulated) between W8 and W22 per organ (FDR < 0.01) [Fig. 2A]. **From the organ-specific DEGs, we identified those that had a log-fold change (logFC) > 1 or logFC < -1** [Fig. 2B and 2C; Supplementary Table 2]**. Interestingly, the number of organ-specific DEGs (excluding X-linked genes to avoid sex bias) was small, with the exception of the upregulated (up-)DEGs in the intestine (n = 203) and maternal endometrium (n = 146). Surprisingly, some organs, like the lung, showed no organ-specific DEGs between W8–22 and many others (liver, pancreas, stomach, kidney and muscle) showed less than 10 organ-specific DEGs, suggesting that either between W8–22 the transcriptional state of the organ-specific progenitors remains similar within each organ or that, alternatively, the bulk sequencing of the organ is masking the cellular maturation of organ-specific progenitors.**

We examined whether the changes in expression of organ-specific DEGs were accompanied by changes in DNA methylation in their respective loci. This was done by calculating the difference in beta values between W8–12 and W20–22 (delta beta) of the CpGs present in the proximal promoters (PP) (-1.5 kb to +0.5 kb) and gene bodies (GB) (+0.5 kb to 3' untranslated

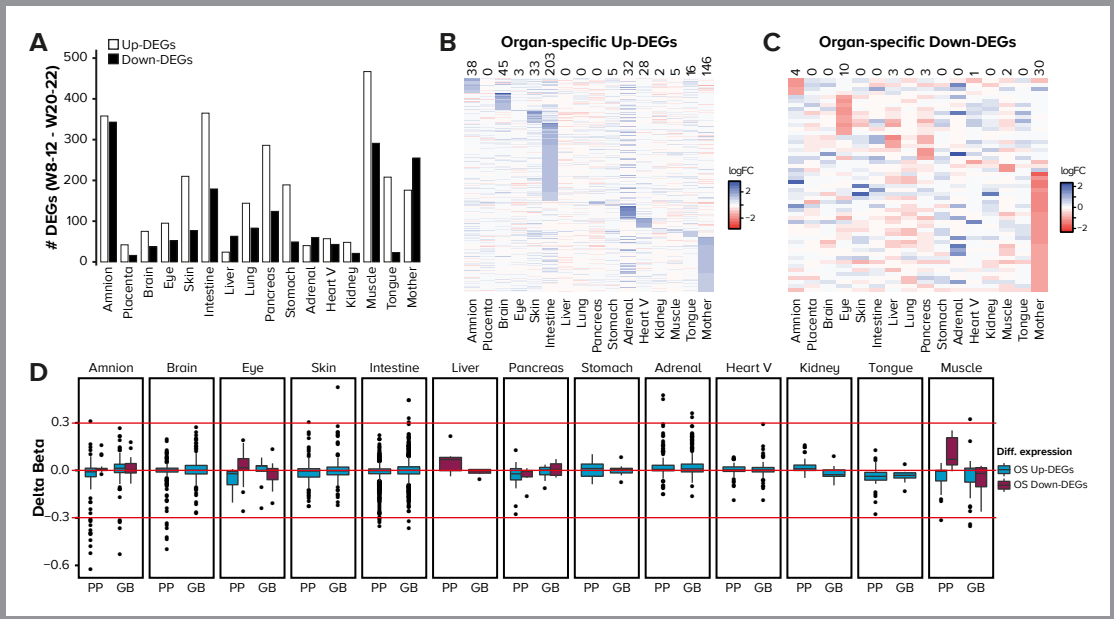
A

	Gestational age (weeks)	Extraembryonic				Ectoderm				Endoderm					Mesoderm								Other	
		Amnion	Chorion	Placenta	Umbilical cord	Brain	Eye	Skin	Spinal cord	Intestine	Liver	Lung	Pancreas	Stomach	Adrenal	Gonad	Heart atrium	Heart ventricle	Kidney	Skeletal muscle	Spleen	Tongue	Mat. endometrium	
RNA sequencing (NGS DeepSAGE)	**W8-12**	2	0	2	2	2	2	2	2	1	2	2	3	2	1	2	0	2	2	2	0	2	1	
	W16-18	3	1	3	0	2	2	3	0	2	2	2	2	2	1	2	2	2	2	2	2	3	1	Roost et al. 2015
	W20-22	2	2	2	0	1	2	2	0	3	2	2	2	2	1	0	2	1	2	2	2	1	1	
DNA methylation (Illumina 450K array)	**W8-12**	3	1	2	1	2	1	2	1	2	2	2	3	2	3	2	2	2	2	3	1	2	0	
	W16-18	3	1	0	1	1	1	1	0	1	1	1	3	1	3	1	1	1	1	3	1	1	0	Slieker et al. 2015
	W20-22	2	1	2	0	2	1	2	0	2	2	2	2	2	3	2	2	2	2	3	1	2	0	



[Fig. 1]

Transcriptional and DNA methylation profiles of a collection of organs of first and second trimester. **[A]** The collection of 111 samples analysed for gene expression (by next generation sequencing (NGS) DeepSAGE) and 105 samples analysed for DNA methylation (by Illumina 450k array). See also Supplementary Table 1. Mat., maternal. **[B]** Multidimensional scaling (Euclidian distance) of the transcriptional profiles of the 111 samples. The colors represent the different germ layers, whereas the shapes indicate the gestational age. The blue circle highlights liver samples, the green circle the spinal cord and brain samples, the red circle most extraembryonic samples and the purple circle the muscular cluster including the heart, muscle and tongue samples. See also Supplementary Fig. 2A. **[C]** Multidimensional scaling (Euclidian distance) of the DNA methylation profiles of the 105 samples. The colors represent the different germ layers, whereas the shapes indicate the gestational age. The red circle highlights the chorion and placenta samples. See also Supplementary Fig. 2B. **[D]** Hierarchical clustering (Euclidian distance) of the 105 DNA methylation samples. **[E]** Multidimensional scaling (Euclidian distance) of the DNA methylation profiles excluding the extraembryonic samples. The colors represent the different germ layers, whereas the shapes indicate the gestational age. Dashed arrows represent the trend of DNA methylation dynamics during development in similar organs.


A
Up-DEGs
Down-DEGs
DEGs (W8-12 - W20-22)
0
100
200
300
400
500
Amnion
Placenta
Brain
Eye
Skin
Intestine
Liver
Lung
Pancreas
Stomach
Adrenal
Heart V
Kidney
Muscle
Tongue
Mother
B
Organ-specific Up-DEGs
38 0 45 3 33 203 0 0 0 5 32 28 2 5 16 146
logFC
2
−2
C
Organ-specific Down-DEGs
4 0 0 10 0 0 3 0 3 0 0 1 0 2 0 30
logFC
2
0
−2
D
Amnion
Brain
Eye
Skin
Intestine
Liver
Pancreas
Stomach
Adrenal
Heart V
Kidney
Tongue
Muscle
Delta Beta
0.3
0.0
−0.3
−0.6
PP
GB
Diff. expression
OS Up-DEGs
OS Down-DEGs

[Fig. 2]

Organ-specific upregulated and downregulated genes and associated changes in DNA methylation.

[A] Numbers of differentially expressed genes (FDR < 0.01) between W8-12 and W20-22 in different organs. Heart V, heart ventricle; Mother, maternal endometrium. **[B]** Heatmap of the upregulated genes that were uniquely assigned to one organ. See also Supplementary Table 2. **[C]** Heatmap of the downregulated genes that were uniquely assigned to one organ. See also Supplementary Table 2. **[D]** Boxplots illustrating the organ-specific methylation changes (delta beta) of the nearest proximal promoter (PP) and the gene body (GB) of the loci identified in [B] and [C], excluding the maternal endometrium. The red line indicates a delta beta of -0.3, 0 and 0.3. See also Supplementary Table 3.

region (UTR)) of the organ-specific up-DEGs and down-DEGs [Fig. 2D; Supplementary Table 3]**. The reduced number of organ-specific down-DEGs was still sufficient to correlate with an increase in DNA methylation (positive delta beta) in their PP and a decrease (negative delta beta) in their GB in eye, liver and muscle** [Fig. 2D]**. By contrast, organ-specific up-DEGs were accompanied by a modest decrease of DNA methylation (negative delta beta) in their PP in amnion, eye, pancreas, tongue and muscle; but an increase (positive delta beta) in their GB only in amnion, eye and pancreas** [Fig. 2D]**.**

Exemplifying the correlation between gene expression and DNA methylation, two CpGs in the PP of the muscle up-DEG *C8orf2*, also shown by others to be overexpressed in skeletal muscle [51]**, showed reduced levels of DNA demethylation between W8–12 and W20–22 (-0.31 and -0.19 of delta beta). In the brain, up-DEG *GFAP*, important in the development of the central nervous system** [52, 53] **and up-DEG *CHRM1*, important in schizophrenia** [54]**, showed pronounced demethylation in their PP between W8–12 and W20–22. The PP of *GFAP* contained five CpGs that underwent demethylation (-0.50, -0.43, -0.42, -0.36, -0.35 of delta beta) and the PP of *CHRM1* contained four CpGs that were demethylated (-0.34, -0.26, -0.24, -0.16 of delta beta). In the eye, we observed increased methylation in five CpGs in the PP (+0.19, +0.17, +0.17, +0.13 and +0.11 of delta beta) and increased demethylation in two CpGs in the GB (-0.2 and -0.13 of delta beta) of eye down-DEGs *CRYBB3* and *CRYBA1*, which are structural components of crystalline, between W8–12 and W20–22.**

ISOGENIC SETS OF HUMAN ORGANS REVEAL ORGAN-SPECIFIC DNA METHYLATION PATTERNS

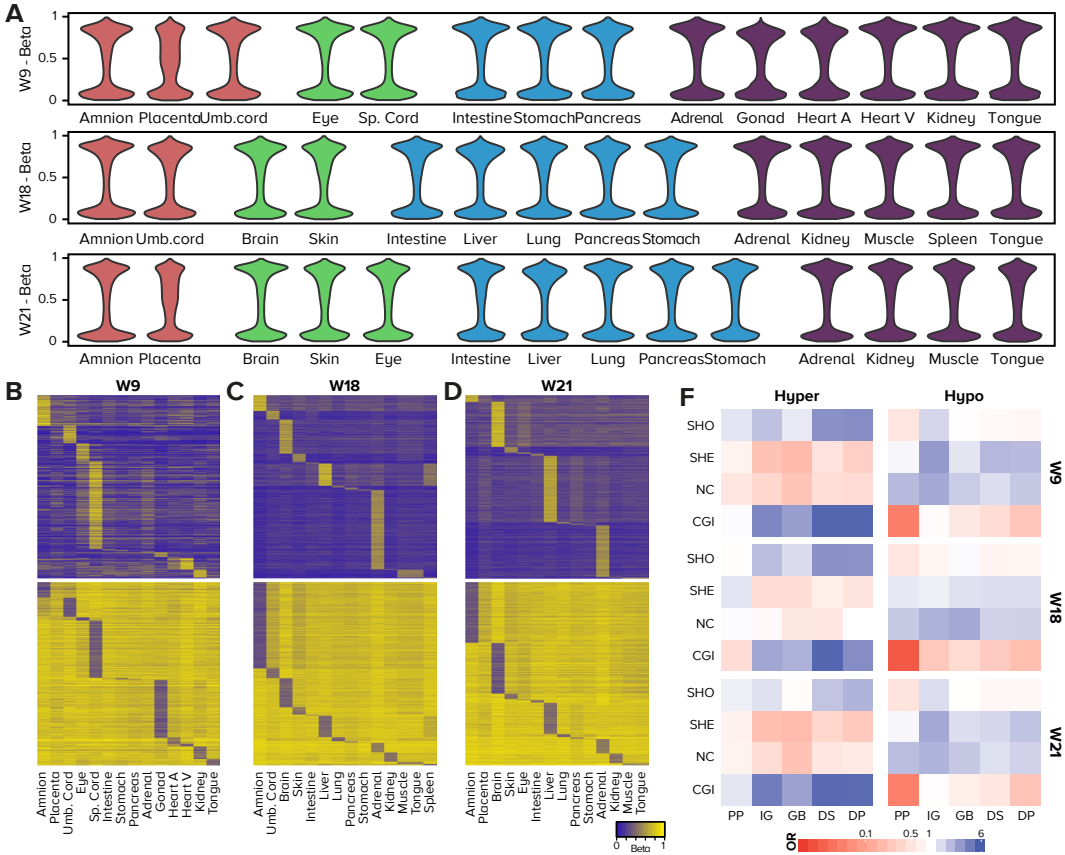
One of the unique features of our datasets is the isogenic nature of many sets of organs [Supplementary Table 1]**, and in particular three sets of DNA methylation of 14 organs from a W9 male, W18 female and W21 male** [Fig. 3A]**. This enabled us to com-**

pare the methylation status of individual CpG between isogenic organs with the assurance that any difference in DNA methylation would be entirely attributable to differences between the organs [Fig. 3A; Supplementary Fig. 3A]**.**

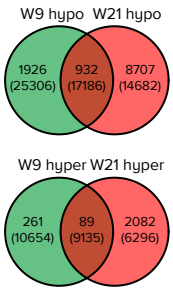
We identified hypermethylated and hypomethylated CpGs, defined as a pairwise difference in beta values of > 0.2 or < 0.2, respectively [Fig. 3B–3D; Supplementary Fig. 3B and 3C]**. The placenta was part of the isogenic W9 and W21 sets and showed high numbers of hypermethylated and hypomethylated CpGs (W9: 19,439 and 39,634 CpGs, W21: 13,260 and 22,311 CpGs, respectively) compared to the other organs** [Supplementary Fig. 3B and 3C]**, most likely reflecting its extraembryonic origin. Therefore, we generated separate plots excluding the hypermethylated and hypomethylated CpGs of the placenta** [Fig. 3B and 3D]**.**

The number of organ-specific hypomethylated CpGs was consistently higher than the number of hypermethylated CpGs [Supplementary Fig. 3D]**, highlighting the importance of hypomethylation as a distinguishing feature between organs [5, 8, 55]. In most individual organs (excluding the placenta), the number of organ-specific hypermethylated and hypomethylated CpGs increased between W9 and W21** [Fig. 3E; Supplementary Fig. 3D]**, suggesting progression in organogenesis and lineage commitment.**

We investigated the genomic location of the organ-specific hypermethylated and hypomethylated CpGs identified (including the placenta) per developmental stage. The organ-specific hypermethylated CpGs were mainly associated with CpG islands (CGIs) and their shores, but only downstream regions ($P < 0.5 \times 10^{-10}$) showed significant enrichment at all three time points. By contrast, the organ-specific hypomethylated CpGs were enriched in non-CGI regions and CGI-shelves consistently at all three time points, particularly when they are associated with distal promoters ($P < 0.5 \times 10^{-10}$) and downstream regions ($P < 0.005$) [Fig. 3F]**.**



E



G

W9	Hyper DMRs	Hypo DMRs
Amnion	4 (72)	5 (457)
Placenta	2371 (19439)	2619 (39634)
Umb. Cord	4 (41)	4 (577)
Eye	3 (43)	1 (98)
Sp. Cord	3 (206)	61 (1687)
Intestine	0 (3)	0 (6)
Stomach	0 (2)	0 (39)
Pancreas	0 (0)	0 (9)
Adrenal	0 (3)	0 (3)
Gonad	0 (11)	47 (1707)
Heart A	0 (2)	0 (182)
Heart V	0 (27)	1 (76)
Kidney	1 (19)	11 (371)
Tongue	0 (2)	5 (188)

W18	Hyper DMRs	Hypo DMRs
Amnion	24 (395)	317 (7908)
Umb. Cord	28 (209)	8 (888)
Brain	54 (852)	107 (2666)
Skin	22 (216)	14 (685)
Intestine	3 (29)	3 (79)
Liver	21 (552)	78 (1975)
Lung	3 (39)	7 (383)
Pancreas	7 (54)	1 (128)
Stomach	1 (19)	0 (10)
Adrenal	173 (1962)	19 (871)
Kidney	0 (11)	42 (864)
Muscle, Tongue	22 (192)	6 (201)
Spleen	1 (22)	4 (129)

W21	Hyper DMRs	Hypo DMRs
Amnion	6 (133)	100 (4087)
Placenta	1692 (13260)	1346 (22311)
Brain	32 (889)	211 (3424)
Skin	8 (112)	6 (367)
Eye	2 (22)	1 (49)
Intestine	5 (42)	4 (186)
Liver	85 (1307)	93 (2115)
Lung	1 (31)	4 (209)
Pancreas	3 (22)	2 (106)
Stomach	1 (9)	0 (22)
Adrenal	91 (1021)	19 (953)
Kidney	0 (11)	28 (667)
Muscle, Tongue	3 (22)	3 (145)

[Fig. 3]

DNA methylation signatures of isogenic sets of organs.

[A] Violin plots showing the distribution of the DNA methylation (beta values) in each of the 14 isogenic organs at W9, W18 and W21. Heart A, heart atrium; Heart V, heart ventricle; Sp. cord, spinal cord. **[B-D]** Heatmaps illustrating hyper- (top) and hypomethylated (bottom) CpGs per organ with a beta value difference of > 0.2 or < 0.2, respectively, at W9 **[B]**, W18 **[C]** and W21 **[D]**. Due to the quantity of hypermethylated and hypomethylated CpGs in the placenta at W9 and W18, those were removed but are given in Supplementary Fig. 3B and 3C. **[E]** Venn diagram illustrating the overlap of the hypermethylated (bottom) and hypomethylated (top) CpGs of the organs at W9 and W21. The numbers in brackets represent the CpGs including the placenta. The overlapping CpGs per organ are given in Supplementary Fig. 3D. **[F]** Genic and CGI-centric annotation of the identified hyper- (left) and hypomethylated (right) CpGs given as odds ratio (OR). CGI, CpG island; DP, distal promoter; GB, gene body; IG, intergenic; NC, non-CGI; PP, proximal promoter; SHE, shelf; SHO, shore. **[G]** Number of organ-specific hypermethylated and hypomethylated differentially methylated regions (DMRs) per isogenic organ. Tongue and muscle were pooled. For the nearest associated loci see Supplementary Table 4. The number of organ-specific hypermethylated and hypomethylated individual CpGs are given in brackets.

In addition to DNA methylation changes in single CpGs in the isogenic datasets, we also inferred the organ-specific differentially methylated regions (DMRs), defined as three consecutive hypermethylated or hypomethylated CpGs within 1 kb of each other [Fig. 3G]**. Considering the limited number of organ-specific hypermethylated and hypomethylated CpGs at W9, many W9 organs contained no organ-specific DMRs** [Fig. 3G]**. However, the number of organ-specific DMRs increased between W9 and W21, and in the second trimester each organ showed at least one organ-specific DMR** [Fig. 3G]**. The skeletal muscle and tongue, both muscular organs, showed similar DNA methylation patterns and were therefore pooled. The organ-specific DMRs were mapped to their nearest loci** [Supplementary Table 4] **and, interestingly, many identified loci showed biological relevance for the corresponding organ. At W21, one of the intestine hypomethylated DMRs was mapped to *TFF1*, a gene that has been reported to be downregulated via DNA methylation in gastric cancer** [56]**. The stomach at W21 on the other hand harbored a hypermethylated DMR associated with *NKX3-2*, which is thought to be involved in stomach development** [57]**.**

GENERATION OF AN ISOGENIC SET OF HUMAN IPSCS FROM SIX DIFFERENT W21 ORGANS

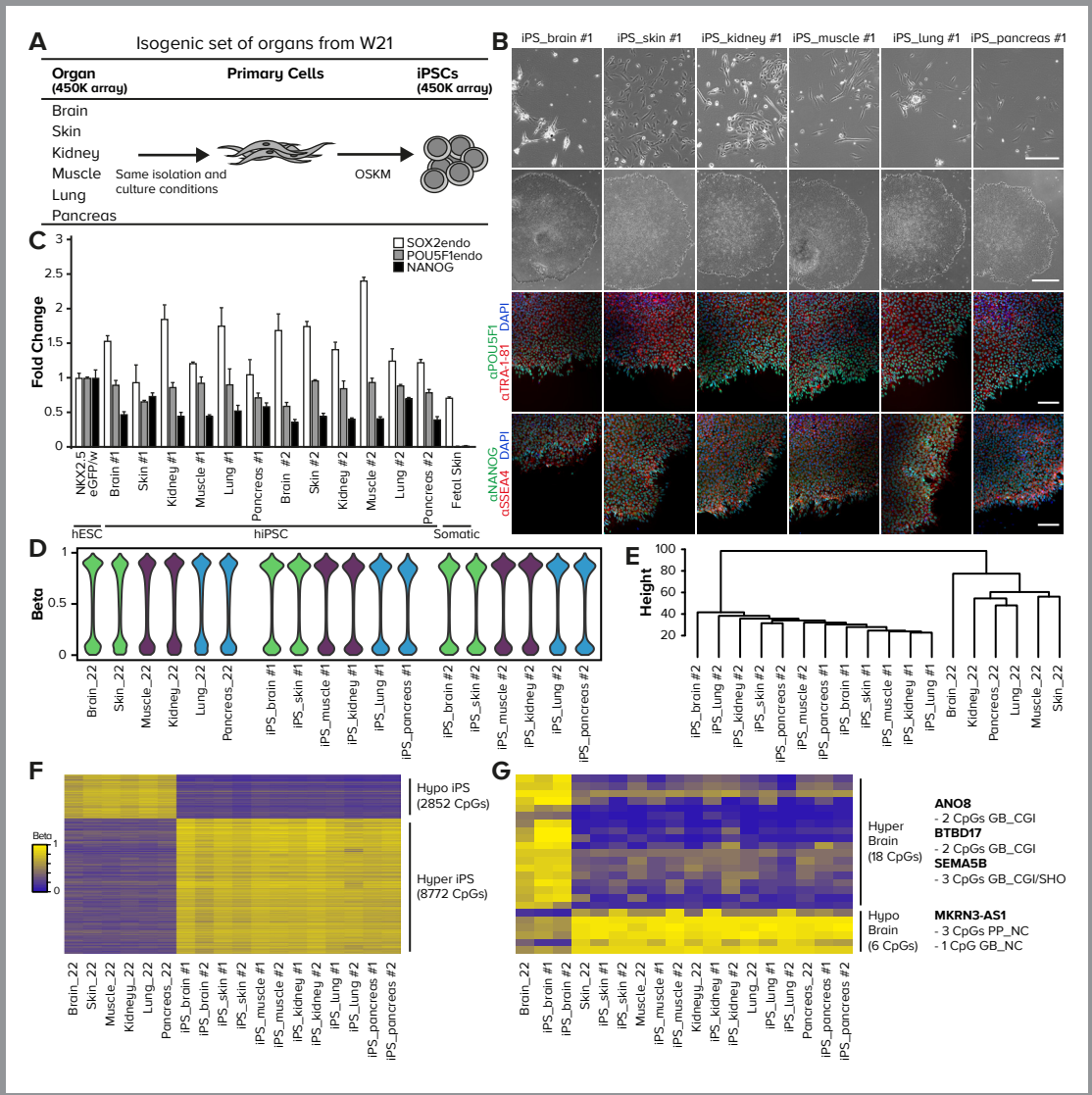
After determining the trajectory of DNA methylation during (isogenic) lineage commitment in development, we investigated the direct effects of reprogramming cells from different isogenic organs to iPSCs with respect to DNA methylation. We developed a protocol to isolate and culture primary cells from different human fetal organs at W21 and reprogrammed primary cells from six different organs representing the three primary embryonic germ layers: brain and skin (ectoderm), kidney and muscle (mesoderm), and lung and pancreas (endoderm) [Fig. 4A and 4B]**. After lentiviral reprogramming using *POU5F1* (or *OCT4*), *KLF4*, *SOX2* and c-*MYC*** [36]**, two clones**

per iPSC line were expanded and characterized [Fig. 4B, top-middle row]**. Immunohistochemical analysis showed that all clones expressed the pluripotency-associated markers POU5F1, TRA-1-81, NANOG and SSEA4** [Fig. 4B, bottom rows; Supplementary Fig. 4A]**. Furthermore, (endogenous) *POU5F1*, *NANOG* and *SOX2* were also detected at the transcriptional level using real-time quantitative PCR** [Fig. 4C]**. From the 12 hiPSC clones generated, 11 clones showed normal karyotype (46,XY) and no major genomic aberrations** [Supplementary Fig. 4B]**. Clone #2 derived from the lung acquired an unbalanced translocation involving chromosome X and 16, der(16)t(X;16), in all cells examined (25 metaphases)** [Supplementary Fig. 4B]**.**

HUMAN IPSCS GENERATED FROM BRAIN TISSUE HARBOR BRAIN-LIKE EPIGENETIC MEMORY

To determine whether differences in residual DNA methylation were evident between each isogenic organ and its derivative hiPSC clones (epigenetic memory), the DNA methylation profiles obtained for each of the hiPSC lines generated (two different clones per line) were compared with the organ of origin. The total DNA methylation distribution showed that the hiPSCs were more hypermethylated than their organ of origin [Fig. 4D]**. Hierarchical clustering using the DNA methylation datasets revealed a separation between the 12 hiPSC clones and their isogenic tissue counterparts** [Fig. 4E]**. We included two previously published 450k datasets of human pluripotent stem cells (in total n = 39 hiPSCs and n = 32 hESCs)** [5, 49] **in the comparison and observed that the different hESCs and hiPSCs lines clustered together** [Supplementary Fig. 4D] **indicating that general patterns of DNA methylation of all lines studied might be comparable.**

To explore this further, we looked for a hiPSC-specific DNA methylation barcode by comparing the CpG methylation levels of our hiPSC clones to that of the six tissues of origin [Fig. 4F]**. We found 8,772 hypermethylated (> 0.2 delta beta)**


A
Isogenic set of organs from W21
Organ (450K array)
Primary Cells
iPSCs (450K array)
Brain
Skin
Kidney
Muscle
Lung
Pancreas
Same isolation and culture conditions
OSKM
B
iPS_brain #1
iPS_skin #1
iPS_kidney #1
iPS_muscle #1
iPS_lung #1
iPS_pancreas #1
αPOU5F1
αTRA-1-81
DAPI
αNANOG
αSSEA4
DAPI
C
SOX2endo
POU5F1endo
NANOG
Fold Change
NKX2.5 eGFP/w
Brain #1
Skin #1
Kidney #1
Muscle #1
Lung #1
Pancreas #1
Brain #2
Skin #2
Kidney #2
Muscle #2
Lung #2
Pancreas #2
Fetal Skin
hESC
hiPSC
Somatic
D
Beta
Brain_22
Skin_22
Muscle_22
Kidney_22
Lung_22
Pancreas_22
iPS_brain #1
iPS_skin #1
iPS_muscle #1
iPS_kidney #1
iPS_lung #1
iPS_pancreas #1
iPS_brain #2
iPS_skin #2
iPS_muscle #2
iPS_kidney #2
iPS_lung #2
iPS_pancreas #2
E
Height
F
Beta
Hypo iPS (2852 CpGs)
Hyper iPS (8772 CpGs)
G
Hyper Brain (18 CpGs)
Hypo Brain (6 CpGs)
ANO8
- 2 CpGs GB_CGI
BTBD17
- 2 CpGs GB_CGI
SEMA5B
- 3 CpGs GB_CGI/SHO
MKRN3-AS1
- 3 CpGs PP_NC
- 1 CpG GB_NC

[Fig. 4]

Epigenetic memory of isogenic human iPSCs from six organs.

[A] Illustration describing the experimental setup for reprogramming. OSKM, reprogramming factors *POU5F1, SOX2, KLF4, MYC*. **[B]** Bright-field pictures of the primary cells (first row, scale bar 50 µm) and the iPSC colonies of clone #1 (second row, scale bar 1 mm) from the six organs. The iPSCs (clones #1) were immunostained for POU5F1 and TRA-1-81 (third row, scale bar 100 µm) and NANOG and SSEA4 (fourth row, scale bar 100 µm). For the single channels and the immunostaining of clones #2 see Supplementary Fig. 4A. **[C]** Real-time quantitative PCR analysis of the 12 iPSCs for endogenous *SOX2*, endogenous *OCT4* and *NANOG*. NKX2.5eGFP/w hESCs were used as positive and fetal skin as negative control. Each bar represents the mean of three technical replicates ± standard deviation. **[D]** Violin plots depicting the distribution of the DNA methylation (beta values) of the 12 iPSCs and their isogenic organ of origin. **[E]** Hierarchical clustering of the 12 iPSCs and their isogenic organ of origin based on the DNA methylation profiles. **[F]** Heatmap of uniquely hypermethylated (top) and hypomethylated (bottom) CpGs in the 12 iPSCs compared to the six organs of origin with a beta value difference of > 0.2 or < 0.2, respectively. See also Supplementary Table 5. **[G]** Heatmap depicting uniquely hypermethylated (top) and hypomethylated (bottom) CpGs in the brain triplet (two iPSC clones and brain) compared to the five triplets with a beta value difference of > 0.2 or < 0.2, respectively. The genic location and gene identity are provided.

and 2,952 hypomethylated (< 0.2 delta beta) CpGs common to all 12 iPSC clones [Supplementary Table 5]**, underlining the role of DNA methylation during reprogramming towards pluripotency** [58]**. The hypermethylated and hypomethylated CpGs identified in our hiPSC clones showed comparable methylation levels in the large majority of the samples from two external datasets analyzed** [5, 49] [Supplementary Fig. 4D]**, suggesting that the DNA methylation pattern of hiPSC-specific CpGs could barcode other human PSCs. Gene ontology analysis revealed that hypermethylated CpGs were enriched in genes regulating cell adhesion, whereas hypomethylated CpGs were enriched in genes regulating embryonic morphogenesis and development.**

Next, we focused on the CpGs that were hypermethylated (> 0.2 delta beta) or hypomethylated (< 0.2 delta beta) in each triplet (organ of origin and the two associated hiPSC clones) compared to the other five triplets, which would reflect organ-specific epigenetic memory. Surprisingly, only the brain-triplet contained distinctive methylated CpGs: 18 hypermethylated and 6 hypomethylated CpGs [Fig. 4G]**. However, 7 of the 18 brain-triplet hypermethylated CpGs were mapped to the gene bodies of *BTBD17*, a protein-coding gene overexpressed in fetal and adult brain [**51, 59]**; *ANO8*, a Ca^{2+}-activated chloride channel expressed in the human brain** [51, 60]**; and *SEMA5B*, a semaphorin associated with axon development** [61]**; and 3 of the 6 brain-triplet hypomethylated CpGs mapped to the proximal promoter of *MKRN3-AS1*, an antisense RNA that is overexpressed in brain and spinal cord [51]. Those genes as well as the genic location of the CpGs, i.e. gene bodies for hypermethylation and proximal promoters for hypomethylation, suggested that there is a small degree of brain-specific residual epigenetic memory in the brain-hiPSCs.**

To test the functionality of the brain-specific residual epigenetic memory in the brain-hiPSCs, we differentiated both clones of the brain-hiPSCs and skin-hiPSCs to the neural

lineage [Supplementary Fig. 4E]**. After 7 days of differentiation, the brain- and skin-hiPSCs generated similar percentages of embryoid bodies composed of at least 50% of neural rosettes. However, the brain-hiPSC culture contained more TUBB3-positive neural derivatives and less SOX9-positive neural progenitor cells than the skin-hiPSC culture, suggesting higher neural differentiation propensity** [Supplementary Fig. 4F and 4G]**. Indeed, after 12 days of differentiation, the brain-hiPSC culture contained more GFAP-positive neural derivatives than the skin-hiPSC culture** [Supplementary Fig. 4G]**. In summary, both the brain-hiPSCs and skin-hiPSCs differentiate to neural derivatives, but the brain-hiPSCs did so with slightly higher propensity.**

DISCUSSION

In this study, we generated a human atlas consisting of large number of DNA methylation (n = 105 samples) and transcriptome (n = 111 samples) profiles of human fetal organs from 18 individuals at matched gestational time points (W8–12, W16–18, W20–22).

The analyses of the transcriptional datasets allowed us to identify DEGs that were upregulated or downregulated in an organ- and developmental stage-specific manner. We further compared these differences in gene expression with changes in DNA methylation per organ analysed. Organ-specific down-DEGs, in particular in the eye, muscle and liver, associated with increased DNA methylation in their proximal promoters and decreased DNA methylation in their gene bodies, in line with previous findings using hESCs, fibroblasts and monocytes [62]**. Organ-specific up-DEGs were accompanied by decreased DNA methylation in their proximal promoters and increased DNA methylation in their gene bodies in amnion, eye and pancreas. This modest correlation between up-DEGs and DNA methylation changes may be partly explained by the fact that many CpGs in proximal promoters are unmethylated, independent of transcriptional state of each loci.**

Our isogenic sets of 14 organs from W9, W18 and W21 provided us with the organ-specific DNA methylation pattern and revealed its dynamics between W9 and W21. Future studies using further improvements in technology to investigate genome-wide DNA methylation and transcription simultaneously at the single-cell level will increase the resolution from organ to specific progenitor cell types of interest [63, 64]**.**

We succeeded in deriving hiPSCs from six isogenic organs (brain, skin, kidney, muscle, lung and pancreas) from W21, a gestational time point with pronounced organ-specific differences in both hypermethylation and hypomethylation of DMRs.

Our results suggest that the 12 hiPSC clones derived were comparable with respect to global DNA methylation to other human PSCs and different from the organs of origin.

Finally, we investigated whether DNA methylation would relate isogenic derivative hiPSC clones to their organs of origin (epigenetic memory). We observed that the brain-hiPSC clones retained 18 hypermethylated and 6 hypomethylated CpGs that were brain-specific. This could be an underestimate due to the nature of the platform used. Notably, the nearest loci of those CpGs did not appear to be random and 11 of them were associated with four genes (*ANO8, BTBD17, SEMA5B, MKRN3-AS1*) having a neural context. In addition, the fact that the other organ-derived hiPSCs did not show residual DNA marks relating them to their organ of origin, but rather resembled skin-hiPSCs may suggest that the reprogrammed cells may have been fibroblasts or stromal cells independent of the organ of origin. In future studies, we would consider reprogramming either more homogeneous somatic cells or defined cell types. Nevertheless, when subjected to the same directed differentiation protocol, the brain-hiPSCs and skin-hiPSCs showed different propensity to form neural derivatives. This difference in differentiation propensity, perhaps caused by the limited epigenetic memory of the brain-hiPSCs, suggested that this phenomenon should be more extensively studied as it may have implications in the use of human iPSCs in biomedical applications [65, 66]**: the tissue of choice from which to derive patient-specific hiPSCs and the retention of specific DNA methylation marks may determine the success of each specific application (type of drug testing, type of disease modeling or tissue repair).**

SUPPLEMENTARY INFORMATION

The Supplementary Tables are available in the online version of the paper.

ACCESSION NUMBERS

Data has been deposited in the NCBI's Gene Expression Omnibus under the accession number GSE76641.

ACKNOWLEDGEMENTS

We would like to acknowledge H. Locher for the help with tissue dissection, the Centre for Contraception, Abortion and Sexuality (CASA) in Leiden and The Hague for the collection of the human fetal material, C. Freund for advice and support regarding reprogramming, V. Schwach for providing RNA from hESC-NKX2.5eGFP/w, R. Ramakrishnan for testing primers and W. Arindrarto for technical support on the bioinformatics analyses. S.C.d.S.L. is supported by the Netherlands Organisation for Scientific Research (NWO, ASPASIA 015.007.037) and the Interuniversity Attraction Poles (IAP, P7/07); M.S.R. and C.L.M. are supported by the Bontius Stichting; F.C. and E.J.P.d.K. by the Bontius Stichting, Stichting DON, the Dutch Diabetes Research Foundation and JDRF; B.T.H. and R.C.S. by the European Union's Seventh Framework Program IDEAL (FP7/2007–2011) under grant agreement No. 259679.

AUTHOR CONTRIBUTIONS

M.S.R., S.M.C.d.S.L., C.L.M., F.C. and E.J.P.d.K. conceived and designed the experiments. M.S.R., R.C.S., L.v.I., M.M.G.F., N.H., M.B., H.E.D.S. and K.S. performed the experiments. M.S.R., R.C.S., M.B., C.L.M., K.S., B.T.H. and S.M.C.d.S.L. analysed the data. M.S.R., C.L.M. and S.M.C.d.S.L. wrote the manuscript. All authors read and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

[1] Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR *et al*: **The NIH Roadmap Epigenomics Mapping Consortium**. *Nat Biotechnol* 2010, **28**(10):1045–1048.

[2] Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E *et al*: **Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver**. *Genome Res* 2009, **19**(6):1044–1056.

[3] The ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**(7414):57–74.

[4] FANTOM Consortium and the RIKEN PMI and CLST (DGT): **A promoter-level mammalian expression atlas**. *Nature* 2014, **507**(7493):462–470.

[5] Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Muller FJ, Wang YC, Boscolo FS *et al*: **Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives**. *Cell Stem Cell* 2012, **10**(5):620–634.

[6] Roadmap Epigenomics Consortium: **Integrative analysis of 111 reference human epigenomes**. *Nature* 2015, **518**(7539):317–330.

[7] Roost MS, van Iperen L, Ariyurek Y, Buermans HP, Arindrarto W, Devalla HD, Passier R, Mummery CL, Carlotti F, de Koning EJ *et al*: **KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas**. *Stem Cell Reports* 2015, **4**(6):1112–1124.

[8] Slieker RC, Roost MS, van Iperen L, Suchiman HE, Tobi EW, Carlotti F, de Koning EJ, Slagboom PE, Heijmans BT, Chuva de Sousa Lopes SM: **DNA Methylation Landscapes of Human Fetal Development**. *PLoS Genet* 2015, **11**(10):e1005583.

[9] Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CC, O'Donovan MC, Bray NJ, Mill J: **Methylomic trajectories across human fetal**

brain development. *Genome Res* 2015, **25**(3):338–352.

[10] Takasato M, Er PX, Chiu HS, Maier B, Baillie GJ, Ferguson C, Parton RG, Wolvetang EJ, Roost MS, Chuva de Sousa Lopes SM *et al*: **Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis**. *Nature* 2015, **526**(7574):564–568.

[11] van den Berg CW, Okawa S, Chuva de Sousa Lopes SM, van Iperen L, Passier R, Braam SR, Tertoolen LG, del Sol A, Davis RP, Mummery CL: **Transcriptome of human foetal heart compared with cardiomyocytes from pluripotent stem cells**. *Development* 2015, **142**(18):3231–3238.

[12] Yan L, Guo H, Hu B, Li R, Yong J, Zhao Y, Zhi X, Fan X, Guo F, Wang X *et al*: **Epigenomic landscape of human fetal brain, heart, and liver**. *J Biol Chem* 2015.

[13] Lokk K, Modhukur V, Rajashekar B, Martens K, Magi R, Kolde R, Koltsina M, Nilsson TK, Vilo J, Salumets A *et al*: **DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns**. *Genome Biol* 2014, **15**(4):r54.

[14] Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B: **Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues**. *Nat Genet* 2013, **45**(10):1198–1206.

[15] Rakyan VK, Down TA, Thorne NP, Flicek P, Kulesha E, Graf S, Tomazou EM, Backdahl L, Johnson N, Herberth M *et al*: **An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs)**. *Genome Res* 2008, **18**(9):1518–1529.

[16] Slieker RC, Bos SD, Goeman JJ, Bovee JV, Talens RP, van der Breggen R, Suchiman HE, Lameijer EW, Putter H, van den Akker EB *et al*: **Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array**. *Epigenetics Chromatin* 2013, **6**(1):26.

[17] Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S: **Induction of pluripotent stem cells from adult human fibroblasts by defined factors**. *Cell* 2007, **131**(5):861–872.

[18] Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors**. *Cell* 2006, **126**(4):663–676.

[19] Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R *et al*: **Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution**. *Cell Stem Cell* 2007, **1**(1):55–70.

[20] Choi J, Lee S, Mallard W, Clement K, Tagliazucchi GM, Lim H, Choi IY, Ferrari F, Tsankov AM, Pop R *et al*: **A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs**. *Nat Biotechnol* 2015.

[21] Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiuwu O, Richter L, Zhang J *et al*: **Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures**. *Cell Stem Cell* 2009, **5**(1):111–123.

[22] Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC: **Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells**. *PLoS One* 2010, **5**(2):e8975.

[23] Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR: **Transcriptional signature and memory retention of human-induced pluripotent stem cells**. *PLoS One* 2009, **4**(9):e7076.

[24] Nishizawa M, Chonabayashi K, Nomura M, Tanaka A, Nakamura M, Inagaki A, Nishikawa M, Takei I, Oishi A, Tanabe K *et al*: **Epigenetic Variation between Human Induced Pluripotent Stem Cell Lines Is an Indicator of Differentiation Capacity**. *Cell Stem Cell* 2016.

[25] Hiler D, Chen X, Hazen J, Kupriyanov S, Carroll PA, Qu C, Xu B, Johnson D, Griffiths L, Frase S *et al*: **Quantification of Retinogenesis in 3D Cultures Reveals Epigenetic Memory and Higher Efficiency in iPSCs Derived from Rod Photoreceptors**. *Cell Stem Cell* 2015, **17**(1):101–115.

[26] Hu Q, Friedrich AM, Johnson LV, Clegg DO: **Memory in induced pluripotent stem cells: reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation**. *Stem Cells* 2010, **28**(11):1981–1991.

[27] Hargus G, Ehrlich M, Arauzo-Bravo MJ, Hemmer K, Hallmann AL, Reinhardt P, Kim KP, Adachi K, Santourlidis S, Ghanjati F *et al*: **Origin-dependent neural cell identities in differentiated human iPSCs in vitro and after transplantation into the mouse brain**. *Cell Rep* 2014, **8**(6):1697–1703.

[28] Bar-Nur O, Russ HA, Efrat S, Benvenisty N: **Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells**. *Cell Stem Cell* 2011, **9**(1):17–23.

[29] Quattrocelli M, Swinnen M, Giacomazzi G, Camps J, Barthelemy I, Ceccarelli G, Caluwe E, Grosemans H, Thorrez L, Pelizzo G *et al*: **Mesodermal iPSC-derived progenitor cells functionally regenerate cardiac and skeletal muscle**. *J Clin Invest* 2015, **125**(12):4463–4482.

[30] Kim K, Zhao R, Doi A, Ng K, Unternaehrer J, Cahan P, Huo H, Loh YH, Aryee MJ, Lensch MW *et al*: **Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells**. *Nat Biotechnol* 2011, **29**(12):1117–1119.

[31] Vitaloni M, Pulecio J, Bilic J, Kuebler B, Laricchia-Robbio L, Izpisua Belmonte JC: **MicroRNAs contribute to induced pluripotent stem cell somatic donor memory**. *J Biol Chem* 2014, **289**(4):2084–2098.

[32] Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T *et al*: **Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells**. *Nat Biotechnol* 2010, **28**(8):848–855.

[33] Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ *et al*: **Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells**. *Nat Cell Biol* 2011, **13**(5):541–549.

[34] Rouhani F, Kumasaka N, de Brito MC, Bradley A, Vallier L, Gaffney D: **Genetic background drives transcriptional variation in human induced pluripotent stem cells**. *PLoS Genet* 2014, **10**(6):e1004432.

[35] Tiemann U, Wu G, Marthaler AG, Scholer HR, Tapia N: **Epigenetic Aberrations Are Not Specific to Transcription Factor-Mediated Re-**

programming. *Stem Cell Reports* 2016, **6**(1):35–43.

[36] Dambrot C, van de Pas S, van Zijl L, Brandl B, Wang JW, Schalij MJ, Hoeben RC, Atsma DE, Mikkers HM, Mummery CL *et al*: **Polycistronic lentivirus induced pluripotent stem cells from skin biopsies after long term storage, blood outgrowth endothelial cells and cells from milk teeth**. *Differentiation* 2013, **85**(3):101–109.

[37] Elliott DA, Braam SR, Koutsis K, Ng ES, Jenny R, Lagerqvist EL, Biben C, Hatzistavrou T, Hirst CE, Yu QC *et al*: **NKX2-5(eGFP/w) hESCs for isolation of human cardiac progenitors and cardiomyocytes**. *Nat Methods* 2011, **8**(12):1037–1040.

[38] Moretti A, Bellin M, Welling A, Jung CB, Lam JT, Bott-Flugel L, Dorn T, Goedel A, Hohnke C, Hofmann F *et al*: **Patient-specific induced pluripotent stem-cell models for long-QT syndrome**. *N Engl J Med* 2010, **363**(15):1397–1409.

[39] Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F: **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**. *Genome Biol* 2002, **3**(7):RESEARCH0034.

[40] Szuhai K, Tanke HJ: **COBRA: combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping**. *Nat Protoc* 2006, **1**(1):264–275.

[41] Duggal G, Warrier S, Ghimire S, Broekaert D, Van der Jeught M, Lierman S, Deroo T, Peelman L, Van Soom A, Cornelissen R *et al*: **Alternative Routes to Induce Naive Pluripotency in Human Embryonic Stem Cells**. *Stem Cells* 2015, **33**(9):2686–2698.

[42] Hansen KD, Aryee M: **Minfi: Analyze Illumina's 450K methylation arrays.** *R package version* 2013, **1.16.0**.

[43] van Iterson M, Tobi EW, Slieker RC, den Hollander W, Luijk R, Slagboom PE, Heijmans BT: **MethylAid: visual and interactive quality control of large Illumina 450k datasets**. *Bioinformatics* 2014, **30**(23):3435–3437.

[44] Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM, Hansen KD: **Functional normalization of 450k methylation array data improves replication in large cancer studies**. *Genome Biol* 2014, **15**(12):503.

[45] Genome of the Netherlands C: **Whole-genome sequence variation, population structure and demographic history of the Dutch population**. *Nat Genet* 2014, **46**(8):818–825.

[46] Wickham H: **ggplot2: Elegant Graphics for Data Analysis**: Springer New York; 2009.

[47] Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**(1):139–140.

[48] Robinson MD, Smyth GK: **Small-sample estimation of negative binomial dispersion, with applications to SAGE data**. *Biostatistics* 2008, **9**(2):321–332.

[49] Johannesson B, Sagi I, Gore A, Paull D, Yamada M, Golan-Lev T, Li Z, LeDuc C, Shen Y, Stern S *et al*: **Comparable frequencies of coding mutations and loss of imprinting in human pluripotent cells derived by nuclear transfer and defined factors**. *Cell Stem Cell* 2014, **15**(5):634–642.

[50] Schroeder DI, Blair JD, Lott P, Yu HO, Hong D, Crary F, Ashwood P, Walker C, Korf I, Robinson WP *et al*: **The human placenta methylome**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(15):6037–6042.

[51] Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ *et al*: **Human genomics. The human transcriptome across tissues and individuals**. *Science* 2015, **348**(6235):660–665.

[52] Eng LF, Ghirnikar RS: **GFAP and astrogliosis**. *Brain Pathol* 1994, **4**(3):229–237.

[53] Roessmann U, Velasco ME, Sindely SD, Gambetti P: **Glial fibrillary acidic protein (GFAP) in ependymal cells during development. An immunocytochemical study**. *Brain Res* 1980, **200**(1):13–21.

[54] Scarr E, Craig JM, Cairns MJ, Seo MS, Galati JC, Beveridge NJ, Gibbons A, Juzva S, Weinrich B, Parkinson-Bates M *et al*: **Decreased cortical muscarinic M1 receptors in schizophrenia are associated with changes in gene promoter methylation, mRNA and gene targeting microRNA**. *Translational psychiatry* 2013, **3**:e230.

[55] Colaneri A, Wang T, Pagadala V, Kittur J, Staffa NG, Jr., Peddada SD,

Isganaitis E, Patti ME, Birnbaumer L: **A minimal set of tissue-specific hypomethylated CpGs constitute epigenetic signatures of developmental programming**. *PLoS One* 2013, **8**(9):e72670.

[56] Feng G, Zhang Y, Yuan H, Bai R, Zheng J, Zhang J, Song M: **DNA methylation of trefoil factor 1 (TFF1) is associated with the tumorigenesis of gastric carcinoma**. *Mol Med Rep* 2014, **9**(1):109–117.

[57] Verzi MP, Stanfel MN, Moses KA, Kim BM, Zhang Y, Schwartz RJ, Shivdasani RA, Zimmer WE: **Role of the homeodomain transcription factor Bapx1 in mouse distal stomach development**. *Gastroenterology* 2009, **136**(5):1701–1710.

[58] Nishino K, Toyoda M, Yamazaki-Inoue M, Fukawatase Y, Chikazawa E, Sakaguchi H, Akutsu H, Umezawa A: **DNA methylation dynamics in human induced pluripotent stem cells over time**. *PLoS Genet* 2011, **7**(5):e1002085.

[59] Safran M, Chalifa-Caspi V, Shmueli O, Olender T, Lapidot M, Rosen N, Shmoish M, Peter Y, Glusman G, Feldmesser E *et al*: **Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE**. *Nucleic acids research* 2003, **31**(1):142–146.

[60] Flegel C, Schobel N, Altmuller J, Becker C, Tannapfel A, Hatt H, Gisselmann G: **RNA-Seq Analysis of Human Trigeminal and Dorsal Root Ganglia with a Focus on Chemoreceptors**. *PLoS One* 2015, **10**(6):e0128951.

[61] O'Connor TP, Cockburn K, Wang W, Tapia L, Currie E, Bamji SX: **Semaphorin 5B mediates synapse elimination in hippocampal neurons**. *Neural development* 2009, **4**:18.

[62] Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J *et al*: **Dynamic changes in the human methylome during differentiation**. *Genome Res* 2010, **20**(3):320–331.

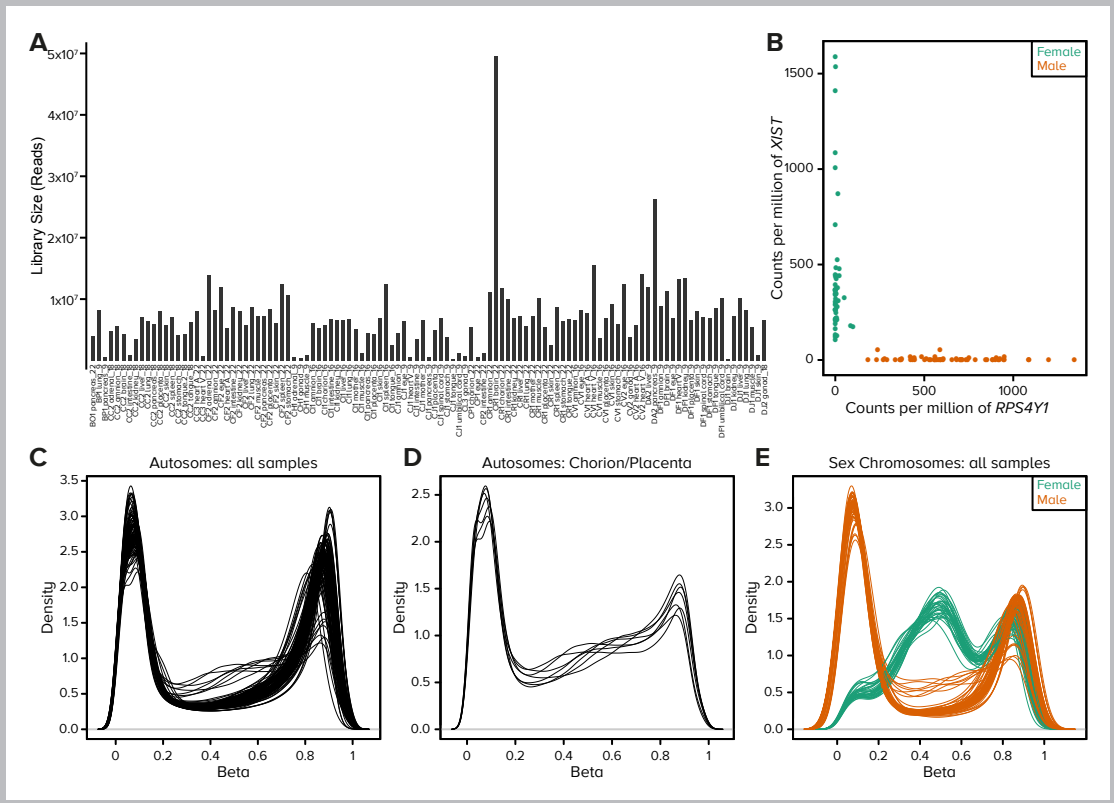
[63] Hou Y, Guo H, Cao C, Li X, Hu B, Zhu P, Wu X, Wen L, Tang F, Huang Y *et al*: **Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas**. *Cell research* 2016, **26**(3):304–319.

[64] Hu Y, Huang K, An Q, Du G, Hu G, Xue J, Zhu X, Wang CY, Xue Z,

Fan G: **Simultaneous profiling of transcriptome and DNA methylome from a single cell**. *Genome Biol* 2016, **17**(1):88.

[65] Hewitt KJ, Garlick JA: **Cellular reprogramming to reset epigenetic signatures**. *Molecular aspects of medicine* 2013, **34**(4):841–848.

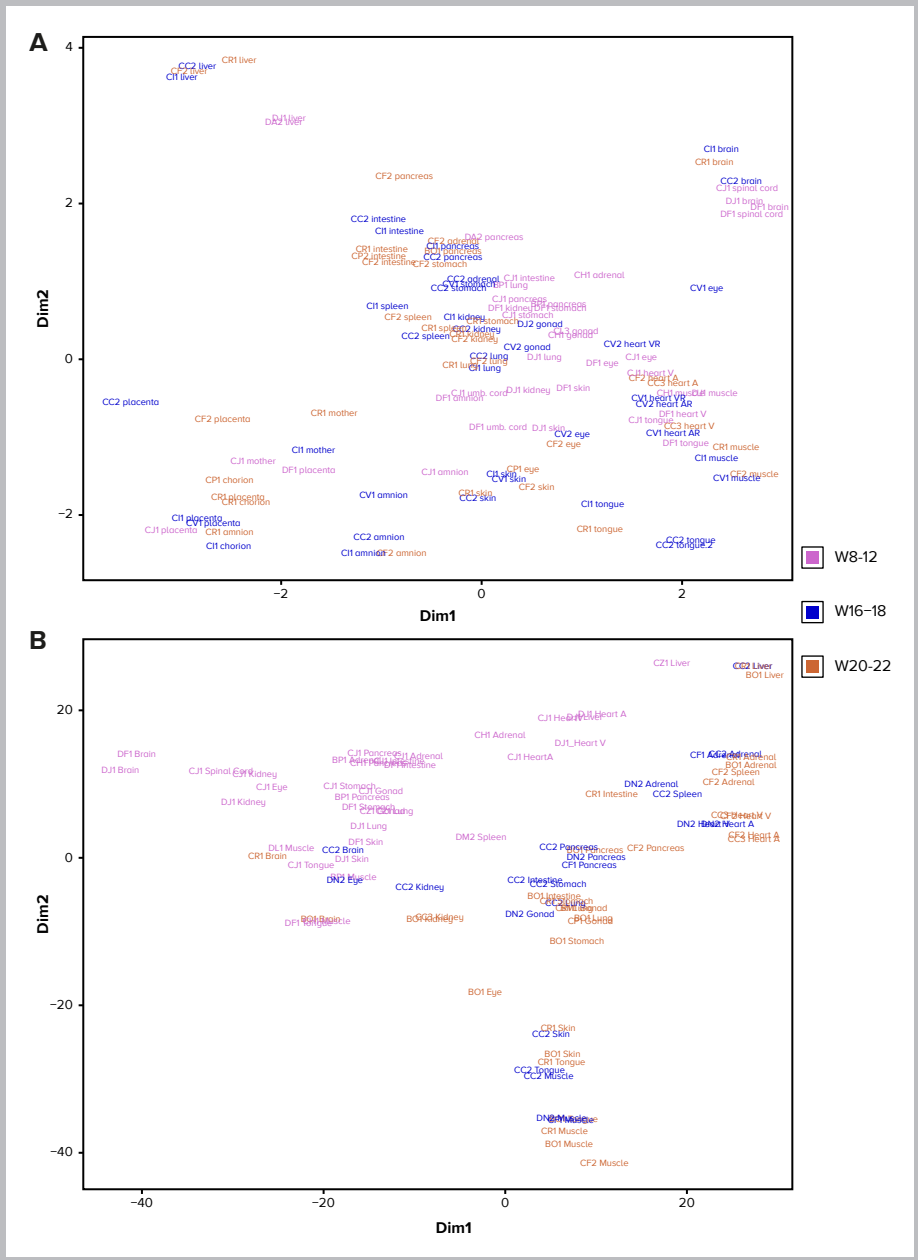
[66] Firas J, Liu X, Polo JM: **Epigenetic memory in somatic cell nuclear transfer and induced pluripotency: evidence and implications**. *Differentiation* 2014, **88**(1):29–32.


A
Library Size (Reads)
5x10⁷
4x10⁷
3x10⁷
2x10⁷
1x10⁷
B
Female
Male
Counts per million of XIST
Counts per million of RPS4Y1
1500
1000
500
0
0
500
1000
C
Autosomes: all samples
Density
Beta
D
Autosomes: Chorion/Placenta
Density
Beta
E
Sex Chromosomes: all samples
Female
Male
Density
Beta

[Supplementary Fig. 1]

Quality control of transcriptome and DNA methylome data.

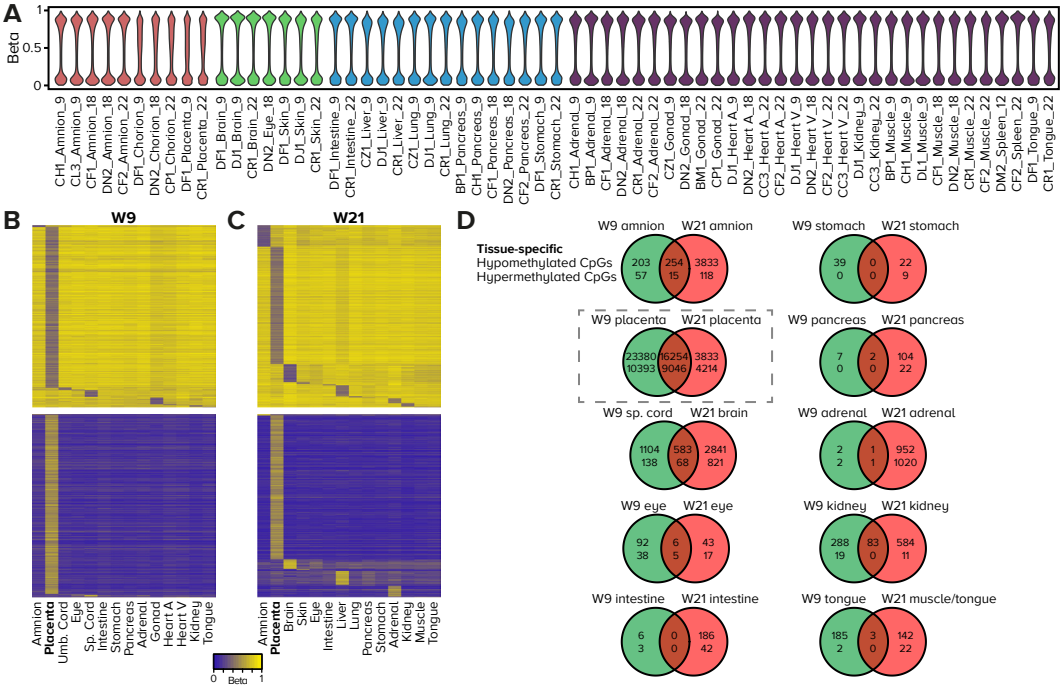
[A] Library sizes of the 111 NGS transcriptional profiles. **[B]** Expression levels (counts per million, CPM) of the female gene XIST and the male gene RPS4Y1. **[C–D)** Density plots of the beta values (450k DNA methylation) of the autosomes of all samples **[C]** and only of the chorion and placenta **[D]**. **[E]** Density plot of the beta values (450k DNA methylation) of the sex chromosomes of all samples separated by sex (male and female).


A
Dim2
Dim1
W8-12
W16–18
W20-22
B
Dim2
Dim1

[Supplementary Fig. 2]

Multidimensional scaling.

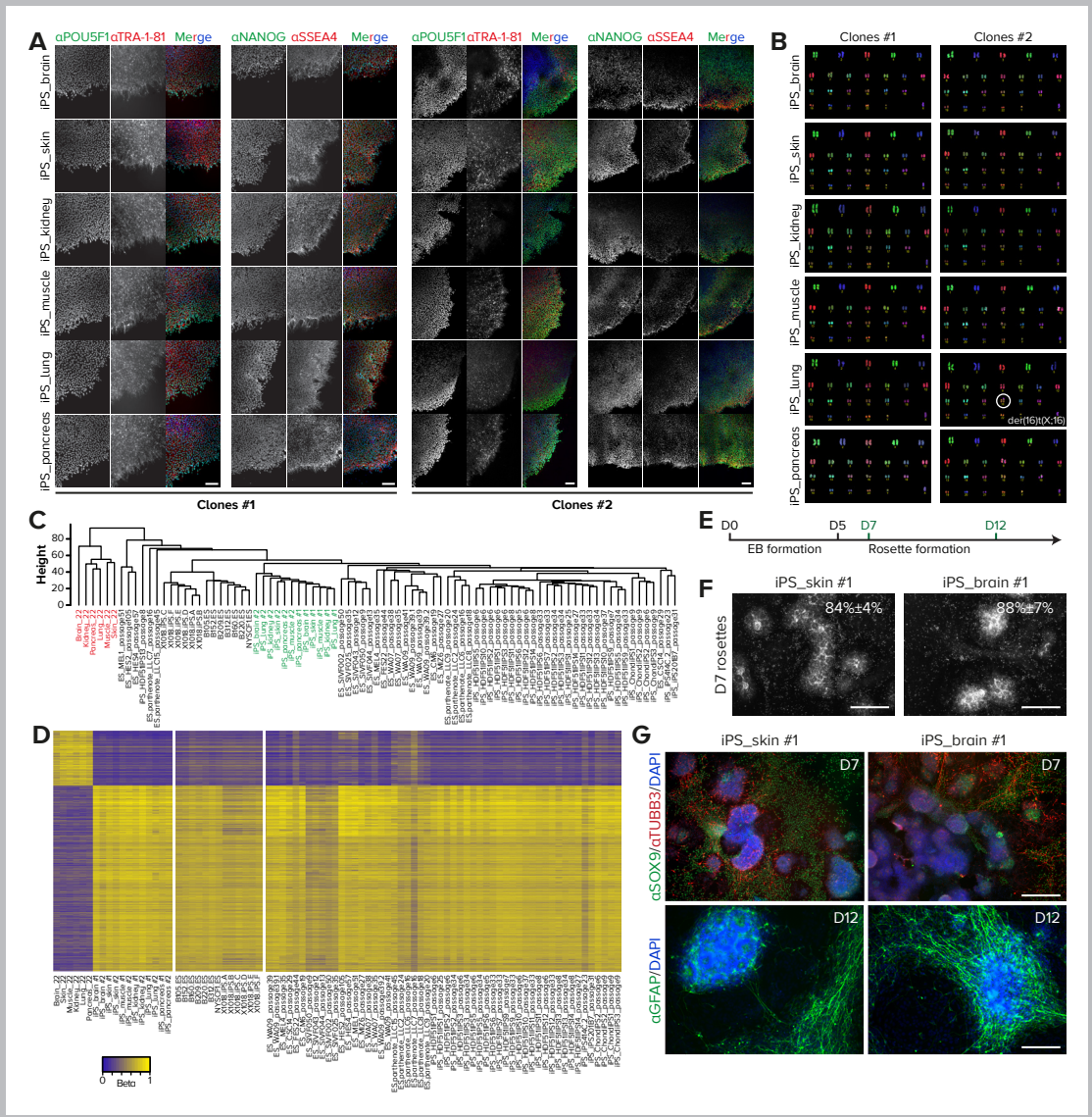
[A–B] Multidimensional scaling (Euclidian distance) of the 111 transcriptional profiles (including extra-embryonic samples) **[A]** and the 87 DNA methylation profiles of the embryonic samples only (excluding extraembryonic samples) **[B]**. Each sample is named as «embryo ID code_tissue name» and the color code depicts the gestational age group (W8–12 in pink; W16–18 in blue, W20–22 in orange).


A
Beta
1
0.5
0
CH1_Amnion_9
CL3_Amnion_9
CF1_Amnion_18
DN2_Amnion_18
CF2_Amnion_22
DF1_Chorion_9
DN2_Chorion_18
CP1_Chorion_22
DF1_Placenta_9
CR1_Placenta_22
DF1_Brain_9
DJ1_Brain_9
CR1_Brain_22
DN2_Eye_18
DF1_Skin_9
DJ1_Skin_9
CR1_Skin_22
DF1_Intestine_9
CR1_Intestine_22
CZ1_Liver_9
DJ1_Liver_9
CR1_Liver_22
CZ1_Lung_9
DJ1_Lung_9
CR1_Lung_22
BP1_Pancreas_9
CH1_Pancreas_9
CF1_Pancreas_18
DN2_Pancreas_18
CF2_Pancreas_22
DF1_Stomach_9
CR1_Stomach_22
CH1_Adrenal_9
BP1_Adrenal_9
CF1_Adrenal_18
DN2_Adrenal_18
CR1_Adrenal_22
CF2_Adrenal_22
CZ1_Gonad_9
DN2_Gonad_18
BM1_Gonad_22
CP1_Gonad_22
DJ1_Heart A_9
DN2_Heart A_18
CC3_Heart A_22
CF2_Heart A_22
DJ1_Heart V_9
DN2_Heart V_18
CF2_Heart V_22
CC3_Heart V_22
DJ1_Kidney_9
CC3_Kidney_22
BP1_Muscle_9
CH1_Muscle_9
DL1_Muscle_9
CF1_Muscle_18
DN2_Muscle_18
CR1_Muscle_22
CF2_Muscle_22
DM2_Spleen_12
CF2_Spleen_22
DF1_Tongue_9
CR1_Tongue_22
B
W9
C
W21
Amnion
Placenta
Umb. Cord
Eye
Sp. Cord
Intestine
Stomach
Pancreas
Adrenal
Gonad
Heart A
Heart V
Kidney
Tongue
Amnion
Placenta
Brain
Skin
Eye
Intestine
Liver
Lung
Pancreas
Stomach
Adrenal
Kidney
Muscle
Tongue
0
Beta
1
D
Tissue-specific
Hypomethylated CpGs
Hypermethylated CpGs
W9 amnion
W21 amnion
203
57
254
15
3833
118
W9 stomach
W21 stomach
39
0
0
0
22
9
W9 placenta
W21 placenta
23380
10393
16254
9046
3833
4214
W9 pancreas
W21 pancreas
7
0
2
0
104
22
W9 sp. cord
W21 brain
1104
138
583
68
2841
821
W9 adrenal
W21 adrenal
2
2
1
1
952
1020
W9 eye
W21 eye
92
38
6
5
43
17
W9 kidney
W21 kidney
288
19
83
0
584
11
W9 intestine
W21 intestine
6
3
0
0
186
42
W9 tongue
W21 muscle/tongue
185
2
3
0
142
22

[Supplementary Fig. 3]

DNA methylation signatures during human fetal development.

[A] Violin plot depicting the distribution of the DNA methylation levels of the remaining samples that were excluded from the three isogenic sets shown in Fig. 3A. **[B–C]** Heatmaps displayed in Fig. 3B and 3D including the probes of the placenta. **[D]** Venn diagrams showing the overlap between W9 and W21 of tissue-specific hypomethylated CpGs (top) and of hypermethylated CpGs (bottom). The dashed box highlights the large number of tissue specific CpGs in the placenta.


A
αPOU5F1 αTRA-1-81 Merge
αNANOG αSSEA4 Merge
iPS_brain
iPS_skin
iPS_kidney
iPS_muscle
iPS_lung
iPS_pancreas
Clones #1
Clones #2
B
Clones #1
Clones #2
der(16)t(X;16)
C
Height
D
Beta
E
D0
D5
D7
D12
EB formation
Rosette formation
F
iPS_skin #1
iPS_brain #1
D7 rosettes
84%±4%
88%±7%
G
iPS_skin #1
iPS_brain #1
αSOX9/αTUBB3/DAPI
αGFAP/DAPI
D7
D12

[Supplementary Fig. 4]

Isogenic sets of human iPSCs.

[A] Single and merged channels of the immunofluorescence of each of the 12 iPSC clones for POU5F1 and TRA-1-81, and NANOG and SSEA4. Scale bars 100 µm. **[B]** Karyograms of the 12 iPSC clones using COBRA. **[C]** Hierarchical clustering of the 12 iPSC clones (in green) and their tissues of origin (in red) together with two external 450k datasets [5, 49]. **[D]** DNA methylation levels of two external datasets regarding the hiPSC-specific hypomethylated (2852 CpGs) and hypermethylated (8772 CpGs) CpGs depicted in Fig. 4F. **[E]** Cartoon depicting the neural induction protocol followed and time points of analysis (in green). **[F]** Dark field picture of embryoid bodies containing rosettes present in culture after 7 days of neural differentiation. Scale bars: 500 µm. **[G]** Immunofluorescence for neural markers SOX9, TUBB3 and GFAP after 7 (D7) and 12 (D12) days of neural differentiation. Scale bars: 500 µm (D7) and 200 µm (D12).

CHAPTER 6
GENERAL DISCUSSION

VASCULARIZATION OF THE HUMAN PANCREAS EARLY DURING FETAL DEVELOPMENT ILLUSTRATES ITS IMPORTANCE FOR TRANSPLANTATION AND STEM CELL DIFFERENTIATION

One of the major findings of this thesis with respect to the pancreas was the confirmation of the essential role of the vasculature in development. Vascularization of the pancreas including the islets of Langerhans is of paramount importance for the physiology of the whole pancreas, and has effects on islet transplantation as well as on protocols for differentiating pluripotent stem cells towards insulin-producing β-cells [1, 2]**. Most insights into how the blood vascular system is established during development have been derived from model organisms such as rodents and chick. However, knowledge about human development is also very important since endothelial cells have functions other than just the transport of oxygen and nutrients between blood and tissue-specific cells, including direct reciprocal signaling between endothelial and tissue-specific cells** [3]**, which may be species-specific. It has been shown that the blood vasculature is particularly important at two stages of pancreas development at least in mice: (1) the dorsal aorta is necessary for the specification of the pancreatic buds from the foregut epithelium** [4, 5]**; and (2) later during branching morphogenesis when blood vessels control the development of the exocrine compartment via inhibition of acinar differentiation** [6]**. Although the contribution of blood vessels to the formation of the islets of Langerhans and the arrangement of the endocrine cells during pancreas organogenesis remains unknown, their function in adult islets is irreplaceable due to the endocrine function of β-cells** [7]**. Islets of Langerhans harbor a highly organized, dense microvasculature, which is essential for the core functions of β-cells, namely sensing glucose levels and secreting**

insulin [8, 9]**. Besides interaction via signaling molecules, there is also paracrine communication between the endothelial and endocrine cells mediated through a basement membrane (BM)** [10, 11]**. The BM is also involved in proliferation, survival and motility of β-cells. Besides the blood vasculature, there is a second vascular system, the lymphatics, that perfuses the body and thus also the pancreas. Although also important, much less is known about it, partly due the lack of reliable markers** [12]**, but it is known to have functions in facilitating transport of cells of the immune system and draining fluid excess from tissues. The discovery of two markers, namely LYVE1 and Podoplanin (PDPN), has advanced research of the lymphatic system in recent years** [13]**, but in the human pancreas this is mostly limited to the adult** [14, 15]**.**

We addressed in chapter 2 the highly dynamic progress of angiogenesis and lymphangiogenesis in the peri- and intra-pancreatic mesenchyme (PPM and IPM) between 9 and 22 weeks of gestation (W9–W22). Blood vessels populated the PPM and the IPM as early as W9, in contrast to lymphatic vessels: although they were present in the PPM at W9, they were not observed in the IPM until W14.5. Differences between the two vascular networks were also observed regarding two aspects of maturation, namely the association with smooth muscle cells and the formation of a BM consisting of collagen type IV. Whereas mature blood vessels were found in the PPM and IPM as early as W9, lymphatic vessels were again delayed. A BM and smooth muscle cells around lymphatic vessels in the PPM were observed at W11 and W14.5, respectively, whereas there were no mature vessels in the IPM until W17. Interestingly, the first islets of Langerhans appeared at W12 and by W14.5 they have already established their extensive microvasculature. In contrast, there was no islet penetration by lymphatic vessels at W12, but 3D-reconstructions revealed small lymphatic capillaries close to islets between W17 and W22. This study was the first systematic examination of the establishment of the

two vascular systems in the human fetal pancreas. We showed that they both undergo their own, probably developmentally independent, colonization of the pancreas and maturation.

In contrast to the whole pancreas, where perfusion of blood is restored rapidly after organ transplantation, the microvasculature of the islets of Langerhans is severely disrupted during the isolation process [1]. Insufficient access to the blood flow in the post-transplant period has been shown to be a major contributor to the low survival of the transplanted islets [16]. Although transplantation sites close to vascular blood supply improve survival, the damaging hypoxic conditions in the core of the islets due to the disrupted microcirculation remain [17]. Furthermore, it has been reported that increased VEGF-A expression stimulates revascularization of the transplanted islets and may improve the outcome of islet transplantation [9, 18]. The immediate establishment of a microcirculation, as we observed in chapter 2, underlines the paramount importance of the blood supply for the endocrine cells. Culturing islets after isolation from the pancreas using 3D-scaffolds may stimulate revascularization and improve the outcome of the transplantation [19, 20]. To successfully culture islets using 3D-scaffolds would require engineering such that it would mimic the *in vivo* niche, which is substantially determined by the interaction with neighboring cells and the extracellular matrix (ECM) [21]. In this context, the peri-insular and the vascular BMs of the islets are very important for the endocrine function, and their restoration after isolation and transplantation is necessary [11, 22]. The major components of the pancreatic BMs are laminins and collagen type IV [23]. Integrins are receptors consisting of αβ heterodimers that bind to the components of the ECM, and thereby provide physical attachment to the ECM but also allow signal transduction from the ECM to the cells [24]. Further insights about favorable ECM compositions may be gained by investigating the integrin composition of endocrine cells. It has been shown that human adult and fetal islets are positive for the β1, α1,

α3, α5, α6 and αV subunits, and that they are involved in β-cell survival and function [10, 25–28]**. In chapter 2, we also investigated the BMs formed by the blood and lymphatic vessels in the pancreatic mesenchyme. But, there is still a lack of comprehensive studies on the dynamics of the BM formation of the islet microcirculation. Future research should focus on the dynamics of the formation of islet BMs, particularly if they are of different composition at different stages of development. One intriguing detail of human islets is the presence of a double BM: in contrast to rodents that only harbor an endothelial BM surrounding the microvessels, human islets contain a double BM with one derived from endothelial cell and one from endocrine cells** [29]**. The double membrane function remains unclear, but it may indicate which physiological requirements in terms of BM signals are needed for the isolated islets during the culture period prior to transplantation or at the transplantation site.**

An alternative to primary islet transplantation to treat diabetes is the use of pluripotent stem cells (PSCs), which can be induced to form insulin-secreting β-cells. To date however, these cells have been functionally immature exhibiting a fetal phenotype [30, 31]**. The early vascularization of the pancreas and the islets in development underlines the importance of endothelial signals, either directly or through BMs, for the differentiation towards functional endocrine cells. And it may provide vital clues on how to promote maturation of PSC-derived insulin-secreting β-cells in culture or post transplantation. Previously, differentiation protocols have tried to imitate the physiological environment by co-culture with endothelial cells or different ECM materials** [32, 33]**. Nowadays, more sophisticated methods mimicking organogenesis have been developed: (1) 3D-culture allowed the self-organization of cells derived from PSCs to form organoids, including pancreatic organoids** [34, 35]**, (2)** ***in vitro*****-derived organ buds that generated functional organs upon transplantation** [36]**. Organoids in particular still show limited vascularization due to simple physical restraints.**

Therefore, 3D-scaffolds mimicking the physiological ECM may help to grow larger and more complex structures from organoids allowing a fully functional vascular network to develop.

It has been shown that lymphatic vessels invade the pancreas in order to drain fluid from the interstitial space [15]**. These lymphatic vessels reside in the inter- and intralobular space, and islets do not harbor a lymphatic vascular network. We show in chapter 2 that both, the PPM and the IPM, contain lymphatic vessels by W22. Furthermore, 3D-reconstruction confirmed that islets are not penetrated by lymphatic vessels; however, close proximity was observed with developing islets between W17 and W22. Importantly, we only observed lymphatic capillaries and pre-collecting vessels, and it remains to be determined at what stage mature collecting vessels are formed. Finally, we showed that PDPN and LYVE1 provide a good combination of markers to investigate lymphangiogenesis** [13]**. Interestingly, similar findings have recently turned out to apply to the heart** [37]**. Pancreatic pathologies, particularly cancer and pancreatitis, are very severe. In order to provide the best treatment to patients, detailed understanding of the lymphatic system is indispensable and more research is needed** [14]**. Interestingly, it has been observed that islets transplanted under the renal capsule, in contrast to endogenous islets, showed lymphatic vascularization** [38]**. Although its function remains to be determined, this further illustrates that understanding the lymphatic system is paramount to evaluate the spread of tumors as well as immune reactions during transplantation.**

TRANSCRIPTIONAL AND DNA METHYLATION PROFILES DURING HUMAN FETAL DEVELOPMENT AND THEIR IMPLICATIONS FOR STEM CELLS

Knowledge of the development of specific organs and cell types *in vivo* is of key importance for stem cell biologists in order to develop proper differentiation protocols. As shown in chapter 2, knowledge of the niche composition, in which a particular cell type differentiates, is important in terms of growth factors and ECM components that need to be present in a PSC differentiation protocol. Secondly, characterizing organs and cells at the molecular level at different stages of development allows benchmarking the outcome of differentiation at different stages. The primary and ultimate readout of stem cell differentiations is functional testing of the stem cell derivatives. However, since it has been shown that the transcriptional profiles of differentiated cells resemble their fetal, and not adult, counterpart, determining the gene expression and epigenetic patterns represents an essential information set that is complementary to functional characterization [30, 39]**.**

TRANSCRIPTIONAL SIGNATURES DURING HUMAN FETAL DEVELOPMENT AND HOW THEY COMPARE TO STEM CELL DERIVATIVES

Significant advances in technology over the last several years and better access to human fetal and adult material have allowed research on the identification of tissue-specific transcriptional states using microarray and next-generation sequencing (NGS) [40–43]**. This knowledge on the physiological context of gene expression is particularly important to identify aberrant patterns in diseases such as cancer** [44, 45]**. However, in light of the immature state of all stem cell derivatives to date, the field**

would benefit from the generation of genome-wide transcriptional data during human fetal development. In chapter 3, we generated a comprehensive transcriptional atlas of 21 different fetal organs using NGS. Apart from the large number of organs, the presence of three different time points during first and second trimester, namely W9, W18 and W22, is another outstanding feature of this dataset. Furthermore, we developed an algorithm called KeyGenes, which determines the transcriptional identity of a training set on the basis of a minimal number of genes and predicts the identity of queried samples, referred to as a test set. By dividing our atlas into a training set and a test set, we identified and validated classifier genes that were able to characterize the 21 organs of our atlas. Intriguingly, this set of classifier genes was comprised of only about 90 genes involved in a wide variety of biological processes. These combinations of genes per organ were even capable of distinguishing between first and second trimester fetuses in the majority of organs illustrating the dynamic nature of the gene expression landscape during development. However, the fact that KeyGenes identified adult samples with high prediction scores also shows that there is a basic organ transcriptional signature as early as W9, which is still present and characteristic of each organ in adulthood.

The human fetal atlas was pivotal in being able to use KeyGenes for assessing the cellular outcome of differentiation protocols from PSCs, since the comparison to adult transcriptional data only may be misleading and does not allow developmental staging. Although the prediction was effective on both FACS-sorted cells and specific cell populations, the highest identity scores were obtained with intestinal organoids. This illustrates how the self-organization capability of cells in advanced organoid technology is able to generate more complex structures that actually resemble organs [34]. The applicability of KeyGenes to benchmark PSC-derived organoids was further confirmed by Takasato *et al.* [46]. There, we generated kidney or-

ganoids from human PSCs and were able to show that morphologically they followed human nephrogenesis. However, based on the transcriptional identity, we showed using KeyGenes that the kidney organoids could be assigned high identity scores for kidney, but most importantly they were almost indistinguishable from human first trimester kidney. This again confirmed the immaturity of the stem cell derivatives [30, 39]**.**

The transcriptional atlas generated in chapter 3 provides a very valuable resource for researchers from a wide variety of biological and medical fields. It will help increase knowledge of the molecular basis of human development and what it takes to form mature cells and organs from progenitor cells. Three particular features are outstanding: (1) the number of tissues allowing comparative studies that are context-dependent, (2) the material was derived from elective abortions and only few individuals minimizing genetic variability, and (3) the different time points during first and second trimester, which enables the identification of gene expression trajectories over time. The combination of our fetal data with existing adult datasets expands the latter in order to evaluate to what extent fetal signatures are maintained in adult organs [40–43]**. Although ethically and practically difficult, it is desirable that the gap between our last time point (W22) and adulthood is filled, particularly with third trimester tissue. Furthermore, a drawback of our material is that we collected entire organs and most tissues are heterogeneous in cell composition. However, that KeyGenes was capable to predict the identity of specific parts of organs and cell populations suggests that a basic transcriptional barcode exists. Nevertheless, separating organs into distinct anatomical and functional regions by microdissection will generate an increasingly comprehensive and complex atlas. In the future, advancing technology of single-cell RNA-sequencing will not only allow characterization of the functional cell types of an organ but also shed light on different progenitor cells from which organs develop** [47]**.**

We showed that the training set used by KeyGenes can be expanded with adult data in order to assign the developmental stage to the cellular outcome of stem cell differentiation. However, we did encounter problems when we applied KeyGenes on PSC-derived endocrine cells, which illustrates the limitation of using whole organs: it is difficult to predict the identity of a cell that makes up only a small fraction of the whole organ. However, the flexibility of KeyGenes also allowed us to circumvent this problem by incorporating the transcriptional data of adult islets, which provided the biological context for this test set. The outstanding feature of flexibility of the training set opens a wide variety of exciting possibilities for our tool with regard to the progressing single-cell RNA-sequencing technology [47]**. For organs like the pancreas, harboring two different compartments, endocrine and exocrine, with many different functional cells, this is of particular interest.**

COMPUTATIONAL TOOLS IN STEM CELL RESEARCH

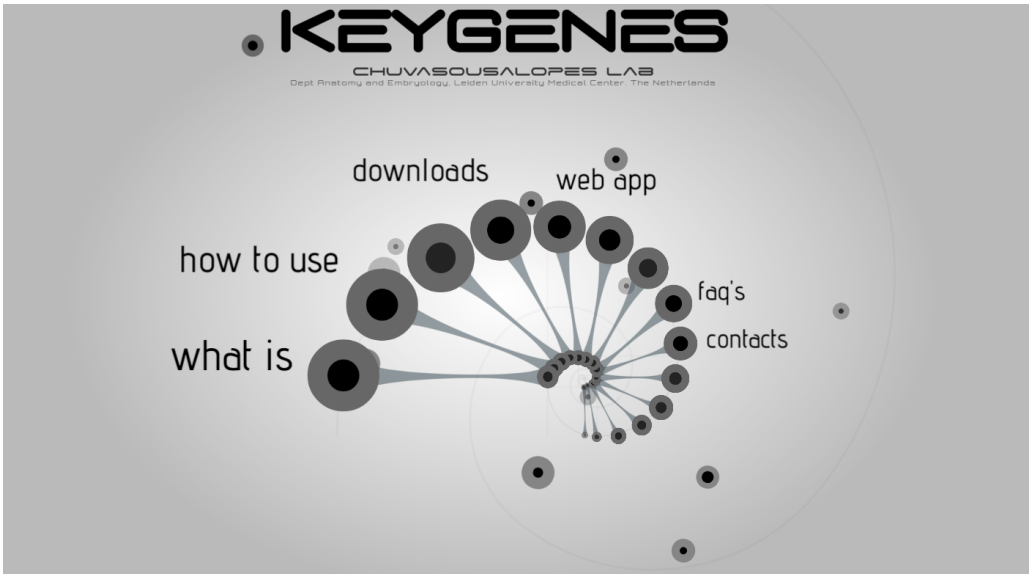
In order to provide accessibility to KeyGenes for the community, we designed the website www.keygenes.nl [Fig. 1A]**. We not only made the actual algorithm and the entire collection of different training sets used by KeyGenes available for download** [Fig. 1B]**, but we additionally provided an easy-to-use online web application for users with transcriptional data but no experience in bioinformatics** [Fig. 1C]**.**

The recent progress in the generation of genome-wide sequencing data at the DNA and RNA level including epigenetics («big data») poses the challenge of analyzing the growing amount of data in a biologically meaningful way. In order to process their own data but also other data related to their research available on public databases, scientists depend more and more on algorithms incorporated in computational platforms. Two applications have been particularly interesting for stem cell researchers [48]**: (1) several bioinformatic tools have been developed to identify appropriate genetic**

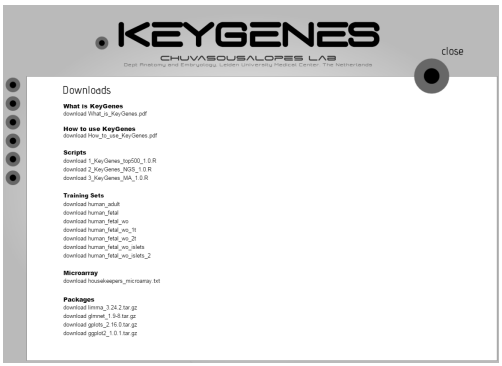
determinants, such as transcription factors, to differentiate PSCs or directly convert somatic cells to the required cell type [49–52]**; (2) there is significant interest in bioinformatic tools capable of evaluating fidelity in the identity of reprogrammed or differentiated cells. PluriTest was among the first of these tools; it uses genome-wide data to evaluate and quantify the pluripotent state of PSCs** [53]**. The ScoreCard and TeratoScore are both tools that assess the differentiation capacity of PSCs based on the analysis of embryoid body differentiation and teratoma formation, respectively** [54–56]**. CellNet was the first platform that enabled comparison of microarray data of reprogrammed, directed differentiated or directly converted cells to training data of 20 cells and adult tissues in order to quantify the identity of the *in vitro* generated cells** [57]**. We show in chapter 3 that KeyGenes can be used to monitor and quantify directed differentiation of PSCs and that it has three major advantages: first, in contrast to the other tools using qPCR and microarray data, KeyGenes gives accurate predictions with microarray and NGS data. Secondly, using our fetal data and (external) adult data, KeyGenes is capable of assigning developmental stages to stem cell progeny. Thirdly, the flexibility regarding the training set and the open source nature makes KeyGenes a powerful tool for many applications and gives the user the possibility of adapting it to specific research questions.**

With the technological progress and the declining costs of RNA sequencing, the significance of «big data» in stem cell research will continue to increase. Particularly due to single-cell sequencing, new challenges in terms of analysis but also new possibilities will emerge. RaceID was the first algorithm designed for single cells exclusively and is used to identify yet unknown cell types based on the transcriptional identity of a single cell population [58]**. Further development of this algorithm led to a tool called StemID, which is capable of identifying yet unknown stem cells** [59]**. This tool is particularly useful to find evidence for the presence of multipotent, adult stem**

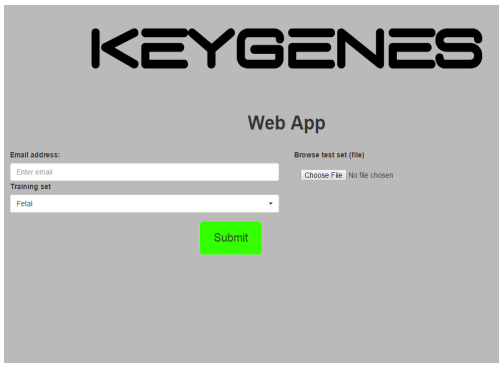
A


KEYGENES
CHUVASOUSALOPES LAB
Dept Anatomy and Embryology, Leiden University Medical Center, The Netherlands
downloads
web app
how to use
faq's
what is
contacts

B


KEYGENES
CHUVASOUSALOPES LAB
close
Downloads
What is KeyGenes
download What_is_KeyGenes.pdf
How to use KeyGenes
download How_to_use_KeyGenes.pdf
Scripts
download 1_KeyGenes_top500_1.0.R
download 2_KeyGenes_NGS_1.0.R
download 3_KeyGenes_MA_1.0.R
Training Sets
download human_adult
download human_fetal
download human_fetal_wo
download human_fetal_wo_TI
download human_fetal_wo_2t
download human_fetal_wo_islets
download human_fetal_wo_islets_2
Microarray
download housekeepers_microarray.txt
Packages
download limma_3.24.2.tar.gz
download gtimer_1.9-8.tar.gz
download gdata_2.16.0.tar.gz
download ggplot2_1.0.1.tar.gz

C


KEYGENES
Web App
Email address:
Enter email
Training set
Fetal
Browse test set (file)
Choose File
No file chosen
Submit

[Fig. 1]

KeyGenes website.

[A] The home page containing the different options of the KeyGenes website. **[B]** All the documents, scripts and data to download from the KeyGenes website. **[C]** The web application of KeyGenes.

cells, and its value has been demonstrated on the adult pancreas. Single-cell analysis may also improve KeyGenes and broaden its application range. Using transcriptional data of single cells from the different organs at different stages of development as training sets would increase the significance of KeyGenes predictions and its value for stem cell research. However, the generation of a single cell atlas of all the functional cells of the organs of our training set would be an enormous task in terms of cell isolation, sequencing and data analysis, and the associated costs would also be significant.

DNA METHYLATION AND GENE EXPRESSION TRAJECTORIES AND ORGAN-SPECIFIC SIGNATURES DURING HUMAN FETAL DEVELOPMENT

Following on from the work in chapter 3, we conducted an in-depth analysis of the DNA methylation profiles of four human tissues (amnion, muscle, adrenal and pancreas) at W9, W18, and W22 of human fetal development in chapter 4. Hypomethylation has been described as organ-discriminating feature in adults and a selection of second trimester human organs [60–62]**. Using three time points and biological replicates, we identified specific sets of hypomethylated CpGs for each organ suggesting organ-specific epigenetic identity as early as W9. In chapter 5, we extended the dataset of genome-wide DNA methylation data to 21 fetal organs. Since inter-individual DNA methylation differences may influence the analysis of the data, one feature of this dataset is particularly important: it comprises one isogenic set of 14 tissues per time point (W9, W18 and W21). Using these isogenic organ sets, we showed that each organ harbors specific sets of hypermethylated and hypomethylated CpGs at every time point, which is intriguing considering the large number of organs that were compared to each other. The observation that there were fewer tissue-specific signatures of hypermethylated CpGs confirmed that hypomethylation is indeed the predominating feature capable**

of distinguishing different tissues as early as the first trimester of human development. Furthermore, taking advantage of the three time points, we show in chapter 4 how dynamic the human DNA methylome during fetal development is. Around 11.5% of the CpGs present on the Illumina 450k array platform underwent either a 20% gain or 20% loss of DNA methylation. Interestingly, loss of methylation was rather tissue-specific and observed near genes involved in the function of the corresponding tissue, whereas gain of methylation was shared by the four investigated tissues and associated with developmental genes and programs. The observation in chapter 5 that the hyper- and hypomethylated signatures at W9 and W22 showed very little overlap per tissue further suggests that the human methylome in all of the 21 tissues undergoes extensive remodeling. The regulation of enhancers and their role in tissue specificity have just begun to be explored [63–65]**. Underlining the biological relevance of the hypomethylated and dynamic CpGs identified in chapter 4, we show that these CpGs were enriched for enhancers. Moreover, we showed that regions with a loss of methylation were associated with open chromatin and active histone modifications, whereas regions that gain methylation were depleted of open chromatin and active histone modifications. This illustrates that the two epigenetic marks DNA methylation and histone modifications cannot be considered separately when investigating the tissue-specific regulation of gene expression. It will be important to take a global chromatin state involving DNA methylation and histone modifications into account at all the regulatory regions in the human genome** [64, 66]**.**

The transcriptional data generated in chapter 3 allowed us to compare how DNA methylation changes are accompanied by up- or downregulation of the associated genes. This comparison is particularly powerful because the two datasets consist of the same organs at the same time points. In chapter 4, we show that the genes associated with embryonic

morphogenesis near CpGs undergoing gain of methylation in the four investigated tissues were downregulated between W9 and W22. On the other hand, functional genes that were upregulated in a tissue correlated with a loss of methylation. In chapter 5, we identified the unique differentially expressed genes first, and investigated the methylation changes of proximal promoters and gene bodies separately. Generally, the DNA methylation changes of proximal promoters and gene bodies were rather small, although there were some tissues including the eye, liver and muscle showing a more pronounced association. It has been shown that DNA methylation regulates gene expression during myoblast differentiation indicating an important role of DNA methylation during development [67]. We primarily investigated whole organs, which may partly explain the rather low correlation between DNA methylation changes and gene expression changes. As the observed DNA methylation and gene expression levels are actually the average of different cells within the organ, it is likely that our observations are actually underestimations. This is further underlined by the strong correlation between DNA methylation and gene expression observed in the skeletal muscle, which is relatively homogenous compared to other tissues. Furthermore, the somewhat small DNA methylation changes observed in chapter 5 may be partly explained by the focus on proximal promoters and gene bodies: many promoters harbor CpG islands, which are devoid of methylation and they rarely become methylated [68]. Gene bodies, on the other hand, are underrepresented on the Illumina 450k array platform, so that changes may not be picked up for this technical reason [69]. Moreover, the enrichment of active and repressive histone modifications at CpGs with a loss and gain of methylation observed in chapter 4 suggests that DNA methylation is not the only regulator of gene expression, but acts in concert with other epigenetic mechanisms.

As mentioned above, it is crucial to be able to characterize the outcome of stem cell differentiations and to com-

pare it to the desired cell type *in vivo*. It has been shown that the DNA methylome of stem cells is subjected to changes during *in vitro* differentiation [61, 70–72]. Therefore, knowledge about the molecular signatures of the different organs and cell types during development and at the mature, adult stage is essential to properly benchmark stem cell derivatives. It has been shown that the DNA methylation trajectories of the *in vitro* differentiation to liver cells differed from the liver development *in vivo* illustrating the necessity of accurately monitoring stem cell differentiation [73]. We observed in chapter 4 that the nearest genes of CpGs with a loss and a gain of DNA methylation were associated with organ-specific and developmental functions, respectively, which has also been reported during *in vitro* differentiations [67, 72, 74, 75]. This illustrates that observations of biological processes during fetal development can be traced back to *in vitro* differentiations and can serve as reference to assess the validity of the outcome. A further example for this is *GFAP*, a gene identified in chapter 5 as specifically differentially upregulated in the fetal brain between W9 and W22 and marked by demethylation of five CpG in its proximal promoter in the same time period. *GFAP* has been described as being involved in brain development and, moreover, it is commonly used to characterize the neural progeny from PSCs [76–78].

Our datasets are a valuable resource, and our analyses proved their significance with regard to developmental biology but also stem cell biology. However, there are limitations that need to be addressed in the future: the platform we used for the investigation of DNA methylation, the Illumina 450k array, is limited to the CpGs present on the array. Although we confirmed some of our findings with external data generated with whole genome bisulfite sequencing (WGBS) in chapter 4, it would have been better to extend our DNA methylation data to all CpGs in the genome and include non-CpG methylation and DNA hydroxymethylation. Although we investigated other epigenetic marks using ChIP-seq data on histone modifi-

cations from external sources including H3K4me1, -me2, -me3, H3K9ac, H3K27ac, H3K9me3, H3K27me3, and H3K36me3, it would be preferable to generate ChIP-seq data from the same material regarding source, time points and preparation in order to create a comprehensive and directly comparable picture of the epigenetic landscape. By using adult data, we show in chapter 4 that the most changes of the DNA methylation landscape occurs between the first and second trimester. However, data from the third trimester and matched adult samples to provide a complete epigenetic and transcriptional atlas would be informative. New technologies with increased sequencing depth and coverage and improved detection of non-coding RNAs will expand our knowledge of the tissue- and cell-specific (epi)genetic makeup. The biggest impact on the generation of genomic and epigenomic data will be the further developments in single-cell sequencing at both RNA and DNA levels [47, 79, 80]**. This will revolutionize our understanding of development by pushing resolution to the single-cell level. The ultimate goal will be to create a collection of both transcriptional and epigenetic data of all different cells of an organ and not only of the whole organ. It has been shown that even endothelial cells from different organs contain an organ-specific transcriptional signature illustrating the importance of characterizing all cells of an organ individually** [81]**. However, this technology still has limitations that need to be resolved: there are technical difficulties when generating and analyzing this kind of data, it is expensive considering the large number of different cells in the human body, the organs need to be gently dissociated and the cells individually sorted. How many yet unknown and rare cell types there are in the body will undoubtedly reveal new insights in tissue function. However, how they are identified and characterized still needs to be addressed.**

EPIGENETIC MEMORY

When iPSCs were first generated, much effort went into investigating their similarities with ESCs [82]**. However, the observation that there were sometimes residual transcription patterns from the reprogrammed cells led to the idea that iPSCs may retain an epigenetic memory** [83, 84]**. Indeed, epigenetic marks, particularly regarding DNA methylation, have been confirmed and in some cases this introduced a differentiation bias** [85–92]**. However, recent studies suggest that the two types of PSCs (ESCs and iPSCs) may indeed be indistinguishable and that it was the genetic background of the different lines that was confounding the results in the previous studies** [93–95]**. Despite the contradictory evidence, it is still controversial and often matter of debate whether an epigenetic memory may be beneficial or should be avoided** [96–98]**. Since pluripotency is a transient state during development, the establishment of ESCs from the inner cell mass of preimplantation embryos is *per se* artificial and instable. Therefore, it may be favorable to get as close as possible to the ESC state, and an epigenetic memory may introduce more instability to the culture. Furthermore, the possibility of an epigenetic memory needs to be kept in mind and may influence the outcome when iPSCs are used in the context of disease modelling. On the other hand, the presence of an epigenetic memory, or more specifically its potential impact on the differentiation capacity, may help to generate desired cell types more efficiently or even functionally more mature. Moreover, research on epigenetic memory may increase knowledge not only about the epigenetic mechanism of reprogramming, but also how epigenetic modifiers may improve the process of cellular reprogramming** [99]**.**

Due to the inconsistent findings described above and the potential impact of an epigenetic memory, there is

still a need for more research. To address this, we took advantage of our unique fetal material and we derived hiPSCs from six different organs from the same individual and characterized two clones per isogenic cell line in chapter 5. Since it has been shown that many parameters including reprogramming methods and culture conditions may have an impact on the epigenome [100, 101]**, we developed a single protocol with the same isolation method and culture conditions for the derivation of the primary cells from the six organs. These cells were then reprogrammed with the same lentiviral method and the iPSCs generated were kept in the same culture conditions. This unique experimental setup allowed us to exclude the majority of confounding parameters like genetic background and culture conditions and to assess the presence of a *bona fide* epigenetic memory. We investigated the DNA methylation landscapes of the six triplets (two iPSC clones and associated organ of origin) with the Illumina 450k array and observed that the different cell hiPSC lines were very much alike and there was no indication of a «global» epigenetic memory. However, when filtering for CpGs that have a DNA methylation difference of > 0.2 or < 0.2 in one triplet compared to the other five, we observed that the two iPSC clones from the brain retained 18 hyper- and 6 hypomethylated CpGs from the fetal brain. Intriguingly, 11 of them are associated with the proximal promoters and gene bodies of four genes that have been described in a neuronal context, suggesting that at least the iPSCs derived from brain harbor a very local but genuine epigenetic memory. Strikingly, the brain-derived iPSC clones generated more GFAP-positive cells when subjected to neural differentiation compared to the skin-derived clones, suggesting a possible functional implication of the epigenetic memory.**

One drawback of our study in chapter 5 is that there was no cell sorting step before reprogramming, i.e. we do not know exactly which specific cell types were reprogrammed. Although the cell populations derived from the six organs

showed different morphologies, it is theoretically possible that organ-specific fibroblasts were reprogrammed. The neuronal context of the majority of hyper- and hypomethylated CpGs that were found in the iPSCs derived from brain suggest that in that case we indeed reprogrammed a neural population of cells. However, regarding the other five organs, we cannot exclude the possibility that only (organ-specific) fibroblasts were reprogrammed. If this was the case, this would be valuable information as well since it has been shown that fibroblasts from different anatomical regions harbour different gene expression signatures [102]**. If we reprogrammed only fibroblasts, our results would suggest that the different positional background of the fibroblasts does not have an impact on the epigenome regarding DNA methylation, since all those clones from the five organs, except the brain, were practically indistinguishable. Another concern is the integrating nature of the lentiviral construct to reprogram: it has been reported recently that integrating reporgramming methods may influence gene expression, the epigenome as well as the differentiation capacity more than non-integrating systems** [103]**. However, the lack of origin-specific DNA methylation marks in 10 of 12 iPSC lines and the small numbers of hyper- and hypomethylated CpGs in the two clones derived from brain, and particularly the neuronal context, argue against a random technical impact on the epigenome. It is important to point out that the lack of an epigenetic memory in the iPSCs lines from five out of six organs could also be a technical underestimation since the Illumina 450k does not cover all the CpGs. Also the conclusion regarding the drawbacks of the study could be affected by this platform-related limitation. Moreover, an epigenetic memory does not have to be necessarily related to DNA methylation; it could be constituted as well at different epigenetic levels such as histone modifications and micro RNAs** [89, 104]**.**

For future studies, it will be important to reprogram a defined starting cell population, which could be achieved

by FACS. Fetal tissue progenitor cells from different organs could be interesting cells to reprogram, since it has been reported that mouse iPSCs derived from hepatoblasts harbour a transcriptional memory, in contrast to the ones derived from adult hepatocytes [105]**. Furthermore, it will be paramount to not only extend the analysis to the DNA methylation patterns of the whole genome, but to include other epigenetic marks including histone modifications and others. Lastly, it remains to be investigated further what the mechanism of the impact of the differentially methylated CpGs found in the brain-derived iPSCs on the neural differentiation capacity exactly is.**

CONCLUDING REMARKS

In this thesis, we have demonstrated that our access to human fetal material derived from elective abortions generated new and valuable insights into aspects of developmental as well as stem cell biology. Investigating (lymph-)angiogenesis in the human fetal pancreas in chapter 2 underlined the importance of the two vascular systems with regard to transplantation and stem cell differentiation. In chapter 3, 4 and 5, we characterized the fetal pancreas, amongst other organs, at different time points in terms of gene expression programs and DNA methylation signatures. This provides a valuable source of information to unravel developmental trajectories of the human pancreas and to benchmark the outcome of β-cell differentiations from PSCs. Furthermore, we generated iPSC lines from the fetal pancreas in chapter 5. However, we did not observe an epigenetic memory retained from the pancreas in these lines, and it remains to be investigated how pancreas-derived hiPSCs behave in a β-cell differentiation protocol.

REFERENCES

[1] Pepper AR, Gala-Lopez B, Ziff O, Shapiro AM: **Revascularization of transplanted pancreatic islets and role of the transplantation site**. *Clinical & developmental immunology* 2013, **2013**:352315.

[2] McKnight KD, Wang P, Kim SK: **Deconstructing pancreas development to reconstruct human islets from pluripotent stem cells**. *Cell Stem Cell* 2010, **6**(4):300–308.

[3] Cleaver O, Melton DA: **Endothelial signaling during development**. *Nature medicine* 2003, **9**(6):661–668.

[4] Lammert E, Cleaver O, Melton D: **Induction of pancreatic differentiation by signals from blood vessels**. *Science* 2001, **294**(5542):564–567.

[5] Yoshitomi H, Zaret KS: **Endothelial cell interactions initiate dorsal pancreas development by selectively inducing the transcription factor Ptf1a**. *Development* 2004, **131**(4):807–817.

[6] Pierreux CE, Cordi S, Hick AC, Achouri Y, Ruiz de Almodovar C, Prevot PP, Courtoy PJ, Carmeliet P, Lemaigre FP: **Epithelial: Endothelial cross-talk regulates exocrine differentiation in developing pancreas**. *Developmental biology* 2010, **347**(1):216–227.

[7] Cleaver O, Dor Y: **Vascular instruction of pancreas development**. *Development* 2012, **139**(16):2833–2843.

[8] Ballian N, Brunicardi FC: **Islet vasculature as a regulator of endocrine pancreas function**. *World journal of surgery* 2007, **31**(4):705–714.

[9] Brissova M, Shostak A, Shiota M, Wiebe PO, Poffenberger G, Kantz J, Chen Z, Carr C, Jerome WG, Chen J *et al*: **Pancreatic islet production of vascular endothelial growth factor--a is essential for islet vascularization, revascularization, and function**. *Diabetes* 2006, **55**(11):2974–2985.

[10] Kaido T, Yebra M, Cirulli V, Montgomery AM: **Regulation of human beta-cell adhesion, motility, and insulin secretion by collagen IV and its receptor alpha1beta1**. *J Biol Chem* 2004, **279**(51):53762–53769.

[11] Nikolova G, Jabs N, Konstantinova I, Domogatskaya A, Tryggvason K, Sorokin L, Fassler R, Gu G, Gerber HP, Ferrara N *et al*: **The vascular basement membrane: a niche for insulin gene expression and Beta cell proliferation**. *Developmental cell* 2006, **10**(3):397–405.

[12] Jurisic G, Detmar M: **Lymphatic endothelium in health and disease**. *Cell and tissue research* 2009, **335**(1):97–108.

[13] Ordonez NG: **Immunohistochemical endothelial markers: a review**. *Advances in anatomic pathology* 2012, **19**(5):281–295.

[14] Cesmebasi A, Malefant J, Patel SD, Du Plessis M, Renna S, Tubbs RS, Loukas M: **The surgical anatomy of the lymphatic system of the pancreas**. *Clinical anatomy* 2015, **28**(4):527–537.

[15] O'Morchoe CC: **Lymphatic system of the pancreas**. *Microscopy research and technique* 1997, **37**(5–6):456–477.

[16] Olsson R, Maxhuni A, Carlsson PO: **Revascularization of transplanted pancreatic islets following culture with stimulators of angiogenesis**. *Transplantation* 2006, **82**(3):340–347.

[17] Merani S, Toso C, Emamaullee J, Shapiro AM: **Optimal implantation site for pancreatic islet transplantation**. *The British journal of surgery* 2008, **95**(12):1449–1461.

[18] Zhang N, Richter A, Suriawinata J, Harbaran S, Altomonte J, Cong L, Zhang H, Song K, Meseck M, Bromberg J *et al*: **Elevated vascular endothelial growth factor production in islets improves islet graft vascularization**. *Diabetes* 2004, **53**(4):963–970.

[19] Stendahl JC, Kaufman DB, Stupp SI: **Extracellular matrix in pancreatic islets: relevance to scaffold design and transplantation**. *Cell transplantation* 2009, **18**(1):1–12.

[20] Tuch BE, Gao SY, Lees JG: **Scaffolds for islets and stem cells differentiated into insulin-secreting cells**. *Frontiers in bioscience* 2014, **19**:126–138.

[21] Frantz C, Stewart KM, Weaver VM: **The extracellular matrix at a glance**. *Journal of cell science* 2010, **123**(Pt 24):4195–4200.

[22] Korpos E, Kadri N, Kappelhoff R, Wegner J, Overall CM, Weber E, Holmberg D, Cardell S, Sorokin L: **The peri-islet basement membrane, a barrier to infiltrating leukocytes in type 1 diabetes in mouse and human**. *Diabetes* 2013, **62**(2):531–542.

[23] Otonkoski T, Banerjee M, Korsgren O, Thornell LE, Virtanen I: **Unique basement membrane structure of human pancreatic islets: implications for beta-cell growth and differentiation**. *Diabetes, obesity & metabolism* 2008, **10 Suppl 4**:119–127.

[24] Takada Y, Ye X, Simon S: **The integrins**. *Genome Biol* 2007, **8**(5):215.

[25] Wang R, Li J, Lyte K, Yashpal NK, Fellows F, Goodyer CG: **Role for beta1 integrin and its associated alpha3, alpha5, and alpha6 subunits in development of the human fetal pancreas**. *Diabetes* 2005, **54**(7):2080–2089.

[26] Wang RN, Paraskevas S, Rosenberg L: **Characterization of integrin expression in islets isolated from hamster, canine, porcine, and human pancreas**. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 1999, **47**(4):499–506.

[27] Saleem S, Li J, Yee SP, Fellows GF, Goodyer CG, Wang R: **beta1 integrin/FAK/ERK signalling pathway is essential for human fetal islet cell differentiation and survival**. *The Journal of pathology* 2009, **219**(2):182–192.

[28] Krishnamurthy M, Li J, Fellows GF, Rosenberg L, Goodyer CG, Wang R: **Integrin {alpha}3, but not {beta}1, regulates islet cell survival and function via PI3K/Akt signaling pathways**. *Endocrinology* 2011, **152**(2):424–435.

[29] Virtanen I, Banerjee M, Palgi J, Korsgren O, Lukinius A, Thornell LE, Kikkawa Y, Sekiguchi K, Hukkanen M, Konttinen YT *et al*: **Blood vessels of human islets of Langerhans are surrounded by a double basement membrane**. *Diabetologia* 2008, **51**(7):1181–1191.

[30] Hrvatin S, O'Donnell CW, Deng F, Millman JR, Pagliuca FW, DiIorio P, Rezania A, Gifford DK, Melton DA: **Differentiated human stem cells resemble fetal, not adult, beta cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**(8):3038–3043.

[31] Kieffer TJ: **Closing in on Mass Production of Mature Human Beta Cells**. *Cell Stem Cell* 2016, **18**(6):699–702.

[32] Talavera-Adame D, Wu G, He Y, Ng TT, Gupta A, Kurtovic S, Hwang JY, Farkas DL, Dafoe DC: **Endothelial cells in co-culture enhance embryonic stem cell differentiation to pancreatic progenitors and**

insulin-producing cells through BMP signaling. *Stem cell reviews* 2011, **7**(3):532–543.

[33] Narayanan K, Lim VY, Shen J, Tan ZW, Rajendran D, Luo SC, Gao S, Wan AC, Ying JY: **Extracellular matrix-mediated differentiation of human embryonic stem cells: differentiation to insulin-secreting beta cells**. *Tissue engineering Part A* 2014, **20**(1–2):424–433.

[34] Clevers H: **Modeling Development and Disease with Organoids**. *Cell* 2016, **165**(7):1586–1597.

[35] Boj SF, Hwang CI, Baker LA, Chio, II, Engle DD, Corbo V, Jager M, Ponz-Sarvise M, Tiriac H, Spector MS *et al*: **Organoid models of human and mouse ductal pancreatic cancer**. *Cell* 2015, **160**(1–2):324–338.

[36] Takebe T, Enomura M, Yoshizawa E, Kimura M, Koike H, Ueno Y, Matsuzaki T, Yamazaki T, Toyohara T, Osafune K *et al*: **Vascularized and Complex Organ Buds from Diverse Tissues via Mesenchymal Cell-Driven Condensation**. *Cell Stem Cell* 2015, **16**(5):556–565.

[37] Norman S, Riley PR: **Anatomy and development of the cardiac lymphatic vasculature: Its role in injury and disease**. *Clinical anatomy* 2016, **29**(3):305–315.

[38] Kallskog O, Kampf C, Andersson A, Carlsson PO, Hansell P, Johansson M, Jansson L: **Lymphatic vessels in pancreatic islets implanted under the renal capsule of rats**. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2006, **6**(4):680–686.

[39] Patterson M, Chan DN, Ha I, Case D, Cui Y, Van Handel B, Mikkola HK, Lowry WE: **Defining the nature of human pluripotent stem cell progeny**. *Cell Res* 2012, **22**(1):178–193.

[40] Jongeneel CV, Delorenzi M, Iseli C, Zhou D, Haudenschild CD, Khrebtukova I, Kuznetsov D, Stevenson BJ, Strausberg RL, Simpson AJ *et al*: **An atlas of human gene expression from massively parallel signature sequencing (MPSS)**. *Genome Res* 2005, **15**(7):1007–1014.

[41] Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G *et al*: **A gene atlas of the mouse and human protein-encoding transcriptomes**. *Proceedings of the*

National Academy of Sciences of the United States of America 2004, **101**(16):6062–6067.

[42] Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K *et al*: **Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics**. *Mol Cell Proteomics* 2014, **13**(2):397–406.

[43] Hwang PI, Wu HB, Wang CD, Lin BL, Chen CT, Yuan S, Wu G, Li KC: **Tissue-specific gene expression templates for accurate molecular characterization of the normal physiological states of multiple human tissues with implication in development and cancer studies**. *BMC Genomics* 2011, **12**:439.

[44] Ju W, Greene CS, Eichinger F, Nair V, Hodgin JB, Bitzer M, Lee YS, Zhu Q, Kehata M, Li M *et al*: **Defining cell-type specificity at the transcriptional level in human disease**. *Genome Res* 2013, **23**(11):1862–1873.

[45] Lal A, Lash AE, Altschul SF, Velculescu V, Zhang L, McLendon RE, Marra MA, Prange C, Morin PJ, Polyak K *et al*: **A public database for gene expression in human cancers**. *Cancer Res* 1999, **59**(21):5403–5407.

[46] Takasato M, Er PX, Chiu HS, Maier B, Baillie GJ, Ferguson C, Parton RG, Wolvetang EJ, Roost MS, Chuva de Sousa Lopes SM *et al*: **Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis**. *Nature* 2015, **526**(7574):564–568.

[47] Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA: **The technology and biology of single-cell RNA sequencing**. *Mol Cell* 2015, **58**(4):610–620.

[48] Bian Q, Cahan P: **Computational Tools for Stem Cell Biology**. *Trends in biotechnology* 2016.

[49] Rackham OJ, Firas J, Fang H, Oates ME, Holmes ML, Knaupp AS, The FANTOM Consortium, Suzuki H, Nefzger CM, Daub CO *et al*: **A predictive computational framework for direct reprogramming between human cell types**. *Nat Genet* 2016, **48**(3):331–335.

[50] D'Alessio AC, Fan ZP, Wert KJ, Baranov P, Cohen MA, Saini JS, Cohick E, Charniga C, Dadon D, Hannett NM *et al*: **A Systematic**

Approach to Identify Candidate Transcription Factors that Control Cell Identity. *Stem Cell Reports* 2015, **5**(5):763–775.

[51] Heinaniemi M, Nykter M, Kramer R, Wienecke-Baldacchino A, Sinkkonen L, Zhou JX, Kreisberg R, Kauffman SA, Huang S, Shmulevich I: **Gene-pair expression signatures reveal lineage control**. *Nature methods* 2013, **10**(6):577–583.

[52] Lang AH, Li H, Collins JJ, Mehta P: **Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes**. *PLoS computational biology* 2014, **10**(8):e1003734.

[53] Muller FJ, Schuldt BM, Williams R, Mason D, Altun G, Papapetrou EP, Danner S, Goldmann JE, Herbst A, Schmidt NO *et al*: **A bioinformatic assay for pluripotency in human cells**. *Nature methods* 2011, **8**(4):315–317.

[54] Avior Y, Biancotti JC, Benvenisty N: **TeratoScore: Assessing the Differentiation Potential of Human Pluripotent Stem Cells by Quantitative Expression Analysis of Teratomas**. *Stem Cell Reports* 2015, **4**(6):967–974.

[55] Tsankov AM, Akopian V, Pop R, Chetty S, Gifford CA, Daheron L, Tsankova NM, Meissner A: **A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells**. *Nat Biotechnol* 2015.

[56] Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH *et al*: **Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines**. *Cell* 2011, **144**(3):439–452.

[57] Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ: **CellNet: Network Biology Applied to Stem Cell Engineering**. *Cell* 2014, **158**(4):903–915.

[58] Grun D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, van Oudenaarden A: **Single-cell messenger RNA sequencing reveals rare intestinal cell types**. *Nature* 2015, **525**(7568):251–255.

[59] Grun D, Muraro MJ, Boisset JC, Wiebrands K, Lyubimova A, Dharmadhikari G, van den Born M, van Es J, Jansen E, Clevers H *et al*: **De Novo Prediction of Stem Cell Identity using Single-Cell**

Transcriptome Data. *Cell Stem Cell* 2016.

[60] Lokk K, Modhukur V, Rajashekar B, Martens K, Magi R, Kolde R, Koltsina M, Nilsson TK, Vilo J, Salumets A *et al*: **DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns**. *Genome Biol* 2014, **15**(4):r54.

[61] Nazor KL, Altun G, Lynch C, Tran H, Harness JV, Slavin I, Garitaonandia I, Muller FJ, Wang YC, Boscolo FS *et al*: **Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives**. *Cell Stem Cell* 2012, **10**(5):620–634.

[62] Colaneri A, Wang T, Pagadala V, Kittur J, Staffa NG, Jr., Peddada SD, Isganaitis E, Patti ME, Birnbaumer L: **A minimal set of tissue-specific hypomethylated CpGs constitute epigenetic signatures of developmental programming**. *PLoS One* 2013, **8**(9):e72670.

[63] Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ *et al*: **DNA methylation status predicts cell type-specific enhancer activity**. *EMBO J* 2011, **30**(15):3028–3039.

[64] Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al*: **Mapping and analysis of chromatin state dynamics in nine human cell types**. *Nature* 2011, **473**(7345):43–49.

[65] Zhou VW, Goren A, Bernstein BE: **Charting histone modifications and the functional organization of mammalian genomes**. *Nat Rev Genet* 2011, **12**(1):7–18.

[66] Yen A, Kellis M: **Systematic chromatin state comparison of epigenomes associated with diverse properties including sex and tissue type**. *Nature communications* 2015, **6**:7973.

[67] Miyata K, Miyata T, Nakabayashi K, Okamura K, Naito M, Kawai T, Takada S, Kato K, Miyamoto S, Hata K *et al*: **DNA methylation analysis of human myoblasts during in vitro myogenic differentiation: de novo methylation of promoters of muscle-related genes and its involvement in transcriptional down-regulation**. *Hum Mol Genet* 2015, **24**(2):410–423.

[68] Jones PA: **Functions of DNA methylation: islands, start sites, gene**

bodies and beyond. *Nat Rev Genet* 2012, **13**(7):484–492.

[69] Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M: **Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome**. *Epigenetics* 2011, **6**(6):692–702.

[70] Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R *et al*: **Transcriptional and epigenetic dynamics during specification of human embryonic stem cells**. *Cell* 2013, **153**(5):1149–1163.

[71] Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D *et al*: **Epigenomic analysis of multilineage differentiation of human embryonic stem cells**. *Cell* 2013, **153**(5):1134–1148.

[72] Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J *et al*: **Dynamic changes in the human methylome during differentiation**. *Genome Res* 2010, **20**(3):320–331.

[73] Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E *et al*: **Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver**. *Genome Res* 2009, **19**(6):1044–1056.

[74] Nagae G, Isagawa T, Shiraki N, Fujita T, Yamamoto S, Tsutsumi S, Nonaka A, Yoshiba S, Matsusaka K, Midorikawa Y *et al*: **Tissue-specific demethylation in CpG-poor promoters during cellular differentiation**. *Hum Mol Genet* 2011, **20**(14):2710–2721.

[75] Kim M, Park YK, Kang TW, Lee SH, Rhee YH, Park JL, Kim HJ, Lee D, Lee D, Kim SY *et al*: **Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage**. *Hum Mol Genet* 2014, **23**(3):657–667.

[76] Gorris R, Fischer J, Erwes KL, Kesavan J, Peterson DA, Alexander M, Nothen MM, Peitz M, Quandel T, Karus M *et al*: **Pluripotent stem cell-derived radial glia-like cells as stable intermediate for efficient generation of human oligodendrocytes**. *Glia* 2015, **63**(12):2152–2167.

[77] Eng LF, Ghirnikar RS: **GFAP and astrogliosis**. *Brain Pathol* 1994, **4**(3):229–237.

[78] Roessmann U, Velasco ME, Sindely SD, Gambetti P: **Glial fibrillary acidic protein (GFAP) in ependymal cells during development. An immunocytochemical study**. *Brain research* 1980, **200**(1):13–21.

[79] Trapnell C: **Defining cell types and states with single-cell genomics**. *Genome Res* 2015, **25**(10):1491–1498.

[80] Schwartzman O, Tanay A: **Single-cell epigenomics: techniques and emerging applications**. *Nat Rev Genet* 2015, **16**(12):716–726.

[81] Nolan DJ, Ginsberg M, Israely E, Palikuqi B, Poulos MG, James D, Ding BS, Schachterle W, Liu Y, Rosenwaks Z *et al*: **Molecular signatures of tissue-specific microvascular endothelial cell heterogeneity in organ maintenance and regeneration**. *Developmental cell* 2013, **26**(2):204–219.

[82] Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch R *et al*: **Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution**. *Cell Stem Cell* 2007, **1**(1):55–70.

[83] Marchetto MC, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR: **Transcriptional signature and memory retention of human-induced pluripotent stem cells**. *PLoS One* 2009, **4**(9):e7076.

[84] Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC: **Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells**. *PLoS One* 2010, **5**(2):e8975.

[85] Bar-Nur O, Russ HA, Efrat S, Benvenisty N: **Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells**. *Cell Stem Cell* 2011, **9**(1):17–23.

[86] Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ *et al*: **Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells**. *Nat Cell Biol* 2011, **13**(5):541–549.

[87] Kim K, Zhao R, Doi A, Ng K, Unternaehrer J, Cahan P, Huo H, Loh YH, Aryee MJ, Lensch MW *et al*: **Donor cell type can influence the**

epigenome and differentiation potential of human induced pluripotent stem cells. *Nat Biotechnol* 2011, **29**(12):1117–1119.

[88] Ma H, Morey R, O'Neil RC, He Y, Daughtry B, Schultz MD, Hariharan M, Nery JR, Castanon R, Sabatini K *et al*: **Abnormalities in human pluripotent cells due to reprogramming mechanisms**. *Nature* 2014, **511**(7508):177–183.

[89] Vitaloni M, Pulecio J, Bilic J, Kuebler B, Laricchia-Robbio L, Izpisua Belmonte JC: **MicroRNAs contribute to induced pluripotent stem cell somatic donor memory**. *J Biol Chem* 2014, **289**(4):2084–2098.

[90] Hargus G, Ehrlich M, Arauzo-Bravo MJ, Hemmer K, Hallmann AL, Reinhardt P, Kim KP, Adachi K, Santourlidis S, Ghanjati F *et al*: **Origin-dependent neural cell identities in differentiated human iPSCs in vitro and after transplantation into the mouse brain**. *Cell Rep* 2014, **8**(6):1697–1703.

[91] Quattrocelli M, Swinnen M, Giacomazzi G, Camps J, Barthelemy I, Ceccarelli G, Caluwe E, Grosemans H, Thorrez L, Pelizzo G *et al*: **Mesodermal iPSC-derived progenitor cells functionally regenerate cardiac and skeletal muscle**. *The Journal of clinical investigation* 2015, **125**(12):4463–4482.

[92] Hiler D, Chen X, Hazen J, Kupriyanov S, Carroll PA, Qu C, Xu B, Johnson D, Griffiths L, Frase S *et al*: **Quantification of Retinogenesis in 3D Cultures Reveals Epigenetic Memory and Higher Efficiency in iPSCs Derived from Rod Photoreceptors**. *Cell Stem Cell* 2015, **17**(1):101–115.

[93] Choi J, Lee S, Mallard W, Clement K, Tagliazucchi GM, Lim H, Choi IY, Ferrari F, Tsankov AM, Pop R *et al*: **A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs**. *Nat Biotechnol* 2015.

[94] Rouhani F, Kumasaka N, de Brito MC, Bradley A, Vallier L, Gaffney D: **Genetic background drives transcriptional variation in human induced pluripotent stem cells**. *PLoS Genet* 2014, **10**(6):e1004432.

[95] Tiemann U, Wu G, Marthaler AG, Scholer HR, Tapia N: **Epigenetic Aberrations Are Not Specific to Transcription Factor-Mediated Reprogramming**. *Stem Cell Reports* 2016, **6**(1):35–43.

[96] Brix J, Zhou Y, Luo Y: **The Epigenetic Reprogramming Roadmap in**

Generation of iPSCs from Somatic Cells. *Journal of genetics and genomics = Yi chuan xue bao* 2015, **42**(12):661–670.

[97] Firas J, Liu X, Polo JM: **Epigenetic memory in somatic cell nuclear transfer and induced pluripotency: evidence and implications**. *Differentiation* 2014, **88**(1):29–32.

[98] Sebban S, Buganim Y: **Nuclear Reprogramming by Defined Factors: Quantity Versus Quality**. *Trends in cell biology* 2016, **26**(1):65–75.

[99] Pennarossa G, Zenobi A, Gandolfi CE, Manzoni EF, Gandolfi F, Brevini TA: **Erase and Rewind: Epigenetic Conversion of Cell Fate**. *Stem cell reviews* 2016, **12**(2):163–170.

[100] Chin MH, Pellegrini M, Plath K, Lowry WE: **Molecular analyses of human induced pluripotent stem cells and embryonic stem cells**. *Cell Stem Cell* 2010, **7**(2):263–269.

[101] Nishino K, Toyoda M, Yamazaki-Inoue M, Fukawatase Y, Chikazawa E, Sakaguchi H, Akutsu H, Umezawa A: **DNA methylation dynamics in human induced pluripotent stem cells over time**. *PLoS Genet* 2011, **7**(5):e1002085.

[102] Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY: **Anatomic demarcation by positional variation in fibroblast gene expression programs**. *PLoS Genet* 2006, **2**(7):e119.

[103] Koyanagi-Aoi M, Ohnuki M, Takahashi K, Okita K, Noma H, Sawamura Y, Teramoto I, Narita M, Sato Y, Ichisaka T *et al*: **Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(51):20569–20574.

[104] Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S *et al*: **Distinct epigenomic landscapes of pluripotent and lineage-committed human cells**. *Cell Stem Cell* 2010, **6**(5):479–491.

[105] Lee SB, Seo D, Choi D, Park KY, Holczbauer A, Marquardt JU, Conner EA, Factor VM, Thorgeirsson SS: **Contribution of hepatic lineage stage-specific donor memory to the differential potential of induced mouse pluripotent stem cells**. *Stem Cells* 2012, **30**(5):997–1007.

APPENDIX

SUMMARY

Protocols to differentiate human pluripotent stem cells towards any cell type of interest mostly generate immature and only partially functional cells. In order to establish benchmarks for the outcome of stem cell differentiation protocols, detailed understanding of human organ and cell development *in vivo* is essential. For ethical and logistical reasons, there is a shortage of good quality human fetal tissue; our access to rare first and second trimester human fetuses derived from elective abortions provided the tissue that formed the basis of this thesis.

In chapter 2, the pancreas was used as a paradigm to study human organ development and assess the quality of our fetal material. In a descriptive, histochemical study, we investigated how blood and lymphatic vascular networks develop and their association with basement membranes and smooth muscle cells between gestational weeks 9 and 22 (W9 and W22). Our results suggest that both vascular networks are established in the human fetal pancreas by W22. However, the different timing of colonization of the pancreas indicates that both angiogenesis and lymhangiogenesis, have their own temporal and spatial kinetics of progression.

In Chapter 3, 4 and 5, we analyzed a total of 21 fetal organs and maternal endometrium at three time points (W9, W18 and W22) at the transcriptional and epigenetic level, thereby, providing an atlas of human organ development. In chapter 3, we showed that the expression profiles of only 90 genes are sufficient to determine the molecular identity of the investigated organs. Furthermore, we developed a tool, KeyGenes, that is able to assign identity and developmental stage to queried transcriptional profiles based on our fetal atlas. We showed that KeyGenes can be used to determine the identity of stem cell derivatives and, therefore, is a valuable tool to monitor dif-

ferentiation of many test lineages. Moreover, we observed that differentiated cells often match with their first trimester counterpart illustrating the caveat regarding the immaturity of pluripotent stem cell derivatives. In chapter 4, we analyzed the DNA methylation profiles of four human fetal organs (pancreas, amnion, muscle, adrenal gland). Intriguingly, we were able to distinguish the four organs based on DNA hypomethylation as early as W9. However, we also showed that the DNA methylation landscapes of the four organs were subjected to heavy remodeling between W9 and W22. Interestingly, these changes of DNA methylation occurred preferentially in enhancers, underlining the growing importance of this class or regulatory sequences in the field. Furthermore, the observed gain and loss of DNA methylation correlated well with gene expression and other epigenetic marks such as histone modifications. In chapter 5, we expanded the analysis to DNA methylation signatures of all 21 organs. We demonstrated that organ-specific DNA hyper- and hypomethylation were still sufficient to distinguish the tissues, which is even more meaningful due to the isogenicity of our dataset. Using the transcriptional data from chapter 3, we showed that differentially expressed genes between W9 and W22 correlated with changes of DNA methylation in their gene bodies and proximal promoters.

This rare fetal material also allowed us to investigate the presence of an epigenetic memory from the cell of origin in induced pluripotent stem cells (iPSCs). Since genetic background is one of the major confounder, the generation of isogenic iPSC lines from six fetal organs (brain, skin, kidney, muscle, lung and pancreas) in chapter 5 circumvented this problem. The six iPSC lines had very similar DNA methylation profiles, however, we showed that the two clones derived from the brain harbored 18 hypermethylated and 6 hypomethylated CpGs also found in the fetal brain. Intriguingly, the majority of these CpGs were in the gene bodies and proximal promoters of neuronal genes. Further underlining the *bona fide* nature

of this epigenetic memory, we demonstrated that the brain-iPSC clones appeared to have a differentiation bias towards neural derivatives when comparing neural differentiation of the brain- and skin-iPSC clones.

Together, the work in this thesis showed that our access to human fetal material derived from elective abortions provides novel insights into aspects of developmental as well as stem cell biology. We have demonstrated the importance of detailed profiles of human fetal tissues in understanding the identity of pluripotent stem cell derivatives and which genes are regulated with gestational age that could be used to monitor maturation of pluripotent stem cell derivatives.

NEDERLANDSE SAMENVATTING

Protocollen om humane pluripotente stamcellen te differentiëren naar cel typen van interesse leveren met name immature en slechts gedeeltelijk functionele cellen op. Om referentiepunten vast te stellen voor de uitkomst van stamcel differentiatie protocollen is het van belang om gedetailleerd inzicht te krijgen in de menselijke orgaan- en cel ontwikkeling *in vivo*. Om ethische en logistieke redenen, is er een tekort aan goede kwaliteit humaan foetaal weefsel. Weefsels van eerste- en tweede trimester humane foetussen afkomstig van vrijwillige abortussen uit onze zeldzame databank vormen de basis van dit proefschrift.

In Hoofdstuk 2 gebruikten we de alvleesklier als een model voor humane orgaanontwikkeling en voor het bepalen van de kwaliteit van ons foetaal materiaal. In een beschrijvende, histochemische studie onderzochten we hoe bloed- en lymfatische vasculaire netwerken ontwikkelen en hoe deze samenhangen met basale membranen en gladde spiercellen tussen 9 en 22 weken van de zwangerschap (W9 en W22). Onze resultaten suggereren dat rond W22 beide vasculaire netwerken zijn aangelegd in de foetale alvleesklier. Toch laat het verschil in kolonisatie van de alvleesklier zien dat angiogenese en lymfangiogenese hun eigen temporele en ruimtelijke kinetiek van progressie hebben.

In Hoofdstuk 3, 4 en 5, analyseerden we genexpressie en epigenetische markeringen in totaal 21 foetale organen en het maternale endometrium op drie tijdstippen (W9, W18 en W22), die daarmee samen een humane organogenese atlas vormen. In Hoofdstuk 3 toonden we aan dat een expressie profiel van slechts 90 genen voldoende is om de moleculaire identiteit van onderzochte organen vast te stellen. Daarnaast ontwikkelden we een instrument, KeyGenes, dat in staat is om

zowel de identiteit als het ontwikkelingsstadium vast te stellen door het genexpressie profiel te vergelijken met onze foetale atlas. We lieten zien dat KeyGenes gebruikt kan worden om de identiteit van afgeleiden van stamcellen vast te stellen en is daarmee een waardevol instrument om de differentiatie van cellijnen te monitoren.

We zagen bovendien dat gedifferentieerde cellen vaak overeenkwamen met het overeenkomstige weefsel uit het eerste trimester en dit illustreert hoe immatuur afgeleiden van pluripotente stamcellen zijn.

In Hoofdstuk 4 analyseerden we de DNA-methylatie profielen van vier menselijke foetale organen (pancreas, amnion, spieren, bijnieren). Al in W9 konden we de vier organen onderscheiden op basis van hun DNA methylatie profiel. Daarnaast bleken de vier organen tussen W9 en W22 grote veranderingen te ondergaan in hun DNA methylatie profiel. Interessant is dat deze DNA methylatie veranderingen met name plaatsvonden op *enhancers* en dit laat zien hoe relevant deze regulerende sequenties zijn. De waargenomen toename en verlies van DNA methylatie associeerden met genexpressie en andere epigenetische markeringen zoals histon modificaties.

In Hoofdstuk 5 breidden we de analyse uit door te kijken naar de DNA methylatie profielen van alle 21 organen. Hier lieten we zien dat de orgaan-specifieke hypo- en hypermethylatie nog steeds onderscheidend was tussen weefsels, wat des te bijzonderder is gezien de isogene herkomst van onze weefsels. Gebruikmakend van de genexpressie data uit Hoofdstuk 3, lieten we zien dat differentieel geëxpresseerde genen tussen W9 en W22 correleerden met de DNA methylatie veranderingen in *gene bodies* en *proximale promoters*.

Het zeldzame foetale materiaal maakte het ook mogelijk om de aanwezigheid van een epigenetisch geheugen uit de cel van herkomst in geïnduceerde pluripotente stamcellen (iPSCs) te onderzoeken. Genetische achtergrond is hierbij een van de belangrijkste confounders, maar dit hebben we omzeild

in Hoofdstuk 5 door isogene iPSC lijnen van zes foetale organen (hersenen, huid, nier, spier, long en pancreas) te generen. De zes iPSC cellijnen hadden sterk vergelijkbare DNA methylatie profielen, maar we lieten zien dat de twee klonen afgeleid van de hersenen, 18 gehypermethyleerde en 6 gehypomethyleerde CpGs hadden die ook gevonden werden in foetale hersenen. Intrigerend is dat de meerderheid van deze CpGs in *gene bodies* en *proximale promoters* van neuronale genen lagen. Wat verder het *bona fide* karakter van het epigenetische geheugen onderschrijft, is dat de hersenen-iPSC klonen een voorkeur hadden om te differentiëren in neuronale afgeleiden in vergelijking tot huid-iPSC klonen.

Samengevat laat het werk in dit proefschrift zien dat onze toegang tot humaan foetaal materiaal van vrijwillige abortussen nieuwe inzichten oplevert in zowel ontwikkelings- als stamcel biologie. We hebben het belang laten zien van gedetailleerde profielen van foetale weefsels bij het begrijpen van de identiteit van afgeleiden van pluripotente stamcellen. Ook hebben we laten zien welke genen gereguleerd worden tijdens elke periode van de zwangerschap, welke gebruikt kunnen worden om maturatie te monitoren van afgeleiden van pluripotente stamcellen.

LIST OF PUBLICATIONS

Roost M.S., van Iperen L., de Melo Bernardo A., Mummery C.L., Carlotti F., de Koning E.J.P., Chuva de Sousa Lopes S.M. (2014). Lymphangiogenesis and Angiogenesis during Human Fetal Pancreas Development. *Vascular Cell*. **6:22**.

Heeren A.M., van Iperen L., Klootwijk D.B., de Melo Bernardo A., **Roost M.S.**, Gomes Fernandes M.M., Louwe L.A., Hilders C.G., Helmerhorst F.M., van der Westerlaken L.A., Chuva de Sousa Lopes S.M. (2015). Development of the Follicular Basement Membrane during Human Gametogenesis and Early Folliculogenesis. *BMC Developmental Biology*. **15:4**.

Duggal G., Warrier S., Ghimire S., Broekaert D., van der Jeught M., Lierman S., Deroo T., Peelman L., van Soom A., Cornelissen R., Menten B., Mestdagh P, Vandesompele J., **Roost M.S.**, Slieker R.C., Heijmans B.T., Deforce D., de Sutter P., Chuva de Sousa Lopes S., Heindryckx B. (2015). Alternative Routes to Induce Naïve Pluripotency in Human Embryonic Stem Cells. *Stem Cells*. **33(9):**2686–98.

Roost M.S., van Iperen L., Ariyurek Y., Buermans H.P., Arindrarto W., Devalla H.D., Passier R., Mummery C.L., Carlotti F., de Koning E.J.P., van Zwet E.W., Goeman J.J., Chuva de Sousa Lopes S.M. (2015). KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. *Stem Cell Reports*. **4(6):**1112–24.

Takasato M., Er P.X., Chiu H., Maier B., Baillie G.J., Ferguson C., Parton R.G., Wolvetang E., **Roost M.S.**, Chuva de Sousa Lopes S.M., Little M.H. (2015). Kidney Organoids from Human iPS Cells Contain Multiple Cell Types and Model Human Nephrogenesis. *Nature*. **526(7574):**564–8.

Slieker R.C.*, **Roost M.S.***, van Iperen L., Suchiman H.E.D., Tobi E.W., Carlotti F., de Koning E.J.P., Slagboom E.P., Heijmans B.T., Chuva de Sousa Lopes S.M. (2015). DNA Methylation Landscapes of Human Fetal Development. *PLoS Genetics*. **11(10):**e1005583.
* Shared first authorship

Gomes Fernandes M., Dries R., **Roost M.S.**, Semrau S., de Melo Bernardo A., Davis R.P., Ramakrishnan R., Szuhai K., Maas E., Umans L., Abon Escalona V., Salvatori D., Deforce D., van Criekinge W., Huylebroeck D., Mummery C., Zwijsen A., Chuva de Sousa Lopes S.M. (2016). BMP-SMAD Signaling Regulates Lineage Priming, but Is Dispensable for Self-Renewal in Mouse Embryonic Stem Cells. *Stem Cell Reports*. **6(1):**85–94.

Heeren A.M., He N., de Souza A.F., Goercharn-Ramlal A., van Iperen L., Roost M.S., Gomes Fernandes M.M., van der Westerlaken L.A.J., Chuva de Sousa Lopes S.M. (2016). On the Development of Extra-gonadal and Gonadal Human Fetal Female Germ Cells. *Biology Open*. **5(2):**185–94.

Roost M.S., Slieker R.C., Bialecka M., van Iperen L., Gomes Fernandes M.M., He N., Suchiman H.E.D., Szuhai K., Carlotti F., de Koning E.J.P., Mummery C.L., Heijmans B.T., Chuva de Sousa Lopes S.M. (2017). DNA methylation and transcriptional trajectories during human development and reprogramming of isogenic induced pluripotent stem cells. *Submitted*.

Vertesy A., Arindrarto W., **Roost M.S.**, Reinius B., Ariyurek Y., Kuijk E., Mei L., Sandberg R., van Oudenaarden A., Chuva de Sousa Lopes S.M. (2017). Single-Cell Transcriptional Staging and Parent-of-Origin Allele-Specific Expression Reveal Incomplete Epigenetic Reprogramming in Gonadal Human Female Germ Cells. *Submitted*.

Ghimire S., van der Jeught M., Neupane J., **Roost M.S.**, Anckaert J., Popovic M., van Nieuwerburgh F., Mestdagh P., Vandesompele J., De-

force D., Menten B., Chuva de Sousa Lopes S.M., de Sutter P., Heindryckx B. (2017). Comparative Analysis of the Transcriptional and Functional Profile of Naïve, Primed and Ground State Pluripotency in Mouse Embryonic Stem Cells. *Submitted*.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my two promoters: Professor Christine L. Mummery and Professor Eelco J.P. de Koning. Christine, I am very grateful that you provided the opportunity to complete my PhD thesis in your Department. Your impressive working morale and interpersonal skills have been always exemplary and inspirational for me. Also I appreciated your scientific understanding and critical perspective, which improved my research tremendously. Eelco, although you have been very busy with the patients as a clinician, I enjoyed your genuine curiosity for basic research: Your guidance and advice was indispensable for my thesis.

My sincere thanks also go to my co-promoter Susana M. Chuva de Sousa Lopes in whose group I conducted my research during the four years: Blood, sweat and tears, but a big success after all. Susana, your support, knowledge and experience was the base for that sense of «making things happen» during my PhD project. Your attitude that everything is possible was truly inspirational and is something I will take with me.
I would also like to thank all the (former) members of the group, particularly Maria, Heiko, Nannan, Monika, Ana and Ram.
Your support made it possible to get through all the ups and downs during a thesis. I would especially like to thank Liesbeth van Iperen: Her unique dissection skills provided the groundwork for our research.

I would like to thank my fellow labmates at the Department of Anatomy and Embryology. In particular, the help of Dorien, Christian and Simone regarding cell culture and reprogramming was indispensable. Thank you Kostas, Marcello, Harsha, Sabine, Milena, Cristina, George, Daniel, Atie, Robert, Richard, Catarina, Cheryl, Tim, and everyone else I forgot, for the fruitful discussions and support on various aspects.

I greatly appreciate the input and support of the Island lab at the Department of Nephrology, especially Françoise, Annemieke, Siebe, Rianne, Jeetindra and Maike.

My sincere thanks also go to all the people within the LUMC, whose expertise made this thesis possible: Roderick, Bas and Eka from the Molecular Epidemiology Section; Yavuz, Henk and Emil from the Leiden Genome Technology Center; Jelle and Erik from the Department of Medical Statistics and Bioinformatics; Leon and Wibowo from the Sequencing Analysis Support Core; and Harald, Manuel, Joop and Annelies from the Department of Molecular Cell Biology.

I also would like to thank the people from outside the LUMC who contributed to my thesis: Professor Arnoud Sonnenberg (Netherlands Cancer Institute) and Professor Sólveig Thorsteinsdóttir (University of Lisbon, Portugal), both members of my supervisory committee; Professor Björn Heindryckx and his group (University of Gent, Belgium); Professor Neil. A. Hanley (University of Manchester, United Kingdom), and Professor Melissa Little and Dr. Minoru Takasato (Murdoch Childrens Research Institute, Australia).

Last but not least, I would like to express my gratitude to my family and friends, whose diverse contributions were indispensable for this thesis. In particular, I would like to thank my parents for their support throughout my life. At the end, I owe my deepest gratitude to my girlfriend Anna who shared all the ups and downs during my thesis: your support and unconditional love was the fuel that kept me going. Merci viu mau!

CURRICULUM VITAE

Matthias Shona Roost was born on the 24th of February 1983 in Bern, Switzerland. He completed his matriculation exam at the Freies Gymnasium Bern in 2002. He started his studies in Biochemistry and Molecular Biology at the University of Bern in 2003. In the course of his Bachelor thesis, he conducted a research project addressing the inactivation and structural integrity of the B19 virus under the supervision of Prof. Dr. C. Kempf at the Department of Chemistry and Biochemistry. In 2008, he started a Master's program in Biology at the University of Geneva, Switzerland, where he was introduced to stem cells for the first time. He worked in the laboratory of Dr. M. Jaconi investigating the impact of microRNAs on the differentiation of human embryonic stem cells and obtained his Master's degree in 2010. In the same year, he became a project manager at CELLnTEC Advanced Cell Systems in Bern, Switzerland, assigned with the tasks of setting up a stem cell laboratory and managing a project to improve culture conditions of human embryonic stem cells. In 2012, he started to work as a PhD candidate at the Department of Anatomy and Embryology of the LUMC under the supervision of Prof. dr. C.L. Mummery, Prof. dr. E.J.P. de Koning and Dr. S. Chuva de Sousa Lopes; the results of this research are presented in this thesis.