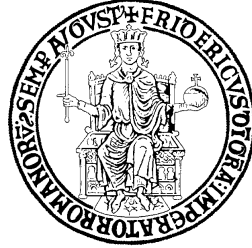# UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II

Dipartimento di Biologia

Ph.D. in Biologia

Insight into Nephrocan Function in Mouse Endoderm Patterning

Tutor
Ch.mo Prof.

Geppino Falco

Candidata

Martina Addeo

Alla mia famiglia…

Index

1.Abstract

Endoderm-derived organs as the liver and pancreas are potential targets for regenerative therapies, and so there is great interest in understanding the pathways that regulate the induction and specification of this germ layer. Currently, the knowledge of molecular mechanisms that guide the *in vivo* endoderm specification is restricted by the lack of early endoderm-specific markers. *Nephrocan* (*Nepn*) is a gene whose expression characterizes the early stages of murine endoderm specification (E7.5-11.5) and encodes a secreted N-glycosylated protein. In the present study, we report the identification of a new transcript variant which is generated through alternative splicing. The new variant was found to have differential and tissue specific expression in the adult mouse. In order to better understand *Nepn* role during endoderm specification, we generated *Nepn* knock-out (KO) mice. *Nepn-/-* mice were born at Mendelian ratios and displayed no evident phenotype compared to WT mice. In addition, we produced a nullizygous mouse embryonic stem cell (mESC) line lacking *Nepn* by applying (CRISPR)/CRISPR-associated systems 9 (Cas9) and employed a differentiation protocol toward endoderm lineage. Our *in vitro* results revealed that *Nepn* loss affects the endoderm differentiation impairing the expression of posterior foregut-associated markers.

2. Introduction

2.1 Developmental principles in stem cells therapies

The ultimate goal of understanding pancreas development is to gain insight into pancreas-associated diseases and to develop novel therapies to treat such diseases. To date, organ transplantation-based therapies are currently limited by the availability of donor-derived tissues. Moreover, to prevent the body from rejecting the transplanted organ, patients must be treated with drugs that suppress the immune system for their entire lives, a regimen that makes them susceptible to a host of other diseases[1]. Embryonic Stem Cells (ESCs) are a promising, renewable source of material for transplantation since they can be expanded indefinitely[2-4] in culture and can differentiate into all cell types of the body, but they are unable to be used *per se* for *in vivo* therapies due to their tumorigenic potential[5]. This limitation could be overcome by optimizing an *in vitro* differentiation protocol to provide an unlimited source of safe progenitor cells, that retain the capability to proliferate and differentiate following the physiological developmental ontogeny. Hence, in the last decade, a great interest has raised the possibility to reprogram a somatic cell into a pluripotent one (iPSCs) resembling the ESCs[6]. This opportunity offers the possible scenario to reprogram a patient's somatic cells, expand them, and then differentiate them following the normal development, into any tissue needed bypassing the immune system and the rejection problems. Although the concept is achievable, there are still several challenges to overcome, for a wide application of cell-based therapies. The idea of expanding

progenitors or adult stem cells *in vitro* has been pursued for different organs with varying degrees of success: in a step-wise differentiation protocol, like the one used in the pancreatic specification, the efficiency of conversion between stages decreases with the progression of pancreatic differentiation[7]. Another open question related to the cell-replacement therapies is the safety of PSC-derived populations and their expanded *in vitro* cell products. Undifferentiated PSCs are tumorigenic and therefore must be removed from their derivative populations before transplantation into patients. A possible way to "purify" an *in vitro* differentiated cell population is to sort the cells through a specific membrane marker, removing the undifferentiated markers expressing cells and retaining the lineage-specific markers expressing cells.

2.2. A Journey From Zygote to Pancreas Budding

The molecular mechanisms controlling the early cell fate decisions in mammals are of major interest to regenerative medicine since guided differentiation of embryonic stem cells (ESCs) toward a specific lineage is an appealing opportunity to provide bases for exploring new therapies. In the early embryo formation, the zygote undergoes a series of rapid sequential mitoses associated with the absence of cell growth: this process is called cleavage. At the stage of 8 cells, this structure is called morula, and the blastomeres face the first cell fate decision. The outer cells become gradually bound together with the formation of tight junction becoming nearly indistinguishable in a process known as compaction, while inner cells form gap junction allowing ions and small molecules sharing. Later the outer cells begin to produce a fluid known as blastocoele given rise to an empty structure, the blastocyst, composed of an outer cell layer called trophoblast, and an inner cell mass (ICM)[8](Figure1a).

The cells of the ICM are defined as pluripotent due to their proficiency to differentiate in all the cells of an adult organism, but not each cell of the ICM will become part of the embryo: the outer cells of the ICM will form the hypoblast (primitive endoderm) while the inner cells will contribute to the epiblast formation (primitive ectoderm), which will give rise to the three primary germ layer[9]. Only the epiblast cells directly in contact with the hypoblast, will give rise to the embryo and this bilayered structure is called germinal disk[10]. The first sign of gastrulation, the process which converts the single-layered blastula into a trilaminar structure known as gastrula, is the formation of the primitive streak (PS) at day E5.5. The epiblast cells near the primitive streak start to proliferate, lose cell-cell adhesion and undergo an epithelial-to-mesenchymal transition (EMT).

4

This "first wave" of cell migration invades the hypoblast and replaces the primitive endoderm cells with the definitive endoderm (DE) once. Later, a "second wave" of migration occurs and the cells take place between the endoderm layer and the ectoderm layer given rise to the third layer: the mesoderm[11] (Figure1b). The mesoderm and the endoderm induction are controlled by an evolutionary conserved gene regulatory network which consists of Nodal growth factor signaling and a core group of downstream transcription factors[12]. The ectoderm layer differentiates to form the nervous system, the epidermis and the cutaneous annexes, whereas the mesoderm layer the skeleton, muscles, kidneys, gonads and the circulatory system, and finally the endoderm layer contributes to the respiratory and gastrointestinal tracts and all of their associated organs such as thyroid, thymus, lungs, liver, biliary system, and pancreas[13].

After gastrulation, a series of morphogenetic movements transform the naïve endoderm into a primitive gut tube that is surrounded by mesoderm. During this period, the gut tube becomes regionalized along the dorsal-ventral (D-V) and anterior-posterior (A-P) axes into broad foregut, midgut, and hindgut domains, characterized by a restricted gene expression patterns. The gut-tube morphogenesis in amniotes begins with the invagination, and rostral-to-caudal movement, of the foregut and hindgut pockets, whereas the intermediate region, the midgut, remains transiently open. In mice, this process begins at ~E8.0, and it is largely complete at E9.5 when the anterior and posterior intestinal portals join at the midgut and organ primordia begin to bud from the gut tube[14]. Organ buds develop as outgrowths of endoderm epithelium that intermingle with the surrounding mesenchyme, and together these proliferate and ultimately differentiate during fetal development into functional organs. The foregut

gives rise to the esophagus, trachea, stomach, lungs, thyroid, liver, biliary system, and pancreas; whereas the midgut forms the small intestine and the hindgut forms the large intestine[14]. The pancreatic development becomes morphologically evident around the embryonic stage 9.0 (E9.0), with the emergence of the epithelial buds on opposing sides of the foregut endoderm. The two pancreatic buds, as a result of gut rotation, eventually fuse into a single organ by E12.5, followed by a cell proliferation which leads to a rapid size increase of the pancreatic buds[15]. This process has been referred to as "primary transition" occurs between E9.5-E12.5 and it is marked by a dramatic morphogenetic change of the pancreatic epithelium in which the main events in this transition include an active proliferation of pancreatic progenitors to generate a stratified epithelium[16]. Starting at E13.5, the epithelium undergoes a striking morphogenetic transformation called the "secondary transition," characterized by a massive differentiation wave and lineage allocation towards the three main pancreatic lineages[16] (Figure2).

Figure1
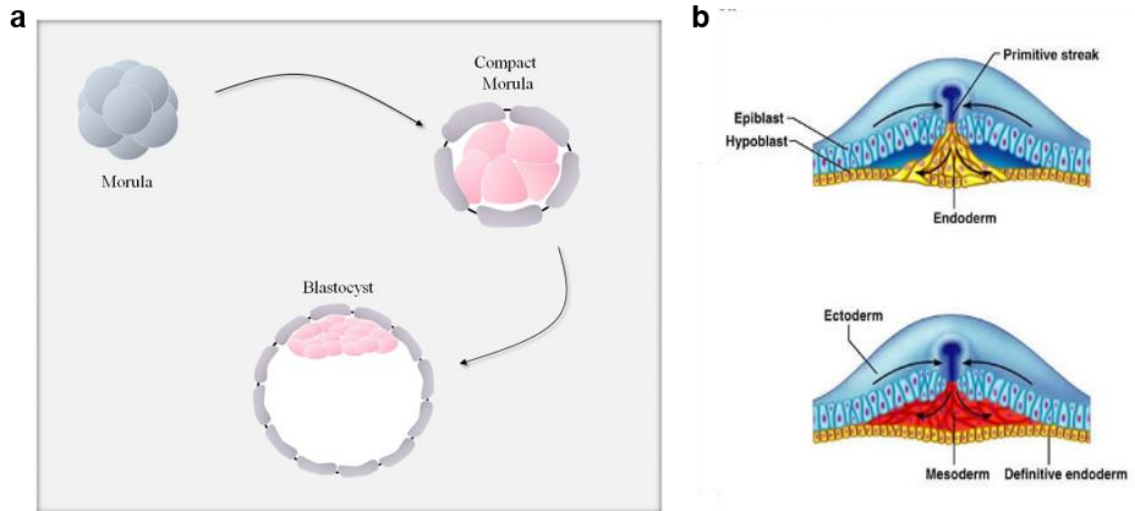
Figure 1. (a) The morula is produced by a series of cleavage divisions of the early embryo, and the blastomeres have no specialization. Later the outer cells become bound tightly together in a process known as compaction. The compact morula develops in the blastocyst, a fluid-filled structure composed of an outer layer of cells, the trophoblast, and an inner cell mass (ICM). **(b).** The primitive streak determines the site of gastrulation and initiates the germ layer formation. The first wave of cell migration gives rise to the definitive endoderm, the second wave the mesoderm, and the remaining epiblast cells will form the ectoderm.

Figure2
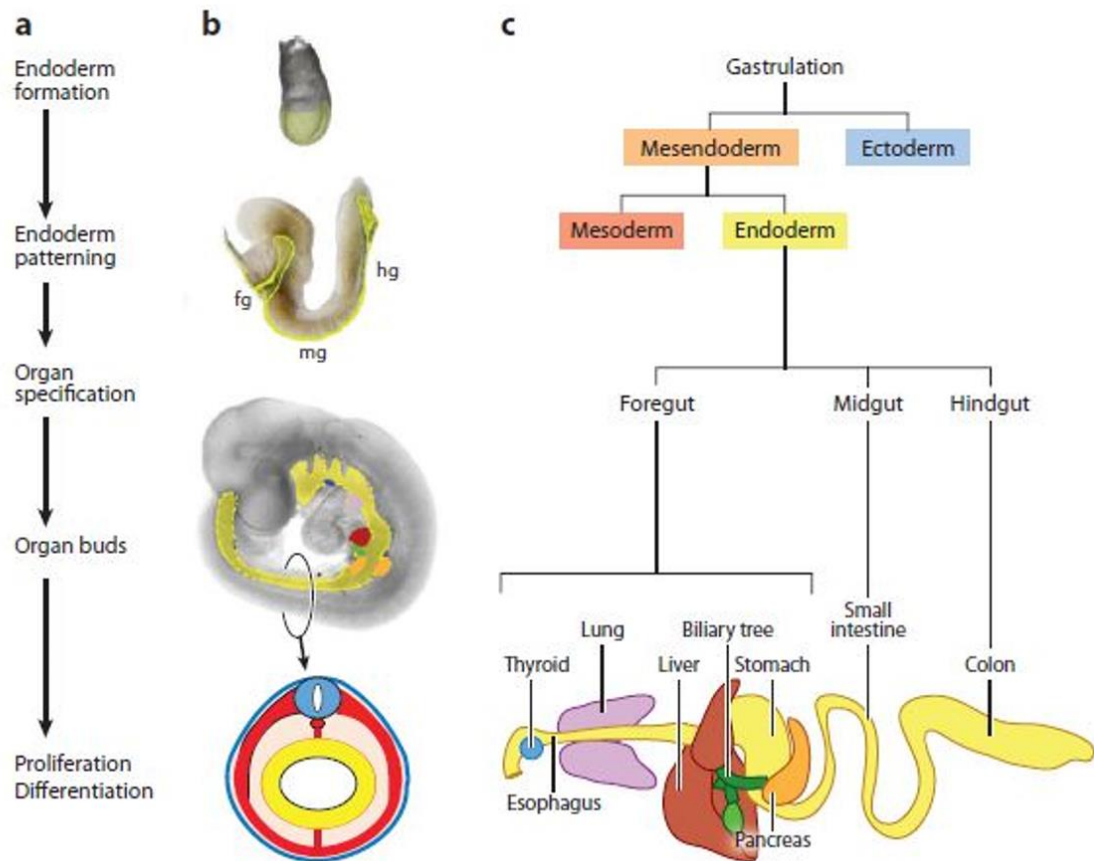
Figure 2. Overview and timeline of endoderm organ formation. (a) The major events in endoderm organ formation are listed in chronological order and **(b)** illustrated with images of mouse embryos at e7.5 (top), e8.5, and e9.5 of development, with the endoderm shaded in (yellow). A schematic of a cross-section through the e9.5 embryo illustrates the characteristic arrangement of the germ layers with the endoderm lining the gut tube (yellow), surrounded by mesoderm (red), and ectoderm (blue). **(c)** Endoderm cell lineages projected onto a schematic of the gastrointestinal tract. Fg, foregut; mg, midgut; hg, hindgut. Adapted from Annu. Rev. Cell Dev. Biol. 2009. 25:221–51.

2.3. Molecular Pathways Governing Endoderm Patterning and Pancreatic Development

During organ formation, cell identity and tissue morphogenesis must be tightly coordinated. These processes are controlled by many growth factor pathways including Nodal, FGF, BMP, Wnt, Retinoic Acid (RA), Hedgehog, and Notch, which play multiple stage-specific roles during endoderm organogenesis.

In all vertebrates, the Nodal signaling pathway is necessary and sufficient to initiate endoderm and mesoderm development. Nodal ligands signal *via* a complex of type I (Alk4 or Alk7) and type II (ActRIIA or ActRIIB) transmembrane serine-threonine kinase receptors; and an EGF-CFC family coreceptor (Cripto). The activated Alk4/7 phosphorylates the cytosolic proteins Smad2 or Smad3. Phosphorylated Smad2/3 then binds to Smad4, and translocates to the nucleus, where it associates with DNA-binding transcription factors such as the Forkhead box h1 (Foxh1/FAST1) or Mix-like homeodomain proteins to stimulate the transcription of mesendoderm genes[17](Figure3). High levels of Nodal signaling promote endoderm development, whereas lower doses specify mesoderm identity, so endoderm cells develop in close proximity to the Nodal source and require a sustained period to be specified[18-20]. Indeed, in mice, the high-nodal environment of the early primitive streak promotes anterior endoderm fate, while cells in the posterior primitive streak are in a lower nodal environment and tend to contribute to posterior endoderm[14]. Nodal signaling promotes the expression of a conserved network of transcription factors which include Mix-like proteins, Foxa2, Sox17, Eomesodermin, and Gata4–6. Together they activate a cascade of gene expression that functions to (a) segregate the endoderm and

mesoderm lineages, (b) commit cells to an endodermal fate, and (c) integrate signaling events that regionalize the nascent endoderm[21]. In particular, *Foxa2* and *Mixl1* are preferentially required for anterior endoderm whereas *Sox17* is required for posterior endoderm, and this correlates with their timing of expression: *Foxa2* is expressed first and *Sox17* is expressed slightly later when the posterior DE emerges from the streak[22-25]. The extrinsic signal that regulates the first transition derives from the notochord and the dorsal aorta: the production of FGF2 and Activinβ2 inhibits the expression of sonic hedgehog (Shh): the exclusion of Shh from the pre-pancreatic region is necessary for the induction of pancreatic markers [26,27].

FGF10 stimulates bud outgrowth and proliferation of the pancreatic progenitor pool marked by the expression of *Pdx1*[28,29]: all pancreatic cell types arise from *Pdx1* positive cells[30]. Pancreatic multipotent progenitors have been identified in the tip and trunk regions of early branching structures of the developing mouse pancreas[31-33]. Based on lineage tracing of CarboxypeptidaseA1 (*Cpa1*)-positive cells marked at E12.5 or earlier, tip multipotent progenitors that co-express *Cpa1*, the transcription factors *Pdx1* and *Ptf1a* and high levels of *cMyc* generate the three major cell types of the pancreas: endocrine, acinar and ductal cells. After E12.5, the pancreatic epithelium differentiates into at least two progenitor regions: the Ptf1a+/Cpa1+ tip that will contribute to forming the acinar cells, and the Sox9+/Hnf1b+/Nkx6.1+ trunk that give rise to a bi-potent progenitor which later will differentiate in ductal or endocrine cells[31-34].

Figure3
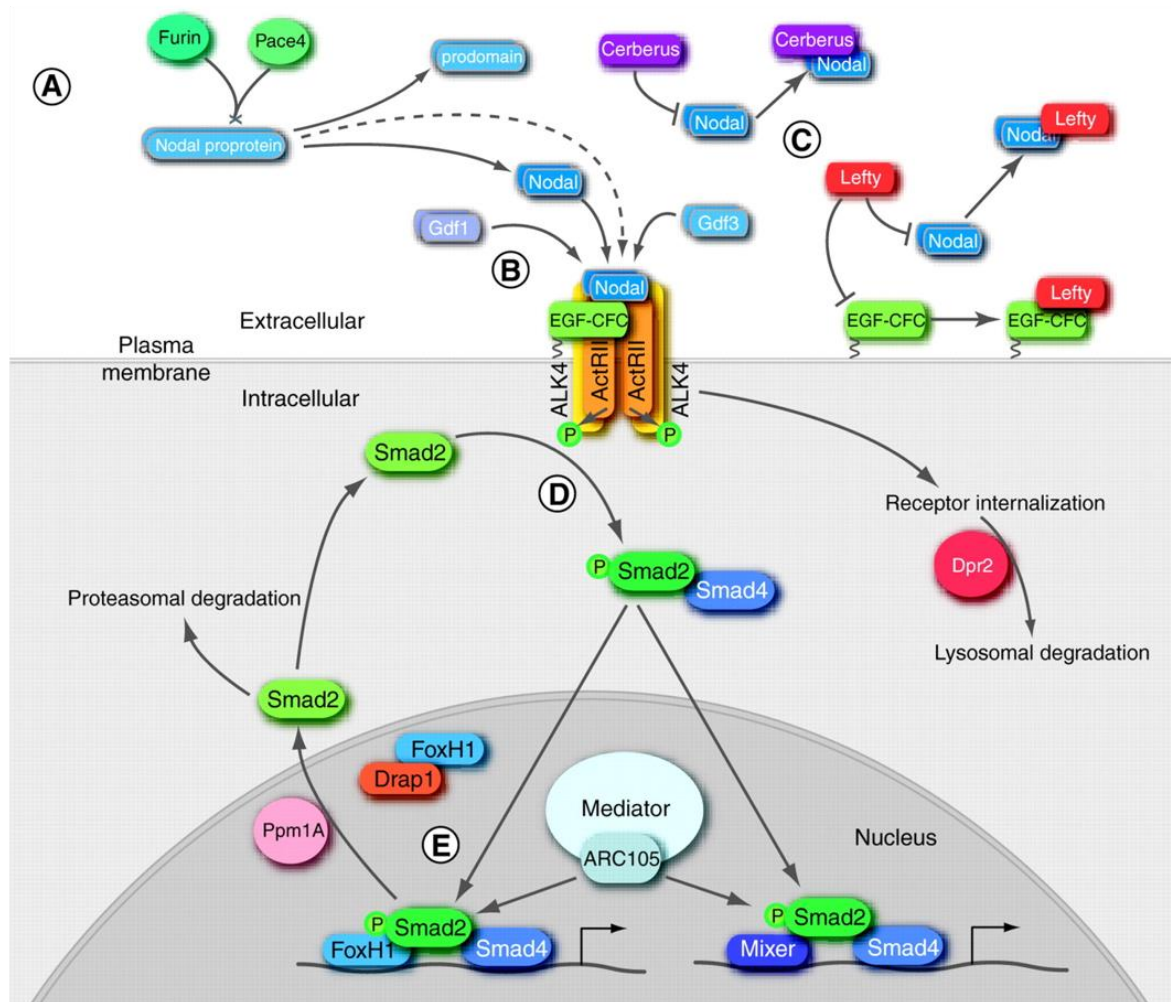
Figure 3. Schematic outline of the Nodal signaling pathway. (**A**) Nodal ligands are expressed as homodimeric proproteins and can be cleaved extracellularly by the proprotein convertases Furin and Pace4. (**B**) Mature Nodal ligands, as well as Gdf1 and Gdf3, can bind to an EGF-CFC co-receptor in a complex with type I receptor (ALK4) and type II receptor (ActRII or ActRIIB) dimers. (**C**) Cerberus and Lefty proteins are soluble antagonists that can interact with Nodal ligands; Lefty proteins can also interact with EGF-CFC co-receptors to inhibit their function. (**D**) Receptor activation leads to the phosphorylation of the type I receptor by the type II kinase, as well as phosphorylation of Smad2 (or Smad3). Activated Smad2 or Smad3 associates with Smad4 and translocates to the nucleus, whereas the receptor complex undergoes internalization into endosomes and can be targeted by Dpr2 for lysosomal degradation. (**E**) Within the nucleus, activated Smad2-Smad4 (or Smad3-Smad4) complexes interact with the winged-helix transcription factor FoxH1 or Mixer homeoproteins on target promoters, leading to transcriptional activation through the interactions with ARC105 and the Mediator complex. Pathway activity can be inhibited by interaction of Drap1 with FoxH1 or by the Smad phosphatase Ppm1A, which promotes the nuclear export of Smad2 and possibly targets it for proteasomal degradation. Adapted from Shen, M. (2007). Nodal signaling: developmental roles and regulation. Development.

2.4. State of art on Nephrocan

The knowledge of genetic mechanisms that guide the *in vivo* endoderm specification is restricted by the lack of early endoderm-specific markers. In order to further improve the comprehension of pancreatic cellular ontogeny, a global expression profile of dorsal pancreatic bud at E10.5 was performed and *Nephrocan* (*Nepn*) was identified among the most enriched genes[35]. *Nepn* is expressed in the definitive endoderm that emerges during gastrulation, as well as its derivative the gut tube epithelium and the early pancreas primordium (E7.5-11.5). *Sox17*, Retinoic Acid (RA) and Nodal/Activin pathway are key players in defining the spatial and temporal expression of *Nepn* in mouse embryo development. *Sox17* is an essential transcription factor (TF) for midgut and hindgut specification and is required to induce *Nepn* expression. The TF has a binding site localized ~700bp upstream *Nepn* TSS whose expression is completely shut off in *Sox17* null mice. RA signaling positively regulates *Nepn* expression while high levels of Nodal/Activin signaling are able to downregulate its expression: the balance between these pathways refine *Nepn* localization in gut development[36].

Nephrocan gene encodes a small leucine-rich repeat protein (SLRP)[37] composed of 17 LRR (Leucine-Rich Repeats) motif, 2 cysteine-rich clusters both in the N and C termini and several potential sites of N-glycosylation. Moreover, a signal peptide in the amino-terminal region of the protein guides the secretion of the protein outside the cell.[38] NEPN has been demonstrated to function as an endogenous inhibitor of the TGF-β/Smad3 signaling pathway, a key regulator in endoderm development. In adult

mice, *Nepn* is expressed mainly in the kidney [38]. In the present study, we report a novel transcript variant of the gene which arises as a result of alternative splicing and is expressed in tissue and developmental stage-specific manner. The newly identified variant was altogether absent between the embryonic stages E8.5-E11.5 and was found to be expressed differentially across different adult tissues. Moreover, we describe *in vivo* and *in vitro* experiments examining *Nepn* involvement in posterior foregut development. We generated and analysed constitutive NepnKOfirst mice (called *Nepn-/-* or *Nepn+/-*). *Nepn−/−* mutants were viable and fertile, exhibited normal body weight and displayed normal glucose tolerance and renal functionality. By using the CRISPR/Cas9 technique[39], we also generated the *Nepn* deficient mESCs. We programmed differentiation of wild-type and mutant mESCs toward the endoderm lineage and demonstrated that the *Nepn* CRISPR knockout impairs the expression of posterior foregut specific markers such as *Gata4/6* and *Sox9*. These experiments will improve the knowledge of mechanisms that guide the in vivo endoderm development which is an important goal to sustain the progression of new therapeutic strategies to be used in the treatment of disease involving endoderm derived organs.

3. Results

3.1. Evidence for a new transcript variant of Nephrocan in mouse.

Nephrocan (*Nepn*) gene is localized on murine chromosome 10 (10qB3) on positive strand, spanning 12,997 bases from position 52,388,864bp to 52,404,613bp (Figure 4a top). To date, only one transcript of 1826bp (NM_025684.2 indicated as *Nepn-201*) has been validated, which contains 3 exons and codes for a protein containing 512 amino acids[38]. Another incomplete transcript is reported on the Ensembl database (Transcript ID: ENSMUST00000219730.1 indicated as *Nepn-202*) consisting of a different exon 1 and a portion of exon 2 with a predicted 201 amino acids polypeptide. Based on this evidence, we hypothesized the presence of a second isoform composed by an alternative exon 1, exon 2 and exon 3. To experimentally validate the existence of this new transcript, we designed two forward primers which recognize specifically the two putative alternative exons 1, and a reverse primer located in exon 3 (Fig 4a bottom). The RT-PCR performed on adult mouse kidney revealed an expected band of 1536bp for *Nepn-201* (hereafter *Nepn* isoform b), and another smaller transcript of 1496bp for *Nepn-202* (hereafter *Nepn* isoform a), indicating that both full-length isoforms exist and are detectable (Figure 4b). As shown in the schematic illustration of *Nepn* genomic locus (Figure 4c), *Nepn* isoform a (1863bp) transcript consists of three exons: exon 1a, exon 2 and exon 3. Almost whole exon 1a consists of a 5'untranslated region (UTR) except for the last two bases, which form the start codon (AUG) with the first base of exon 2. The open reading frame of *Nepn* isoform a, encodes a polypeptide

of 451 amino acids, which differs from *Nepn* isoform b protein for the absence of 62 amino acids at the N-terminus and the consequent loss of a putative signal peptide that directs the protein secretion and cysteine cluster (Figure 4d).

Figure 4
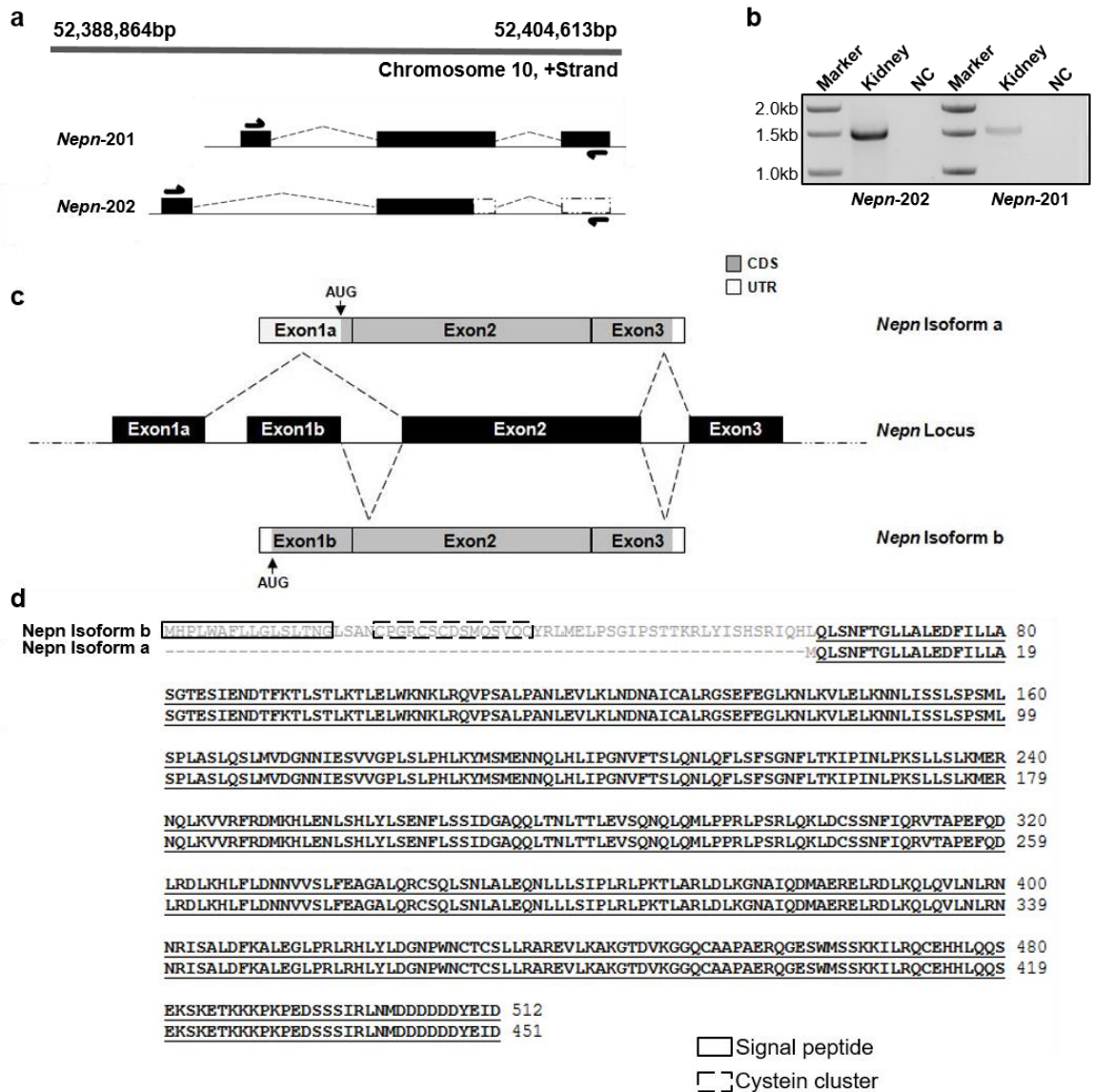
Figure 4. Evidence for a new exon of *Nephrocan* in mouse. (a) The sequence of *Nepn* gene in GenBank runs from 52,388,864 bp to 52,404,613 bp on chromosome 10 (top). Schematic of transcripts reported on Ensemble database and primer designing for RT-PCR (bottom). **(b)** Mouse kidney tissue was used to study the expression of the novel transcript by RT-PCR. NC, negative control with no DNA template. **(c)** Schematic representation of the genomic organization of mouse *Nepn* leading to alternative splicing. Rectangular boxes refer to exons (the size of boxes is indicative of the relative size of exons) and interconnecting lines as introns. The dashed puckered lines show the splicing pattern of the exons. The novel transcript, named *Nepn* Isoform a, is formed by splicing of Exon1a with Exon2 skipping Exon1b. The translation initiation site "AUG" is indicated with a downward arrow in the new transcript. **(d)** Alignment of *Nepn* isoform b (upper sequence) and isoform a (lower sequence) obtained using the Sequence Manipulation Suite programs.

3.2. Nepn isoforms expression pattern in mouse development and tissues.

As the existence of a new *Nepn* transcript has been validated, we analysed the expression pattern during embryogenesis and the tissue-specific distribution of both isoforms. Interestingly, during embryo development, only *Nepn* isoform b transcript is detectable, and in particular, it shows a well-defined time window expression between the embryonic stages E8.5-E11.5 (Figure 5a). Since the restricted expression to the dorsal pancreatic bud (DPB) at the developmental stage E10.5 has been already reported[38], this strengthens the hypothesis that it could play an important role in the pancreatic specification. We also analysed the expression of both isoforms in several adult mouse tissues, deriving from all three germ layers. As shown in Figure 5b, the expression of *Nepn* isoform a was detected in most of the tissues analyzed except for muscle and thyroid, whereas *Nepn* isoform b was found to be expressed in kidney, thyroid, pancreas and, to a much lesser extent, in testis and muscle.

Figure 5
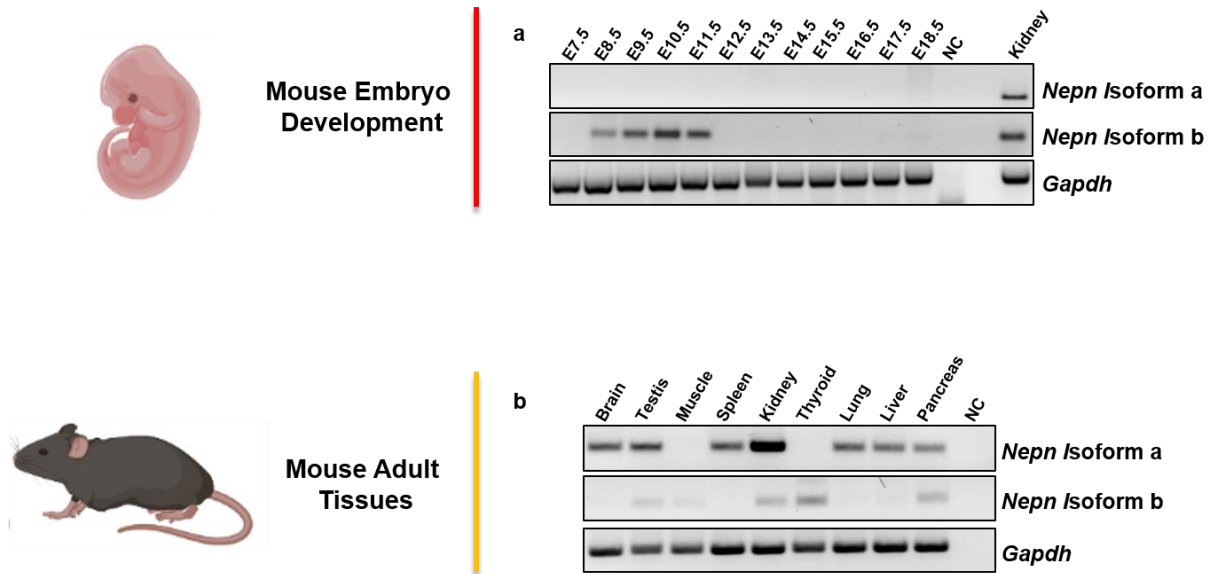
Figure 5. Expression pattern of *Nepn* isoforms. *Nepn* isoforms a and b were amplified by RT-PCR on total RNA from adult mouse organs and embryos homogenate. Two different PCR reactions were performed on the same template cDNA using specific oligos. **(a)** *Nepn* isoforms a and b expression in mouse embryo development; **(b)** *Nepn* isoforms a and b expression in adult mouse tissues. *Gapdh* amplification was performed as a control.

3.3. Nepn knockout first mouse (Nepn^KOfirst) generation.

To shed light on *Nepn* role in pancreatic development, mESCs from Komp Repository, carrying a genetically modified *Nepn* allele (Figure 6a), were microinjected into C57BL/6N blastocysts and transferred to a pseudopregnant recipient mice[40]. Chimeric offspring were then mated with wild-type mice which resulted in a certain percent of heterozygous progeny carrying the transgene in F1. Finally, the heterozygous mice were crossed to obtain homozygous mutant mice in F2 (Figure 6b top). To discriminate homozygous from heterozygous mice, a genotyping analysis was performed using two different forward primers, each one specific for the wild-type or recombinant allele, in combination with a common reverse primer (Figure 6a), which resolved in a different amplicons size (Figure 6c). To avoid the problem of neo toxicity in homozygous *Nepn^KOfirst* mouse, neo expression cassette was removed by crossing muted mouse with C57BL/6N congenic constitutive Cre deleted mice[41]. All experiments illustrated are performed using *Nepn^KOfirstΔneo* mice (so-called *Nepn-/-* or *Nepn+/-*). Analyses of Mendelian ratio from the breeding of *Nepn* heterozygous mutant mice (*Nepn+/-*) show no altered Mendelian ratio suggesting that loss of *Nepn* does not affect mice survival (Figure 6b bottom). Moreover, WT and *Nepn* KO littermates were similar in size, in the appearance of the fur and gross morphology.

In order to further validate the absence of both *Nepn* isoforms in the knockout first mouse, we performed a RT-qPCR on the transcript's portion shared by both isoforms. Since it was still possible to appreciate a residual expression of the gene in *Nepn-/-* mouse (Figure 6d), we performed an RT-PCR to discriminate which of the two isoforms

was still expressed. As shown in figure 6e, it was possible to detect a slight expression of *Nepn* isoform a, whereas *Nepn* isoform b was completely shut down. We cannot rule out that this tiny expression could overcome KO effects.

Figure 6

a

Recombinant allele

Ex1A | Ex1B | FRT | En2SA | IRES | lacZ | pA | loxP | hBactP | neo | pA | FRT | loxP | Ex2 | Ex3

Neo-Fw   562 bp   Nepntt-Rv

Wild-type allele

Ex1A | Ex1B | Ex2 | Ex3

267 bp

Bax1b-5'   Nepntt-Rv

b

Chimeric Mouse   X   Wild Type Mouse

Nepn Heterozygous Mouse   Nepn Heterozygous Mouse

F1   X

Wild Type Mouse   Nepn Heterozygous Mouse   Nepn null Homozygous Mouse

F2

Observed Frequency   29.1%   53.5%   17.4%

c

bp   Nepn +/-   Nepn -/-   Nepn +/+

562   Recombinant allele

267   Wild-type allele

d

Nepn mRNA relative expression

Kidney

□ Nepn+/+
▨ Nepn+/-
■ Nepn-/-

** *

e

Nepn+/+   Nepn+/-   Nepn-/-   NC

Nepn Isoform a

Gapdh

Kidney
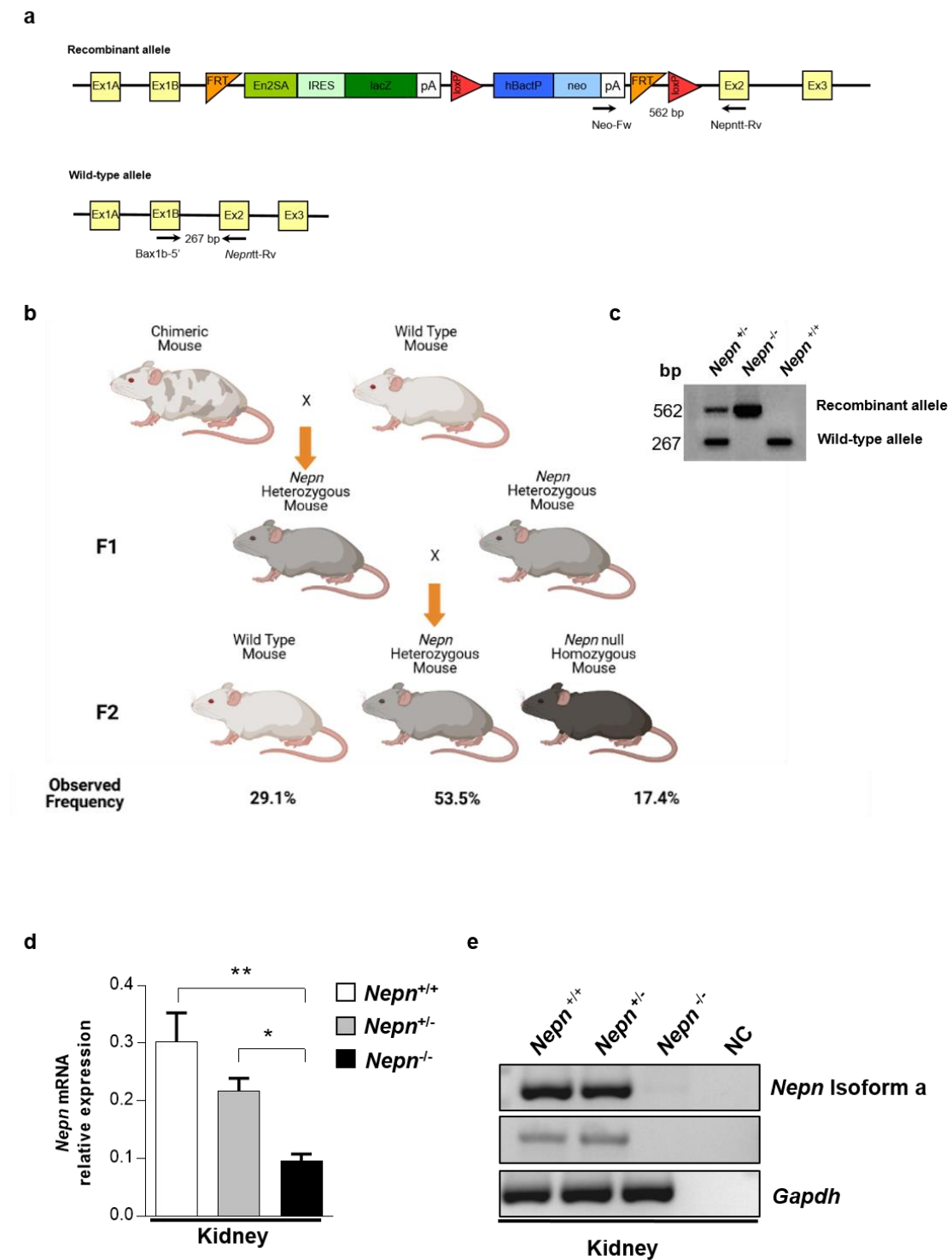
Figure 6. Generation of a mouse model for *Nepn* gene depletion. (a) Graphical representation of *Nepn* mutated locus in mESCs used to generate *Nepn* KO first mouse. Primer designs for RT-PCR are indicated as arrows in the picture. **(b)** Top Schematic representation of *Nepn* null mouse model generation and analyses of Mendelian ratio from the breeding of *Nepn* heterozygous mutant mice (bottom). **(c)** PCR analysis of genomic DNA isolated from mouse ear. The upper band in the panel displays the *Nepn* recombinant allele, while the lower band displays *Nepn* wild-type allele. Thus, the appearance of the upper band alone displays mutated homozygous allele; the lower band alone represents wild-type *Nepn* allele; while both bands together mean that the mouse is *Nepn* heterozygous. **(d)** RT-qPCR was performed to quantify the minimal expression of *Nepn* in *Nepn-/-* mouse and relative controls. The data reported are normalized on *Gapdh* expression. Three replicates for each experimental point were performed. Error bars represent the standard deviation of normalized values (* $p < 0.05$, ** $p < 0.01$). **(e)** *Nepn* isoforms a and b expression in *Nepn-/-* mouse kidney and relative controls.

3.4. Characterization of pancreatic and renal functionality.

Since our main interest is to characterize the role of *Nepn* in pancreatic homeostasis we evaluated the pancreatic functionality of *Nepn* knockout mouse model, performing an intraperitoneal glucose tolerance test (IPGTT)[42]. After intraperitoneal injection, we measured the blood glucose levels at different time points during the following 120min, and as shown in figure 7a top *Nepn-/-* animals have the same glucose induction and clearance curve compared to *Nepn+/+* mice. This points out that the absence of *Nepn* does not result in any type of alteration of glucose metabolism in mice of 4-6 months old. The same measurements were carried out on mice at different ages (7-12 months) but still, no differences were observed (Figure 7a bottom). The intraperitoneal glucose administration can exclude the implication of a response driven by incretin hormones, which are gut peptides that are secreted after nutrient intake and stimulate insulin secretion together with hyperglycemia[43]. We, therefore, examined whether *Nepn* gene depletion may affect body weight control at different ages. *Nepn-/-* mice body weight was measured and compared to *Nepn+/+* mice: both mice cohorts gained similar amounts of weight indicating that the absence of *Nepn* is not directly correlated to metabolism and nutrient assimilation (Figure 7b).

Nepn transcript is predominantly expressed in the kidney in adult mice[38], therefore we investigated the impact of *Nepn* deficiency on renal functionality. *Nepn-/-* mice were accommodated in metabolic cages and urine volume, osmolality, electrolytes, and creatinine clearance were evaluated to assess renal homeostasis. Table 1 summarizes urinary parameter values measured in mice of 4-6 months old, showing the absence

of significant variations between wild-type and mutant mice in renal functionality. The mice were monitored until 12 months, evaluating the same renal parameters to assess that no metabolism changes could occur in aged mice (Table 2). Finally, serum analysis showed no differences in triglycerides, total cholesterol, urea, calcium and phosphorus levels (Figure 8).

Figure 7

a

Intraperitoneal Glucose Tolerance Test (IPGTT)

4-6 Months

7-12 Months

b

Body Weight Examination
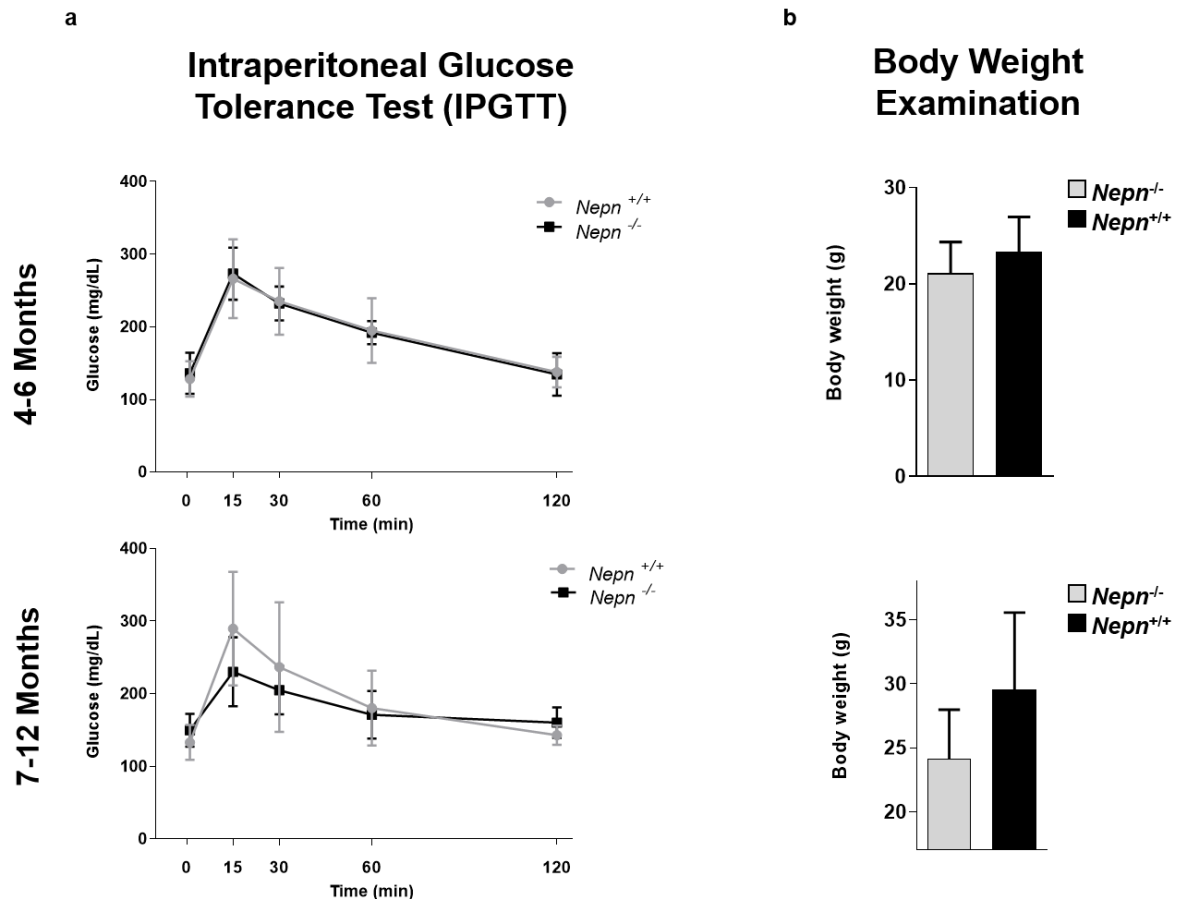
Figure 7. *Nepn* null mice have normal glucose regulatory function and body weight. (**a**) Intraperitoneal Glucose Tolerance Test (IPGTT). Evaluation of blood glucose of *Nepn-/-* mice compared to *Nepn+/+* in a group of 4–6 months and 7-12 month old mice. Blood glucose was measured at the indicated time points after intraperitoneal glucose injections (*Nepn+/+*: seven mice; *Nepn-/-*: seven mice). The plasma glucose concentration peaked at 15 min after the glucose challenge and then gradually returned to normal level throughout the experiment. No significant differences between wild-type and *Nepn* null mice can be observed (p = 0.9, n = 7 for each group). Data are expressed as the mean±SD. (**b**) The body weight of 4 to 6 months old mice and 7 to 12 months old mice from wild-type and *Nepn* null groups. Results are means±SD. T-test revealed no discernible differences between the genotypes (p = 0.26, n = 6 for each group).

Table 1

URINARY PARAMETERS	Nepn $^{+/+}$	Nepn $^{-/-}$
Urinary Volume (ml)	1.86±0.37	1.63±0.25
Creatinine excretion (µmol/g body weight)	0.24±0.03	0.23±0.01
Na$^+$/ creatinine	22.75±4.16	27.92±4.41
K$^+$/ creatinine	20.36±3.95	18.14±2.23
Cl$^-$/ creatinine	62.91±6.43	72.34±3.43
Creatinine clearence	113.16±14.91	97.56±11.33

Table 1. Experimental study of renal functionality. Evaluation of urinary parameters in 24h urine output of *Nepn-/-* mice compared to *Nepn+/+* in a group of 4-6 months old mice.

Table 2

URINARY PARAMETERS	Nepn $^{+/+}$	Nepn $^{-/-}$
Urinary Volume (ml)	2.14±0.78	2.36±0.69
Creatinine excretion (µmol/g body weight)	0.26±0.02	0.26±0.03
Na$^+$/ creatinine	27.42±2.78	24.01±3.15
K$^+$/ creatinine	27.92±2.91	34.66±4.40
Cl$^-$/ creatinine	59.53±6.24	51.50±5.63
Creatinine clearence	120.59±8.83	113.65±15.73

Table 2. Experimental study of renal functionality. Evaluation of urinary parameters in 24h urine output of Nepn-/- mice compared to Nepn+/+ in a group of 7-12 months old mice.
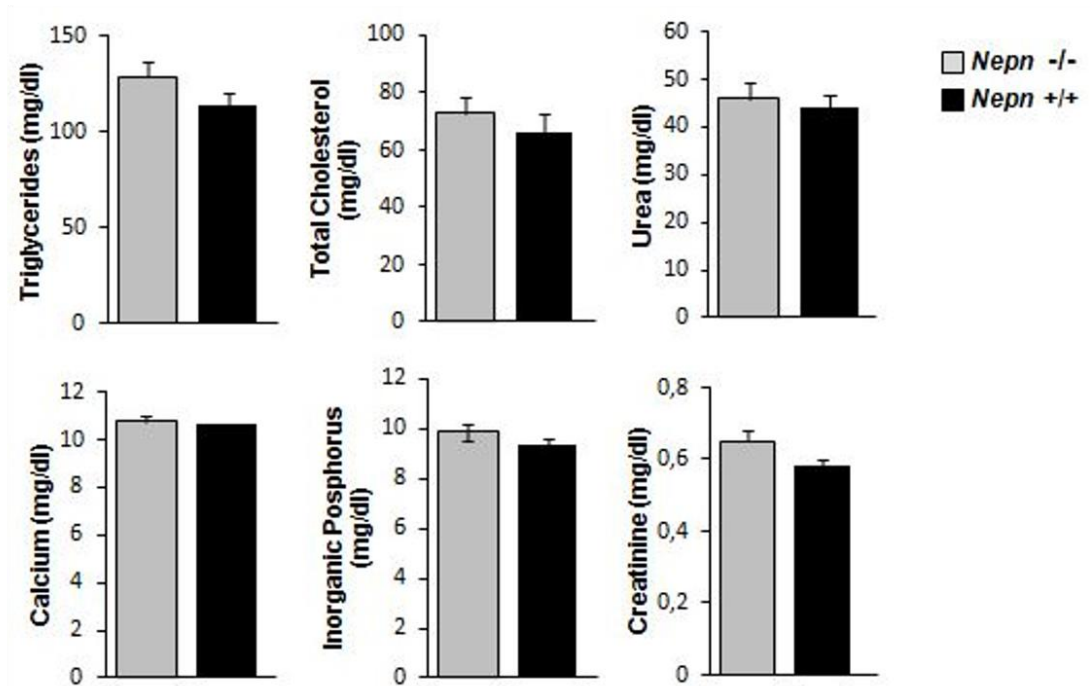
Figure 8

Figure 8: Serum levels parameters in a group of 4-6 months old mice. Triglycerides, total cholesterol, urea, calcium and phosphorus were measured and no significant differences between WT and *Nepn* KO mice can be observed.

3.5. Generation of Nepn deficient mESCs by RNA-guided CRISPR/Cas9.

Since *Nepn-/-* mouse has shown no phenotype, we moved to a new cell system to pinpoint *Nepn* role in endodermal development, generating a homozygous deletion for *Nepn* in mESCs using CRISPR/Cas9 genome editing technology. First, we designed a gRNA to target *Nepn* isoform b (Figure 9a top) which was found to be the exclusive one expressed in mouse development (Figure 5a). Second, the guide RNA was inserted in a construct encoding Cas9 for homologous recombination in Rosa26 genomic locus (Figure 9b bottom) and then electroporated into wild-type mESCs. Finally, after selection, resistant clones were picked and expanded. Sanger-sequencing results revealed, among all picked colonies, a specific clone carrying a biallelic 13bp deletion in Nepn exon1b (Figure 9b). The bi-allelic deletion causes a frameshift mutation few bases downstream ATG, resulting in a truncated peptide of 90 amino acids (Figure 9c). To further confirm the biallelic deletion, the selected clone was tested by RT-PCR using two different primer pairs. The first pair flanks the deleted region, which resolves in a band shift due to the different length in the amplicon size (Figure 9d top); the second pair has the forward primer which anneals on the deleted fragment leading to an amplification product only in the wild-type cell line (Figure 9d bottom). *Nepn* mutant mESCs colonies were indistinguishable in size and shape from unmodified cells (Figure 9e left), as well as growth rate was found to be comparable (data not shown). Moreover, we observed that the *Nepn* isoform b −/− mESCs showed no apparent loss of pluripotency since the expression of *Oct4* and *Nanog* was not

reduced (Figure 9 right), indicating that neither the genetic modification nor the

absence of *Nepn* affected pluripotency and self-renewal properties.

Figure 9

a

gRNA target sequence: 5'-**AGGCTCATGGAACTCCCATC**-3'

b

Nepn
Isoform b⁺/⁺ TGTGATTCTATGCAGTCTGTACAATGCTAC**AGG**CTCATGGAACTCCCATCAGGCATTCCTTCCACCACCAAGAGACTCTACATTAGCCACAGCAGAATCC

Nepn
Isoform b⁻/⁻ TGTGATTCTATGCAGTCTGTACAATGCTAC**AGG**CTCATGGAA-------------TTCCTTCCACCACCAAGAGACTCTACATTAGCCACAGCAGAATCC

**13 bp
deletion**

Nepn Isoform b⁺/⁺

T GT ACAA T GC TACA G GC T CAT GGAAC T CCCAT CA G GCAT T CCT T CCAC CAC CAA GA GA C T CTA CAT T A

Nepn Isoform b⁻/⁻

T GT ACAA T GC TACA G GC T CAT GGAA K T CCT T CCAC CAC CAA GA GA C T CTA CAT T A GC CACA GCA G

c

```
MHPLWAFLLG LSLTNGLSAN CPGRCSCDSM QSVQCYRLME FLPPPRDSTL ATAESSTFSY LTSLDCWLWK ILFCWPVEQS 80
LLKMTPSKL* VH*RL*NFGK IN*DKSPVLS QPTLRC*N*M IMPYVLYVDQ SLKD*RT*KS LNLKTT*SRP CLPACSPHLP 160
VCRV*WWMVT T*SPWLDL*A FRISST*AWR TTNSTLYQGM SSLPYRICSS SVSVVIF*LK FP*IYPNPCY L*RWRETSSK 240
WSGFAI*NTW RTCPIFTCQR TSCLPLMGHS S*PI*QLWRF PKTSFKCCLP GYHRGYRSLI VAVTSFRE*Q HQNSRTSET* 320
NICF*TTMLS ACSRLGPFRG VLSCPTWPWS RICCCLSL*G SLRPWPGWTS RAMPSRTWLK GSSGTSNNCR SSTSGTTGSR 400
P*ISRHWRGC LGSGISTWMG IPGTALAVSY EPGKS*RPRA RT*KGDSAQP QQSDRGRAGC PPRRS*GNVS ITYSRVRKAK 480
RPRRNPNPKT LPASDSTWMT TMTIMK*I 508
```
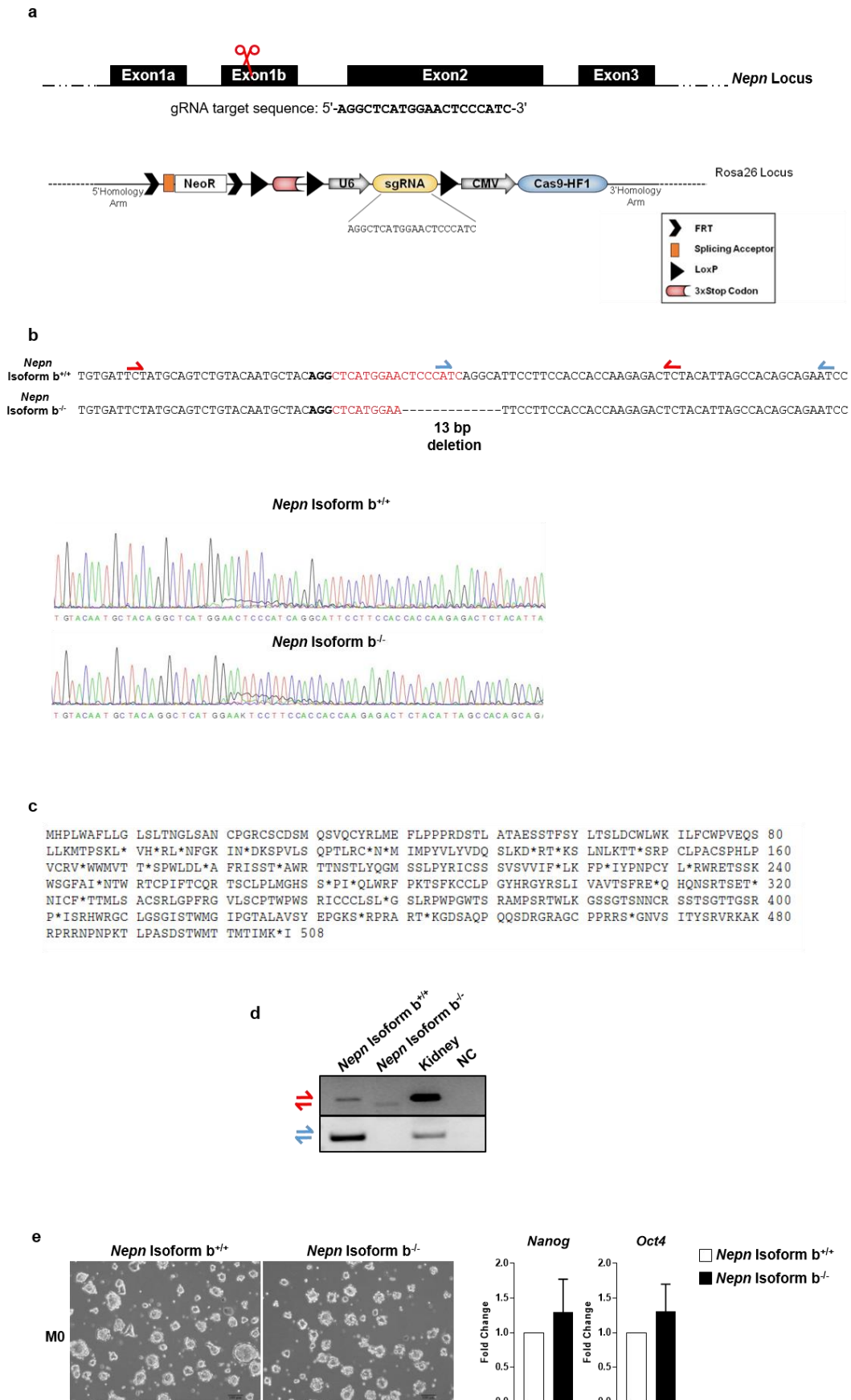
d

e

30

Figure 9. Generation of *Nepn* knockout mESCs. (a) top. Schematic diagram of the location and sequence of gRNA designed to target Exon1b of *Nepn* gene. bottom The selected sgRNA was inserted under the control of U6 promoter, in a construct containing Cas9 expression cassette and Rosa26 homology arms. **(b)** Chromatogram of the representative wild-type and CRISPR/Cas9 *Nepn* mutant clone. The interpretation shows a mutated allele aligned against the wild-type sequence. The bold letters represent the PAM sequence, while the dotted lines indicate deletions. Red and blue half arrows indicate the primer pairs used for the validation of CRISPR/Cas9 deletion. **(c)** In the selected *Nepn-/-* clone a frameshift mutation generates several stop codons (indicates as *) thus inhibiting protein synthesis. **(d)** Identification of a deletion in *Nepn* Isoform b gene sequence. Two PCR were performed once using primers flanking the gRNA target region, which leads to a smaller amplicon in the mutated allele. The second PCR is performed using a forward primer that anneals on the deleted region, resulting in no amplification on the mutated allele. **(e)** (Left) Representative brightfield images of undifferentiated (M0) wild-type mESCs, and *Nepn* Isoform b knockout mESCs colonies. The colonies look alike and cells do not present any differences in generating colonies. Scale bars: 100 μm. (Right) RT-qPCR results show the expression level of stem cell markers (*Oct4*, *Nanog*). No significant differences between wild-type mESCs and *Nepn* Isoform b knockout mESCs can be observed.

3.6. Nepn deficiency impairs endoderm lineage commitment

In vivo Nepn expression is restricted to a small subpopulation of endodermal cells at E7.25, afterwards (E8.0-8.5) was found to be expressed in a wider region known as midgut[36], to then relocate to the limited region of DPB[44]. We, therefore, focused on the possibility that *Nepn* may play a role in endodermal development and used an *in vitro* differentiation protocol, previously established in our laboratory, that is intended to recapitulate the endoderm development *in vivo*. Briefly, definitive endoderm (DE) formation (Metastate 1, M1) is induced during the priming step by Activin A. Sequently, the cells are further directed toward the posterior foregut endoderm (PFE) (Metastate 2, M2) by RA and FGF10 (Figure 10a). We also confirmed that, as *in vivo* also *in vitro*, *Nepn* isoform b is the exclusive isoform expressed through developmental stages (Figure 10b). Finally, we induced mutant and wild-type cells to differentiate, and evaluated the expression of lineage-specific markers. RT-qPCR analyses showed that the expression of DE marker *FoxA2* remained mainly unchanged in *Nepn* isoform b-/- compared to *Nepn* isoform b +/+ suggesting that definitive endoderm induction was not affected by biallelic mutation of *Nepn* (Figure 10c left). Interestingly, although there was no visible morphology difference between wild type and mutant cell line (Figure 10c right), the absence of *Nepn* isoform b appeared to affect the formation of PFE, since its specific markers *Sox9, Gata-4 and Gata-6*, turned out to be expressed to a lesser extent in mutant cell line compared to wild-type (Figure 10c left).
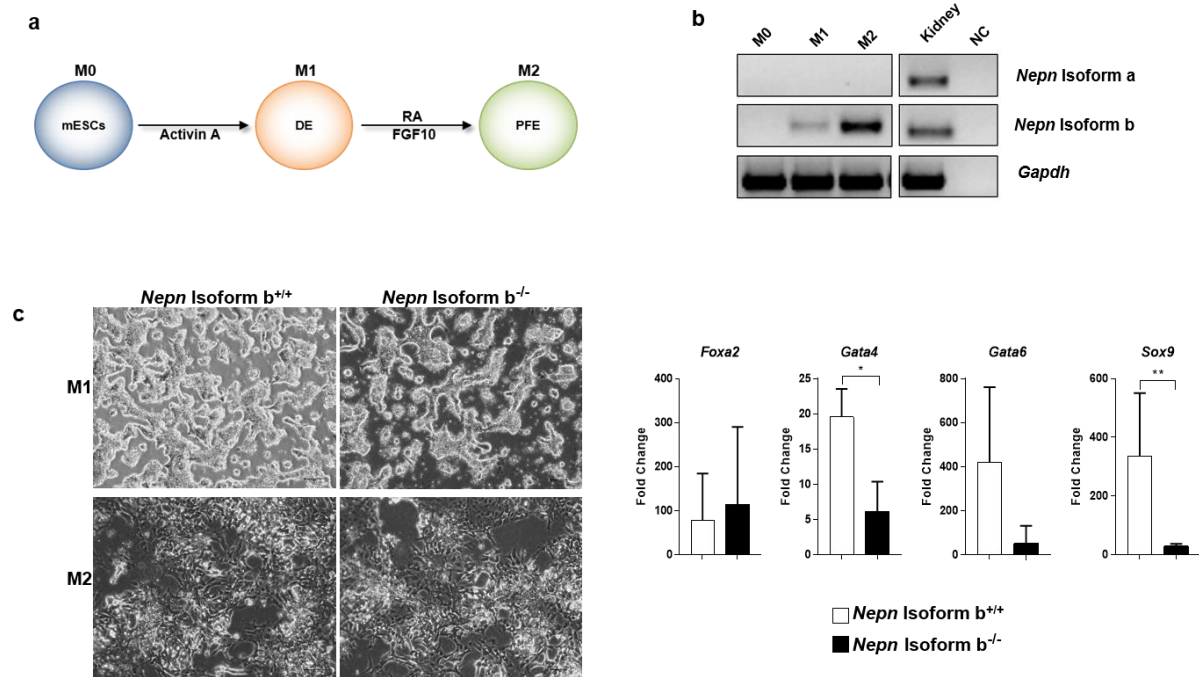
Figure 10

Figure 10. Evaluation on *Nepn* deficiency during endoderm differentiation (a) Schematic protocol of directed differentiation from mESCs into posterior foregut endoderm (PFE). ESCs were plated at 50,000 cells/cm^2 at day0 (M0) in a pro differentiative medium supplemented with Matrigel (200 µg/mL) and Activin A (20 ng/mL) to induce the DE formation (M1). The cells were then treated with Retinoic Acid (5 µM) and FGF10 (10 ng/mL) until the end of protocol to promote the PFE formation (M2). **(b)** *Nepn* isoforms expression through the *in vitro* differentiation was evaluated by RT-PCR. Kidney cDNA was used as a positive control (Ctrl). **(c)** (Left) Cells morphology during *in vitro* endoderm differentiation. Representative brightfield images of wild type and *Nepn* Isoform b knockout differentiated cells at M1 and M2. Images were taken with a Leica DMi8 at 10x magnification. Scale bar: 100µm. (Right) RT-qPCR analysis of the DE marker (*FoxA2*) and PFE markers (*Gata4*, *Gata6* and *Sox9*). The mRNA levels were normalized to *Gapdh* expression and reported as fold change with respect to the value in M1. Values shown are mean±SD, based on triplicate assays. Statistical analyses were performed using Student's t-test, where $p < 0.05$ was considered significant. (* $p < 0.05$, ** $p < 0.01$).

4. Discussion and Future Perspectives

The embryonic definitive endoderm (DE) gives rise to organs of the respiratory and gastrointestinal tract including the pancreas, liver, and epithelia of the colon and lung. Due to a poorly knowledge of molecular mechanisms that guide the *in vivo* endoderm specification, interest in the identification of novel endoderm genes is growing. On the other hand, elucidation of cellular components and genes governing tissue development is critical to understand the cause of organ disorders and cancers and will lead to novel therapies for tissue and organ regeneration. Nepn is a member of a small leucine-repeat (SLRP) family of proteins. It is a secreted, N-glycosylated inhibitor of TGF-β and is expressed in the early stages of murine endoderm specification (E7.5-11.5) [36,38,44]. Here, a novel transcript of mouse *Nepn* gene was identified through bioinformatics tools and molecular biology techniques. This novel transcript, named *Nepn* isoform a, was found to be expressed in different tissues but not during embryo development. The alignment of isoforms protein sequences highlighted the absence of the secretion signal peptide at the N-terminus. Since the cellular localization of a protein plays an important role in defining its function, we speculate that the absence of important domains in the smaller variant might directly affect the secretion capability and activity of this protein. To further characterize *Nepn* role in endoderm patterning, we generated a mouse model with genetic depletion of *Nepn* but found no obvious phenotype alterations, not even pancreas and renal functionality. One possible explanation for this apparent lack of an effect could be the presence of residual

expression of *Nepn* isoform a that could overcome KO effects. Moreover, it is possible that *Nepn* function may instead rely on other SLRP family members, since, e.g., *Decorin* (DCN) has been shown to bind TGF-β as a regulator of growth factor functions[45-47]. At present, 15 SLRP members have been cloned and partially characterized, and most are located in clusters on mouse and human chromosome [48]. Here, we have also generated a nullizygous *Nepn* knock-out mESC line using the CRISPR/Cas9 system. This technology, due to its advantages in the high efficiency and specificity of gene targeting, is largely used in gene-editing [39]. Combining embryonic stem cell technology with CRISPR/Cas9 gene-editing provides an important strategy to clarify the biological role of markers, the understanding of normal embryonic development as well as the pathogenesis of diseases. We applied an *in vitro* differentiation protocol, previously established in our laboratory [44], that allows for robust and efficient generation of endodermal cells within eight days under chemically defined conditions. Interestingly, it is notable that posterior foregut markers, such as *Sox9, Gata4* and *Gata6*, turned out to be expressed to a lesser extent in mutant cell line compared to wild-type. It has been reported that the homozygous deletion of both *Gata4* and *Gata6* is necessary to disrupt pancreas development in mice [49-51]. Moreover, Xuan and coworkers recently demonstrated the importance of GATA4/6-mediated inhibition of hedgehog signaling as a major mechanism regulating pancreatic endoderm specification during patterning of the gut tube [52]. Our findings suggest that *Nepn* KO, interfering with *Gata4* and *Gata6* expression, may have an effect on the posterior foregut (PF) development and on the cells ability to reach posterior foregut metastate. To demonstrate the correlation between *Nepn* loss and the failure of PF induction our future plans foresee carrying out a rescue of the phenotype in the KO

clone to verify if the restoration of the gene is associated with the recovery of the function. Moreover, it would be possible to evaluate the effect of the absence of *Nepn* Isoform b in mouse development using an *in vitro* model called gastruloid[53]. They are three-dimensional structures derived from ESCs which in the appropriate culture conditions are induced to differentiate, summarizing the early stages of embryonic development up to gastrulation and then to the formation of the three embryonic germ layers: ectoderm, mesoderm and endoderm. Finally, it will be interesting to investigate the role of *Nepn* isoform a and characterize the molecular interaction between *Nepn-Gatas* signaling during mouse endoderm development.

5. Materials and Methods

5.1. Ethics statements and animal experiments.

All animal studies were conducted at Biogem S.c.a.r.l. Ariano Irpino, AV (Italy), Preclinical Research and Development Service. Animals have been housed and used following the rules of the Italian laws (DL.vo N° 116 - 27/01/1992 and related) and of the EU directive (2010/63/UE - 22/09/2010) on the protection of animals used for experimental purposes. All the *in vivo* procedures were in compliance with the Guide for the Care and Use of Laboratory Animals (United States National Research Council, 1996). All the *in vivo* experimental activities were evaluated and approved by the Committee for the Ethics of the Experimentations on Animals (CESA) and by Organismo Benessere Animale (OBA) of Biogem with ID: 7717 (Protocol Number: 7F782_20) and were authorized by the Italian Minister of Health (Authorization Number 384/2017-PR) according to the decree nr 100/2006 and later decree 26/2014. Wild type C57BL/6J (herein referred to as B6) mice were purchased from Charles River Laboratories. Animals were housed in an animal house under controlled conditions of temperature (22 ± 1 °C), humidity (55 ± 10%) and lighting on a 12-h light/12-h dark cycle and were supplied with standard rodent food and water ad libitum. Mice for testing were produced by crossing heterozygous females with heterozygous males. Littermate controls were used for all experiments. For tissue collections, all surgery was performed under anesthesia. All efforts were made to minimize suffering.

5.2. Generation of a constitutive Nepn knockout mouse line.

For the generation of the *Nepn^KOfirst* mouse line, one positive ES cell JM8.N4 clones (EPD0686_5_C01), containing a *Nepn* modified locus (*Nepn^KOfirst* allele), were purchased from Knockout Mouse Project (KOMP) Repository. Briefly, this knockout first allele contained a 2080 bp 5' homology arm, a FRT-flanked cassette containing a lacZ reporter gene driven by Nepn promoter, a PGK neomycin resistance gene for positive selection (both separated by a loxP sequence), 2 loxP sites flanking exons 2 and 3 of *Nepn* gene as well as a 6930 bp 3' homology arm. To confirm the presence of the second loxP site flanking the exons 2 and 3 in these ES clones, long-range PCR was used to amplify the 3' modified homology arm using KOMP primers. The presence of the loxP site generates a new restriction site for the restriction endonuclease SacI. Thus, digestion with this enzyme and sequence analysis allowed us to confirm the presence of this modification and therefore the presence of the *Nepn* modified locus. The positive JM8.N4 modified clone containing *Nepn^KOfirst* alleles was subsequently injected into C57Bl/6N blastocysts. Only male mice positive for germ-line transmission were used as the founder of the *Nepn^KOfirst* mouse line. By breeding *Nepn* heterozygous mutant mice, we obtained the constitutive *Nepn* knockout mice. Mouse genotyping was performed with genomic DNA samples isolated from the ear. Standard PCR protocols were used to amplify the wild-type and knockout first alleles of the *Nepn* gene. Mice genotyping was verified by PCR using the following primers: Cds-Neo, Bax1b-5' and Cds-Nepn-tt (Table 3). To remove neo expression cassette, *Nepn^KOfirst* mice were intercrossed with C57BL/6N congenic constitutive Cre deleted mice [doi: 10.1002/gene.10030] and additional rounds of breeding were then performed in order to produce animals that were negative for Cre transgene. Mouse genotyping was

performed with genomic DNA samples isolated from ear biopsies that were taken from each animal at weaning. Standard PCR protocols were used to amplify the wildtype and knockout alleles of the *Nepn* gene. PCR was performed using the following primers: lacZ-(Cre)-Fw, Int1-Fw and Int1-Rv (see Table 3).

5.3. Cell culture and differentiation protocol.

Undifferentiated wild-type and the CRISPR/Cas9 knockout mESCs were cultured on gelatin-coated feeder-free plates in Dulbecco's Modified Eagle Medium (Sigma Aldrich) supplemented with 15% FBS (GE Healthcare), 1,000 units/mL ESGRO leukaemia inhibitory factor (LIF) (Merck Millipore), 1.0 mM sodium pyruvate (Invitrogen), 0.1 mM non-essential amino acids (Invitrogen), 2.0 mM L-glutamine (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma Aldrich) and 500 U ml-1 penicillin/streptomycin (Invitrogen). ESCs were incubated at 37°C in 5% CO_2; the medium was changed daily and cells were split every 2 to 3 days routinely. For the endoderm differentiation, 5×10^4 cells/cm^2 were seeded in 35mm dishes, as a feeder-free monolayer. The differentiation medium consists of DMEM Low Glucose (Lonza), supplemented with 5% FBS (HyClone), 2mM L-Glutamine (Gibco), 1mM non-essential amino acids (Gibco), 100 U-μg/ml Penicillin/Streptomycin (Gibco), 0,1mM β-Mercaptoethanol (Sigma), depleted of LIF, with 200 μg/ml Matrigel (BD Biosciences) and treated sequentially with several molecules at a different time point for 8 days. These factors included Activin A (20 ng/ml, R&D Systems), all-trans Retinoic Acid (RA, 5 μM, Sigma), fibroblast growth factor 10 (Fgf10, 10 ng/ml, R&D Systems), cyclopamine (CYC, 10 μM, Sigma), N-N-(3,5-difluorophenacetyl)-

Lalanylsphenylglycinet- butylesterm (DAPT, 5 µM, Sigma). The medium was changed every 2 days.

5.4. Vector construction.

SpCas9 was used for editing the first exon of the *Nepn* gene. The mammalian codon-optimized Streptococcus pyogenes *Cas9* gene was obtained from the VP12 plasmid (Addgene plasmid ID 72247). It was digested with the two restriction enzymes *NotI* and *PmeI* and cloned into pENTR1A no CCDB plasmid (Addgene plasmid ID:17398). The gRNA expression cassette was amplified from the BPK1520 plasmid (Addgene plasmid ID: 65777), the amplicon was digested with *SpeI* and cloned into the pENTR1A no CCDB/ Cas9 vector. gRNA4 was designed using crispr.mit.edu web tool. Paired oligos corresponding to *Nepn* gRNA4 (5'-AGGCTCATGGAACTCCCATC-3') were cloned into the vector. In vitro recombination between pENTR1A no CCDB/ BPK1520 /Cas9 and pBS-Rosa26/CCDB vectors were obtained using the Gateway LR Clonase II enzyme mix according to the manufacturer's directions (Invitrogen).

5.5. CRISPR/Cas9 genome editing.

pENTR1A no CCDB/ BPK1520 /Cas9-pBS-Rosa26/CCDB construct was electroporated into mESCs. One to two weeks later, colonies were growing in the culture plate. Using 30µl pipette tips the colonies were picked up and placed individually in the 96 well culture plate filled with trypsin. After dissociation into single cells, the colonies were transferred onto gelatin-coated 24-well plates and then

expanded in 10 cm plates separately. Further, the genotyping by PCR/TA-cloning and chromatogram sequencing were used to analyze the mutations and select the positive clones for analysis and differentiation experiments. The primers (5′–3′) used for genotyping are Nepn Crispr F3-R2 (see Table3).

5.6. RNA extraction, RT-PCR analysis and quantitative real time RT–PCR.

Total RNA was extracted using TRIzol reagent (Invitrogen) and cDNA was synthesized using iScript cDNA Synthesis kit according to the manufacturer's instructions (Biorad). 1 µg of total RNA was used for each cDNA synthesis. Primer Express software was used to design the oligo primers setting the annealing temperature to 59–61 °C for all primer pairs. Oligo sequences are reported in Table3. Each PCR reaction was performed with 25 ng of single-stranded cDNA as template and the appropriate set of forward and reverse primers. PCR amplification conditions are reported in Table4. For *Nepn-201* and *Nepn-202* was performed under the following: initial denaturation for 3min at 95°C; repeated 35cycles of denaturation for 30s at 93°C; annealing for 20s at 58°C; and extension for 40s at 72°C. This was followed by final extension at 72°C for 8min. For gene expression analyses, the same amount of cDNA (25 ng) was used for each PCR reaction with each primer pair (forward/reverse primers mix: 0.2 µM, in a final volume of 25 µL). Real time-PCR analysis was performed using the iTaq™ Universal SYBR® Green Supermix (BIORAD) in a 7500 Real-Time PCR System (Applied Biosystems) under the following conditions: 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. The *Gapdh* probe served as a control to normalize the data. The gene

expression experiments were performed in triplicate on three independent experiments and a melting analysis was performed at the end of the PCR run. To calculate the relative expression levels we used the 2−DDCT method [54,55].

5.7. Experimental study for renal functionality.

All experiments were conducted on age and gender-matched animals. Mice were housed individually in metabolic cages for 5 days at 23°C with a 12h dark/light cycle. Food and water were offered ad libitum. After 4 days of adjustment, physiological parameters were evaluated on day 5. 24h urine output was collected under mineral oil to prevent evaporation. Urinary osmolality was measured by Osmometer 3320 (Advanced Instrument, Inc).

5.8. Intraperitoneal Glucose Tolerance Test (IPGTT).

All experiments were conducted on age and gender-matched animals. Mice were housed individually in cages at 23°C with a 12h dark/light cycle and were fasted for 18 hours. After fasting, mice were weighed and received an intraperitoneal injection of glucose (2mg/g of body weight). Blood glucose levels were measured before glucose injection and after 15, 30, 60 and 120 minutes after glucose injection by using the OGCare glucometer (Biochemical Systems International S.r.l., Italy).

5.9. Serum analyses.

Blood samples from mice at different ages (4-6 months, 7-12 months) were collected and centrifuged for 5 minutes at 5000 rpm. Serum supernatants were recovered and analysed by SCIL VITROVET; in particular, triglycerides (mg/dl), total cholesterol (mg/dl), urea (mg/dl), calcium (mg/dl) and phosphorus (mg/dl) were measured.

5.10. Statistics.

Graphpad Prism 5 software was used to perform statistical analysis. Values are presented as mean ± SD. P-values were determined using a two-tail unpaired t-test. $P < 0.05$ was used as a threshold for statistical discernibility.

Table 3

GENE NAME	FORWARD PRIMER 5'-3'	REVERSE PRIMER 5'-3'
Cxcr4	gtaaccaccacggctgtaga	agtagatggtgggcaggaag
Gapdh	cggagtcaacggatttggtcgtat	gaagatggtgatgggcttcc
Gata2	agctcatgactatggcagca	ccggttctgtccattcatct
Nanog	aaccagtggttgaagactagcaatggtc	ttccagatgcgttcaccagatagc
NeuroD	gctccagggttatgagatcg	ctctgcattcatggcttcaa
Oct4	ccgtgtgaggtggagtctggagac	cgccggttacagaaccatactcg
Foxa2	ctgggagccgtgaagatggaag	tccagcgcccacataggatg
Gata4	tcaaaccagaaaacggaagc	ctgctgtgcccatagtgaga
Gata6	caccatcacccgacctactc	gcatgcattgcacaggtaat
Sox9	ggtctgcctggactgtatgtggatg	ctgtccgatgtctctctgcaggag
Hnf6	caaagaggtggcgcagcgtatc	gctctttccgtttgcaggctg
Nkx6.1	gacagcaaatcttcgccctgg	ttctccgaagtcccccttgagcc
Rex1	cagaagaaagcaggatcgcctc	gccacttgtctttgccgttttc
Nepn Iso B	gatgcagttgtgattctatgc	tagactctgttccactggc
Nepn Iso A	agatccttctctgcaatgtgaac	tcttccagagccaacagtcc
Nepn_b FR	tgtgattctatgcagtctgtaca	ggattctgctgtggctaatg
Nepn Crispr F3-R2	aaaagattctgttatcagactccg	cgagggtcctctggatatggaagc
Nepn Crispr Del	tcatggaactcccatcaggc	tagactctgttccactggc
Nepn Crispr Ko	ctcatggaattccttccacc	tagactctgttccactggc
Cds-Neo	gggatctcatgctggagttcttcg	agaattccccttcacatctgcctcc
Bax1b-5'	tggtgagacatttgtagatt	agaattccccttcacatctgcctcc
lacZ	gggatctcatgctggagttcttcg	agaattccccttcacatctgcctcc
lacZ-(Cre)	cgtcagtatcggcggaattcc	tccccttcacatctgcctcc
Int1	gaaactttttcacagtatgg	tccccttcacatctgcctcc

Table3: RT-PCR and qRT-PCR primers (5'-3').

Table 4

GENES	Denaturation (temp./duration)	Cyclic denaturation, annealing and elongation (temp./duration)	No. of cycles	Final elongation and cooling (temp./duration)
Nepn-202 Nepn-201 Gapdh (Fig. 1b)	95°C/5 minutes	95°C/30 seconds 58°C/20 seconds 72°C/45 seconds	35	72°C/5 minutes 4°C/∞
Nepn Isoform a Nepn Isoform b Gapdh (Fig. 2b, 3d, 6b)	95°C/5 minutes	95°C/15 seconds 55°C/15 seconds 72°C/20 seconds	38	72°C/5 minutes 4°C/∞
Recombinant allele Wild-type allele (Fig. 3b)	95°C/5 minutes	95°C/15 seconds 60°C/30 seconds 72°C/45 seconds	10 (-1°C/cycles)	72°C/5 minutes 4°C/∞
		95°C/15 seconds 50°C/30 seconds 72°C/45 seconds	30	72°C/5 minutes 4°C/∞
Nepn Isoform b ⇌ Nepn Isoform b ⇌ (Fig. 6d)	95°C/5 minutes	95°C/30 seconds 58°C/15 seconds 72°C/15 seconds	38	72°C/5 minutes 4°C/∞

Table4: PCR condition used in the study.

6.References

1   Duncan, M. D. & Wilkes, D. S. Transplant-related immunosuppression: a review of immunosuppression and pulmonary infections. *Proceedings of the American Thoracic Society* **2**, 449-455, doi:10.1513/pats.200507-073JS (2005).

2   Tagliaferri, D. *et al.* Retinoic Acid Induces Embryonic Stem Cells (ESCs) Transition to 2 Cell-Like State Through a Coordinated Expression of Dux and Duxbl1. *Frontiers in cell and developmental biology* **7**, 385, doi:10.3389/fcell.2019.00385 (2019).

3   Troiano, A. *et al.* ZSCAN4(+) mouse embryonic stem cells have an oxidative and flexible metabolic profile. *EMBO reports* **21**, e48942, doi:10.15252/embr.201948942 (2020).

4   Napolitano, G. *et al.* A novel member of Prame family, Gm12794c, counteracts retinoic acid differentiation through the methyltransferase activity of PRC2. *Cell death and differentiation* **27**, 345-362, doi:10.1038/s41418-019-0359-9 (2020).

5   Blum, B. & Benvenisty, N. The tumorigenicity of human embryonic stem cells. *Advances in cancer research* **100**, 133-158, doi:10.1016/S0065-230X(08)00005-5 (2008).

6   Fus-Kujawa, A. *et al.* Potential of Induced Pluripotent Stem Cells for Use in Gene Therapy: History, Molecular Bases, and Medical Perspectives. *Biomolecules* **11**, doi:10.3390/biom11050699 (2021).

7   Chmielowiec, J. & Borowiak, M. In vitro differentiation and expansion of human pluripotent stem cell-derived pancreatic progenitors. *The review of diabetic studies : RDS* **11**, 19-34, doi:10.1900/RDS.2014.11.19 (2014).

8   Leung, C. Y. & Zernicka-Goetz, M. Mapping the journey from totipotency to lineage specification in the mouse embryo. *Current opinion in genetics & development* **34**, 71-76, doi:10.1016/j.gde.2015.08.002 (2015).

9   Watson, C. M. & Tam, P. P. Cell lineage determination in the mouse. *Cell structure and function* **26**, 123-129, doi:10.1247/csf.26.123 (2001).

10   Davidson, K. C., Mason, E. A. & Pera, M. F. The pluripotent state in mouse and human. *Development* **142**, 3090-3099, doi:10.1242/dev.116061 (2015).

11   Tam, P. P. & Beddington, R. S. Establishment and organization of germ layers in the gastrulating mouse embryo. *Ciba Foundation symposium* **165**, 27-41; discussion 42-29, doi:10.1002/9780470514221.ch3 (1992).

12   Zorn, A. M. & Wells, J. M. Molecular basis of vertebrate endoderm development. *International review of cytology* **259**, 49-111, doi:10.1016/S0074-7696(06)59002-3 (2007).

13   Kiecker, C., Bates, T. & Bell, E. Molecular specification of germ layers in vertebrate embryos. *Cellular and molecular life sciences : CMLS* **73**, 923-947, doi:10.1007/s00018-015-2092-y (2016).

14   Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annual review of cell and developmental biology* **25**, 221-251, doi:10.1146/annurev.cellbio.042308.113344 (2009).

15   Shih, H. P., Wang, A. & Sander, M. Pancreas organogenesis: from lineage determination to morphogenesis. *Annual review of cell and developmental biology* **29**, 81-105, doi:10.1146/annurev-cellbio-101512-122405 (2013).

16   Pan, F. C. & Wright, C. Pancreas organogenesis: from bud to plexus to gland. *Developmental dynamics : an official publication of the American Association of Anatomists* **240**, 530-565, doi:10.1002/dvdy.22584 (2011).

17   Shen, M. M. Nodal signaling: developmental roles and regulation. *Development* **134**, 1023-1034, doi:10.1242/dev.000166 (2007).

18   Aoki, T. O. *et al.* Molecular integration of casanova in the Nodal signalling pathway controlling endoderm formation. *Development* **129**, 275-286 (2002).

19   Ben-Haim, N. *et al.* The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. *Developmental cell* **11**, 313-323, doi:10.1016/j.devcel.2006.07.005 (2006).

20   Hagos, E. G. & Dougan, S. T. Time-dependent patterning of the mesoderm and endoderm by Nodal signals in zebrafish. *BMC developmental biology* **7**, 22, doi:10.1186/1471-213X-7-22 (2007).

21   Stainier, D. Y. A glimpse into the molecular entrails of endoderm formation. *Genes & development* **16**, 893-907, doi:10.1101/gad.974902 (2002).

22   Dufort, D., Schwartz, L., Harpal, K. & Rossant, J. The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. *Development* **125**, 3015-3025 (1998).

23   Kanai-Azuma, M. *et al.* Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129**, 2367-2379 (2002).

24   Hart, A. H. *et al.* Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. *Development* **129**, 3597-3608 (2002).

25   Sinner, D., Rankin, S., Lee, M. & Zorn, A. M. Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. *Development* **131**, 3069-3080, doi:10.1242/dev.01176 (2004).

26   Hebrok, M., Kim, S. K. & Melton, D. A. Notochord repression of endodermal Sonic hedgehog permits pancreas development. *Genes & development* **12**, 1705-1713, doi:10.1101/gad.12.11.1705 (1998).

27   Kim, S. K. & Melton, D. A. Pancreas development is promoted by cyclopamine, a hedgehog signaling inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13036-13041, doi:10.1073/pnas.95.22.13036 (1998).

28   Deutsch, G., Jung, J., Zheng, M., Lora, J. & Zaret, K. S. A bipotential precursor population for pancreas and liver within the embryonic endoderm. *Development* **128**, 871-881 (2001).

29   Rossi, J. M., Dunn, N. R., Hogan, B. L. & Zaret, K. S. Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes & development* **15**, 1998-2009, doi:10.1101/gad.904601 (2001).

30   Pagliuca, F. W. & Melton, D. A. How to make a functional beta-cell. *Development* **140**, 2472-2483, doi:10.1242/dev.093187 (2013).

31   Schaffer, A. E., Freude, K. K., Nelson, S. B. & Sander, M. Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. *Developmental cell* **18**, 1022-1029, doi:10.1016/j.devcel.2010.05.015 (2010).

32   Kopp, J. L. *et al.* Progenitor cell domains in the developing and adult pancreas. *Cell cycle* **10**, 1921-1927, doi:10.4161/cc.10.12.16010 (2011).

33   Zhou, Q. *et al.* A multipotent progenitor domain guides pancreatic organogenesis. *Developmental cell* **13**, 103-114, doi:10.1016/j.devcel.2007.06.001 (2007).

34   Kopp, J. L. *et al.* Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. *Development* **138**, 653-665, doi:10.1242/dev.056499 (2011).

35   De Angelis, M. T. *et al.* Novel pancreas organogenesis markers refine the pancreatic differentiation roadmap of embryonic stem cells. *Stem cell reviews and reports* **10**, 269-279, doi:10.1007/s12015-013-9489-5 (2014).

36   Hou, J. *et al.* A regulatory network controls nephrocan expression and midgut patterning. *Development* **141**, 3772-3781, doi:10.1242/dev.108274 (2014).

37   Iozzo, R. V. Matrix proteoglycans: from molecular design to cellular function. *Annual review of biochemistry* **67**, 609-652, doi:10.1146/annurev.biochem.67.1.609 (1998).

38   Mochida, Y. *et al.* Nephrocan, a novel member of the small leucine-rich repeat protein family, is an inhibitor of transforming growth factor-beta signaling. *The Journal of biological chemistry* **281**, 36044-36051, doi:10.1074/jbc.M604787200 (2006).

39   Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355, doi:10.1038/nbt.2842 (2014).

40   Kumar, T. R., Larson, M., Wang, H., McDermott, J. & Bronshteyn, I. Transgenic mouse technology: principles and methods. *Methods Mol Biol* **590**, 335-362, doi:10.1007/978-1-60327-378-7_22 (2009).

41   Tang, S. H., Silva, F. J., Tsark, W. M. & Mann, J. R. A Cre/loxP-deleter transgenic line in mouse strain 129S1/SvImJ. *Genesis* **32**, 199-202 (2002).

42   Ayala, J. E. *et al.* Standard operating procedures for describing and performing metabolic tests of glucose homeostasis in mice. *Dis Model Mech* **3**, 525-534, doi:10.1242/dmm.006239 (2010).

43   Nauck, M. A. & Meier, J. J. Incretin hormones: Their role in health and disease. *Diabetes Obes Metab* **20 Suppl 1**, 5-21, doi:10.1111/dom.13129 (2018).

44   De Angelis, M. T. *et al.* Novel pancreas organogenesis markers refine the pancreatic differentiation roadmap of embryonic stem cells. *Stem Cell Rev* **10**, 269-279, doi:10.1007/s12015-013-9489-5 (2014).

45   Hildebrand, A. *et al.* Interaction of the small interstitial proteoglycans biglycan, decorin and fibromodulin with transforming growth factor beta. *Biochem J* **302 (Pt 2)**, 527-534, doi:10.1042/bj3020527 (1994).

46   Takeuchi, Y., Kodama, Y. & Matsumoto, T. Bone matrix decorin binds transforming growth factor-beta and enhances its bioactivity. *J Biol Chem* **269**, 32634-32638 (1994).

47   Qin, P., Haberbusch, J. M., Zhang, Z., Soprano, K. J. & Soprano, D. R. Pre-B cell leukemia transcription factor (PBX) proteins are important mediators for retinoic acid-dependent endodermal and neuronal differentiation of mouse embryonal carcinoma P19 cells. *J Biol Chem* **279**, 16263-16271, doi:10.1074/jbc.M313938200 (2004).

48   Tasheva, E. S., Klocke, B. & Conrad, G. W. Analysis of transcriptional regulation of the small leucine rich proteoglycans. *Mol Vis* **10**, 758-772 (2004).

49   Chia, C. Y. *et al.* GATA6 Cooperates with EOMES/SMAD2/3 to Deploy the Gene Regulatory Network Governing Human Definitive Endoderm and Pancreas Formation. *Stem Cell Reports* **12**, 57-70, doi:10.1016/j.stemcr.2018.12.003 (2019).

50   Zhou, P., He, A. & Pu, W. T. Regulation of GATA4 transcriptional activity in cardiovascular development and disease. *Curr Top Dev Biol* **100**, 143-169, doi:10.1016/B978-0-12-387786-4.00005-1 (2012).

51   Morrisey, E. E. *et al.* GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes Dev* **12**, 3579-3590, doi:10.1101/gad.12.22.3579 (1998).

52   Xuan, S. & Sussel, L. GATA4 and GATA6 regulate pancreatic endoderm identity through inhibition of hedgehog signaling. *Development* **143**, 780-786, doi:10.1242/dev.127217 (2016).

53   El Azhar, Y. & Sonnen, K. F. Development in a Dish-In Vitro Models of Mammalian Embryonic Development. *Frontiers in cell and developmental biology* **9**, 655993, doi:10.3389/fcell.2021.655993 (2021).

54   Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).

55   Lucci, V., Di Palma, T. & Zannini, M. Neuropilin-2 Is a Newly Identified Target of PAX8 in Thyroid Cells. *PLoS One* **10**, e0128315, doi:10.1371/journal.pone.0128315 (2015).