



HAL
open science

Dynamics of the Transcriptional Landscape During Human Fetal Testis and Ovary Development

Estelle Lecluze, Antoine D. Rolland, Panagiotis Filis, Bertrand Evrard, Sabrina Leverrier-Penna, Millissia Ben Maamar, Isabelle Coiffec, Vincent Lavoué, Paul A Fowler, Séverine Mazaud-Guittot, et al.

► **To cite this version:**

Estelle Lecluze, Antoine D. Rolland, Panagiotis Filis, Bertrand Evrard, Sabrina Leverrier-Penna, et al.. Dynamics of the Transcriptional Landscape During Human Fetal Testis and Ovary Development. Human Reproduction, 2020, 35 (5), pp.1099-1119. 10.1093/humrep/deaa041 . hal-02634151

HAL Id: hal-02634151

<https://ehesp.hal.science/hal-02634151>

Submitted on 22 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Dynamics of the transcriptional landscape during human fetal testis and ovary development**

2

3 **Running title:** Human fetal gonad transcriptomes

4

5 Estelle Lecluze¹, Antoine D. Rolland¹, Panagiotis Filis², Bertrand Evrard¹, Sabrina Leverrier-
6 Penna^{1,3}, Millissia Ben Maamar¹, Isabelle Coiffec¹, Vincent Lavoué⁴, Paul A. Fowler², Séverine
7 Mazaud-Guittot¹, Bernard Jégou¹, Frédéric Chalmel^{1,*}

8

9 ¹ Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) -
10 UMR_S 1085, F-35000 Rennes, France.

11 ² Institute of Medical Sciences, School of Medicine, Medical Sciences & Nutrition, University of
12 Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK.

13 ³ Univ Poitiers, STIM, CNRS ERL7003, Poitiers Cedex 9, France.

14 ⁴ CHU Rennes, Service Gynécologie et Obstétrique, F-35000 Rennes, France.

15

16 * To whom correspondence should be addressed

17 **Correspondence:** frederic.chalmel@inserm.fr.

18

19 **Abstract**

20 **STUDY QUESTION:** Which transcriptional program triggers sex differentiation in bipotential
21 gonads and downstream cellular events governing fetal testis and ovary development in humans?

22 **SUMMARY ANSWER:** The characterisation of a dynamically-regulated protein-coding and
23 noncoding transcriptional landscape in developing human gonads of both sexes highlights a large
24 number of potential key regulators that show an early sexually dimorphic expression pattern.

25 **WHAT IS KNOWN ALREADY:** Gonadal sex differentiation is orchestrated by a sexually
26 dimorphic gene expression program in XX and XY developing fetal gonads. A comprehensive
27 characterisation of its noncoding counterpart offers promising perspectives for deciphering the
28 molecular events underpinning gonad development and for a complete understanding of the
29 aetiology of disorders of sex development in humans.

30 **STUDY DESIGN, SIZE, DURATION:** To further investigate the protein-coding and noncoding
31 transcriptional landscape during gonad differentiation, we used RNA-sequencing (RNA-seq) and
32 characterised the RNA content of human fetal testis (N=24) and ovaries (N=24) from 6 to 17
33 postconceptional week (PCW), a key period in sex determination and gonad development.

34 **PARTICIPANTS/MATERIALS, SETTING, METHODS:** First trimester fetuses (6-12 PCW) and
35 second trimester fetuses (13-14 and 17 PCW) were obtained from legally-induced normally-
36 progressing terminations of pregnancy. Total RNA was extracted from whole human fetal gonads
37 and sequenced as paired-end 2x50 base reads. Resulting sequences were mapped to the human
38 genome, allowing for the assembly and quantification of corresponding transcripts.

39 **MAIN RESULTS AND THE ROLE OF CHANCE:** This RNA-seq analysis of human fetal testes
40 and ovaries at seven key developmental stages led to the reconstruction of 22,080 transcripts
41 differentially expressed during testicular and/or ovarian development. In addition to 8,935 transcripts

42 displaying sex-independent differential expression during gonad development, the comparison of
43 testes and ovaries enabled the discrimination of 13,145 transcripts that show a sexually dimorphic
44 expression profile. The latter include 1,479 transcripts differentially expressed as early as 6 PCW,
45 including 39 transcription factors, 40 long noncoding RNAs and 20 novel genes. Despite the use of
46 stringent filtration criteria (expression cut-off of at least 1 fragment per kilobase of exon model per
47 million reads mapped, fold-change of at least 2 and false discovery rate adjusted p-values of less than
48 $< 1\%$) the possibility of assembly artefacts and of false-positive differentially expressed transcripts
49 cannot be fully ruled out.

50 **LARGE SCALE DATA:** Raw data files (fastq) and a searchable table (.xlsx) containing information
51 on genomic features and expression data for all refined transcripts have been submitted to the NCBI
52 GEO under accession number GSE116278.

53 **LIMITATIONS, REASONS FOR CAUTION:** The intrinsic nature of this bulk analysis, i.e. the
54 sequencing of transcripts from whole gonads, does not allow direct identification of the cellular
55 origin(s) of the transcripts characterised. Potential cellular dilution effects (e.g. as a result of distinct
56 proliferation rates in XX and XY gonads) may account for a few of the expression profiles identified
57 as being sexually dimorphic. Finally, transcriptome alterations that would result from exposure to
58 pre-abortive drugs cannot be completely excluded. Although we demonstrated the high quality of the
59 sorted cell populations used for experimental validations using quantitative RT-PCR, it cannot be
60 totally excluded that some germline expression may correspond to cell contamination by, for
61 example, macrophages.

62 **WIDER IMPLICATIONS OF THE FINDINGS:** For the first time, this study has led to the
63 identification of a thousand of protein-coding and noncoding candidate genes showing an early,
64 sexually dimorphic, expression pattern that have not previously been associated with sex
65 differentiation. Collectively, these results increase our understanding of gonad development in

66 humans, and contribute significantly to the identification of new candidate genes involved in fetal
67 gonad differentiation. The results also provide a unique resource that may improve our understanding
68 of the fetal origin of testicular and ovarian dysgenesis syndromes, including cryptorchidism and
69 testicular cancers.

70 **STUDY FUNDING/COMPETING INTEREST(S):** This work was supported by the French
71 National Institute of Health and Medical Research (Inserm), the University of Rennes 1, the French
72 School of Public Health (EHESP), the Swiss National Science Foundation [SNF n° CRS115_171007
73 to B.J.], the French National Research Agency [ANR n° 16-CE14-0017-02 and n°18-CE14-0038-02
74 to F.C], the Medical Research Council [MR/L010011/1 to PAF] and the European Community's
75 Seventh Framework Programme (FP7/2007-2013) [under grant agreement no 212885 to PAF]] and
76 from the European Union's Horizon 2020 Research and Innovation Programme [under grant
77 agreement no 825100 to PAF and SMG]. There are no competing interests related to this study.

78

79

80 **Keywords:** Human gonad development; fetal testis; fetal ovary; sex differentiation; disorders of sex
81 development; transcriptional profiling; novel unannotated transcripts; long noncoding RNAs; bulk
82 RNA-sequencing; proteomics informed by transcriptomics.

83

84 **Introduction**

85 Mammalian ovary and testis development is a unique process compared to other organ development,
86 both developing from a bipotential organ which commits to a different fate following gonadal sex
87 determination and ending up as entirely different organs. This decisive turning point guides the
88 gonad toward one of the two developmental pathways, both including specific cell type development
89 and proliferation. Proper differentiation of these cell lineages determines the reproductive health of
90 the future being. In humans, the gonadal primordium arises from the thickening of the coelomic
91 epithelium at the surface of the mesonephros around 4th postconceptional week (PCW, corresponding
92 to 6 weeks of gestation/amenorrhea), and contains several precursor cell types, notably precursors of
93 supporting and steroidogenic cell lineages, as well as primordial germ cells (PGC) (Wilhelm *et al.*,
94 2013). The expression of the Y-linked transcription factor sex-determining region Y (SRY) during
95 the 6th PCW in supporting cell precursors of the genital ridge triggers the expression of the SRY-box
96 transcription factor 9 (SOX9) transcription factor which subsequently promotes a highly orchestrated
97 gene expression program (Koopman *et al.*, 1991; Vidal *et al.*, 2001; Sekido and Lovell-Badge, 2008;
98 Li *et al.*, 2014; Rahmoun *et al.*, 2017). These molecular events activate the commitment of Sertoli
99 cells, leading to testis cord formation, the appearance of a fetal Leydig cell from 7 PCW onwards,
100 and production of male hormones (androgens, insulin-like 3 protein and anti-Müllerian hormone)
101 that are essential for embryo masculinization. In the absence of SRY the R-spondin 1 (RSPO1)/Wnt
102 family member 4 (WNT4)/ β -catenin pathway and forkhead box L2 (FOXL2) induce another
103 complex cascade of transcriptional events, giving rise to fetal ovaries marked by the differentiation
104 of pre-granulosa cells and the commitment of germ cells into meiosis from 10 PCW onwards (Vainio
105 *et al.*, 1999; Schmidt, 2004; Uda *et al.*, 2004; Ottolenghi *et al.*, 2007; Chassot *et al.*, 2008; Liu *et al.*,
106 2009; Le Bouffant *et al.*, 2010; Childs *et al.*, 2011). To ensure the appropriate commitment to a given
107 fate, the sex-specific pathways antagonize each other to repress the alternative fate (Kim and Capel,

108 2006; Chang *et al.*, 2008; Maatouk *et al.*, 2008; Wilhelm *et al.*, 2009; Kashimada *et al.*, 2011;
109 Jameson *et al.*, 2012a; Greenfield, 2015; Bagheri-Fam *et al.*, 2017). Sex differentiation therefore
110 stems from a critical moment triggering a complex transcriptional landscape that governs gonad
111 specification and organogenesis. Both male and female expression programs are highly dynamic and
112 complex, but many blank areas remain in the map of our understanding, thus preventing full
113 understanding of most disorders of sex development (DSDs) (Eggers *et al.*, 2016). In particular, the
114 noncoding counterpart of the transcriptome is likely to play critical roles in the physiology and the
115 pathophysiology of the human developing gonads (Wu *et al.*, 2016).

116 Many dedicated studies have investigated *in situ* the gonadal expression pattern of specific human
117 genes and/or proteins known to play important roles in sex determination or gonad development
118 (Hanley *et al.*, 1999, 2000; de Santa Barbara *et al.*, 2001; Ostrer *et al.*, 2007; Mamsen *et al.*, 2017).
119 In addition, several valuable genome-wide expression studies have used microarray technology to
120 analyze whole fetal gonads (Small *et al.*, 2005; Fowler *et al.*, 2009; Houmard *et al.*, 2009; Rolland *et*
121 *al.*, 2011; Munger *et al.*, 2013; del Valle *et al.*, 2017; Mamsen *et al.*, 2017) or isolated fetal gonadal
122 cell populations (Nef *et al.*, 2005; Beverdam and Koopman, 2006; Bouma *et al.*, 2007, 2010) in
123 several mammalian species, including human. Recently single-cell transcriptomic analyses of
124 thousands of cells have opened up a new window onto gonadal cell lineages during fetal life (Guo *et*
125 *al.*, 2015, 2017; Li *et al.*, 2017; Stévant *et al.*, 2018, 2019). Although the latest technologies hold
126 great potential for further investigations, they are currently limited in their ability to study alternative
127 splicing, decipher the noncoding expression program or discover novel genes (Haque *et al.*, 2017).
128 Alternatively, “bulk” RNA-sequencing (RNA-seq) can be used to circumvent these issues
129 (Gkountela *et al.*, 2015). Several landmark studies have performed RNA-seq in the mouse to
130 investigate the transcriptome of the fetal Leydig cells (McClelland *et al.*, 2015; Inoue *et al.*, 2016),
131 the regulome of the key gonadal transcription factor SOX9 (Rahmoun *et al.*, 2017), and the

132 transcriptome of the developing fetal testes and ovaries at key developmental stages (Zhao *et al.*,
133 2018). Despite the limited availability of normal fetal human gonads, RNA-seq analyses of gonad
134 development in humans represent a very precious resource for our understanding.

135 In the present study, we performed a strand-specific, ribo-depleted RNA sequencing approach to
136 unravel the protein-coding and noncoding transcriptional landscape of human developing fetal testes
137 and ovaries from 6 to 17 PCW. Selected time-points were chosen to encompass the time window
138 from early transcription of the SRY gene in the male to differentiation and primary sex-specific
139 control of the cell lineages in both gonads. This analysis allowed us to identify a complex sexually
140 and non-sexually dimorphic expression program driving gonad development, with a focus on non-
141 coding genes as well as new unannotated genes that had not previously been described. In particular
142 we have highlighted a core set of transcripts showing a sex-biased expression after the onset of SRY
143 expression within the whole gonad during the 6th PCW that likely plays a critical role in fetal gonad
144 differentiation, and potentially for sex determination in humans. Our study significantly expands
145 knowledge of human gonadogenesis and provides a rich source of data for geneticists and clinicians
146 working in the field of DSDs.

147 **Materials and Methods**

148 Ethical considerations and sample collection

149 *First trimester fetuses*

150 Human fetuses (8-14 GW) were obtained from legally-induced normally-progressing terminations of
151 pregnancy performed in Rennes University Hospital. Tissues were collected with women's written
152 informed consent, in accordance with the legal procedure agreed by the National Agency for
153 Biomedical research (#PFS09-011) and the approval of the Local ethics committee of Rennes
154 Hospital (# 11-48). The termination of pregnancy was induced using a standard combined
155 mifepristone and misoprostol protocol, followed by aspiration. Gestational age was determined by
156 ultrasound, and further confirmed by measurement of foot length for mathematical estimation of fetal
157 development (Evtouchenko *et al.*, 1996; O'Shaughnessy *et al.*, 2019). The gonads were recovered
158 and dissected free of mesonephros in ice-cold phosphate-buffered saline (PBS) using a binocular
159 microscope (Olympus SZX7, Lille, France). The sex of the gonad was determined by morphological
160 criteria, except for fetuses younger than 7 PCW, for which a PCR was performed on genomic DNA
161 using primers specific for SRY (ACAGTAAAGGCAACGTCCAG;
162 ATCTGCGGGAAGCAAAGTGC) (Friel *et al.*, 2002) as well as for amelogenin X-linked (AMELX)
163 and amelogenin Y-linked (AMELY) (CTGATGGTTGGCCTCAAGCCTGTG;
164 GTGATGGTTGGCCTCAAGCCTGTG) (Akane *et al.*, 1992).

165

166 *Second trimester fetuses*

167 Human fetuses (13-14 and 17 PCW) were obtained from pregnant women after legally induced
168 abortions at the Aberdeen Pregnancy Counselling Service. The collection of fetal material was
169 approved by the National Health Service (NHS) Grampian Research Ethics Committees (REC

170 04/S0802/21). In all cases, women seeking elective terminations of pregnancy were recruited with
171 full written, informed consent by nurses working independently of the study at the Aberdeen
172 Pregnancy Counselling Service. Maternal data and medications used were recorded. Only fetuses
173 from normally progressing pregnancies (determined at ultrasound scan prior to termination) from
174 women over 16 years of age were collected following termination induced by a standard combined
175 mifepristone and misoprostol protocol, as detailed previously (O'Shaughnessy *et al.*, 2007; Fowler *et*
176 *al.*, 2008). Fetuses were transported to the laboratory within 30 min of delivery, weighed, sexed and
177 the crown-rump length recorded. Fetal tissues were snap-frozen in liquid nitrogen and stored at -
178 80°C.

179

180 RNA extraction, library construction and RNA-sequencing

181 Total RNA was extracted from human fetal gonads using the RNeasy mini Kit (Qiagen, Hilden,
182 Germany), quantified using a NanoDrop™ 8000 spectrophotometer (Thermo Fisher Scientific,
183 Waltham, MA, USA) and quality controlled using a 2100 Electrophoresis Bioanalyzer (Agilent
184 Technologies, Santa Clara, CA, USA). Only RNA extracts with a high quality RNA integrity number
185 (average value = 9.8 - ranging from 8.6 to 10) were included. Libraries suitable for strand-specific
186 high throughput DNA sequencing were then constructed, essentially as previously described (Jégou
187 *et al.*, 2017; Rolland *et al.*, 2019) using “TruSeq Stranded Total RNA with Ribo-Zero Gold Prep Kit”
188 (catalog # RS-122-2301, Illumina Inc., San Diego, CA, USA). The libraries were finally loaded in
189 the flow cell at 7 pM concentration and clusters were generated in the Cbot and sequenced in the
190 Illumina HiSeq 2500 as paired-end 2x50 base reads following Illumina's instructions. Image analysis
191 and base calling were performed using RTA 1.17.20 and CASAVA 1.8.2.

192

193 Read mapping, transcript assembly and quantification with the Tuxedo suite

194 *Assembly of a unique set of human reference transcripts*

195 Ensembl (Yates *et al.*, 2016) and RefSeq (Pruitt *et al.*, 2014; Brown *et al.*, 2015) transcript
196 annotations (GTF format) of human genome (release hg19) were downloaded from the UCSC (Speir
197 *et al.*, 2016) in June 2015, and were merged with Cuffcompare (Pollier *et al.*, 2013). This non-
198 redundant annotation was then used as the human reference transcripts (HRT) as previously
199 published (Chalmel *et al.*, 2014). A non-redundant dataset of human splice junctions (HSJ) was also
200 extracted from alignments of human transcripts and expressed sequence tags on the human genome
201 as provided by UCSC.

202 *Read mapping*

203 Reads from each individual sample were aligned to the hg19 release of the human genome with
204 TopHat2 (version 2.0.12) (Trapnell *et al.*, 2012) using previously published approaches (Pauli *et al.*,
205 2012; Trapnell *et al.*, 2012; Chalmel *et al.*, 2014; Zimmermann *et al.*, 2015). Briefly, TopHat2
206 program was first run for each fastq file, using HRT and HSJ datasets to improve read mapping.
207 Exonic junction outputs were then merged and added to the HSJ set. TopHat2 was next run a second
208 time using the new HSJ dataset to produce a final alignment file (BAM format) for each. Finally,
209 BAM files corresponding to the same experimental condition (fetal testes or ovaries, at a given time-
210 point) were merged and sorted with the Samtools suite (version 2.19.0) (Li *et al.*, 2009).

211 *Transcriptome assembly and quantification*

212 The Cufflinks suite (version 2.2.1, default settings) (Pollier *et al.*, 2013) was used to assemble
213 transcript fragments (or transfrags) for each experimental condition based on merged BAM files.
214 Resulting assembled transcripts were further merged into a non-redundant set of transfrags which
215 were further compared to the HRT dataset with the Cuffcompare program. Finally, Cuffquant was

216 used to estimate abundance of each transcript in each individual sample as fragments per kilobase of
217 exon model per million reads mapped (FPKM), and the normalization of the expression values across
218 samples was performed by Cuffnorm.

219 *Refinement of assembled transcripts*

220 A four-step strategy was used to filter out dubious transcripts as previously described (Prensner *et*
221 *al.*, 2011; Chalmel *et al.*, 2014) (Fig. S1). First, based on the Cuffcompare comparison, we only
222 selected transfrags defined as complete match ('='), potentially novel isoform ('j'), falling entirely
223 within a reference intron ('i') or an intergenic region ('u'), or showing exonic overlap with reference
224 on the opposite strand ('x'). Next, assembled transcripts of less than 200 nucleotides or that were
225 undetectable (<1 FPKM) were discarded. Finally, all transfrags that were annotated as either novel
226 isoforms or novel genes (class codes "j", "i", "u" or "x") and that did not harbor at least two exons
227 (multi-exonic) were filtered out.

228 *Principal component analysis*

229 Principal component analysis (PCA) was performed based on the expression values of refined
230 transcripts with the FactoMineR package to graphically evaluate the distribution of sequenced
231 samples (Lê *et al.*, 2008).

232

233 Coding potential analysis of assembled transcripts

234 DNA sequences of the refined transcripts were extracted with TopHat's gffread tool. As
235 recommended in (Chocu *et al.*, 2014; Zimmermann *et al.*, 2015), the protein-coding potential of each
236 transcript was estimated with an empirical integrative approach based on four predictive tools:
237 Coding-Potential Assessment Tool (CPAT, coding probability > 0.364) (Wang *et al.*, 2013),
238 HMMER (E-value < 10⁻⁴) (Finn *et al.*, 2011), Coding Potential Calculator (CPC, class "coding")

239 (Kong *et al.*, 2007) and txCdsPredict (score >800) (Kuhn *et al.*, 2013). Finally, we considered
240 assembled transcripts to likely encode for proteins if their nucleic sequences were considered as
241 protein-coding by at least two predictive tools.

242

243 Proteomics informed by transcriptomics strategy

244 As previously published in (Chocu *et al.*, 2014; Rolland *et al.*, 2019), we made use of a Proteomics
245 Informed by Transcriptomics (PIT) approach (Evans *et al.*, 2012) to provide evidence at the protein
246 level for assembled transcripts. First, we assembled a customized non-redundant protein database by
247 merging the UniProt (Pundir *et al.*, 2015) and the Ensembl (Yates *et al.*, 2016)(downloaded 2015/10)
248 proteome databases together with the set of predicted proteins derived from the assembled
249 transcripts. Briefly, the refined transcript sequences were translated into the three-first open reading
250 frames with the EMBOSS's Transeq program (Rice *et al.*, 2000) and only the amino acid sequences
251 of at least 20 residues were selected.

252 We next made use of the human fetal gonad MS/MS proteomics datasets available from the Human
253 Proteome Map (Kim *et al.*, 2014). First, 131 raw data files (corresponding to three fetal ovary
254 samples and two fetal testis samples) were downloaded from the PRIDE database (accession number
255 PXD000561) (Vizcaíno *et al.*, 2016) and converted into mgf format with ProteoWizard (Adusumilli
256 and Mallick, 2017). Subsequent analyses were performed with SearchGUI (Vaudel *et al.*, 2011)
257 (version 3.2.20) and PeptideShaker (Vaudel *et al.*, 2015)(version 1.16.8). A concatenated
258 target/decoy database was created from the enriched reference proteome with SearchGUI. Cross-
259 peptide identification was then performed with X!Tandem, Open Mass Spectrometry Search
260 Algorithm (OMSSA), and MSGF+ tools, with these following parameters: precursor ion tolerance
261 units set at 10 ppm; fragment tolerance set at 0.05 Da; carbamidomethylation of cysteine defined as a

262 fixed modification; oxidation of methionine and acetylation of protein N-term defined as a variable
263 modification; only tryptic peptides with up to two missed cleavages; and minimum peptide length set
264 to six amino acids. All peptides with at least one validated peptide spectrum match (PSM) and a
265 confidence score greater than 80% with PeptideShaker, were kept for further analyses. Finally, only
266 identifications with a false discovery rate (FDR) <1% were selected.

267

268 Statistical transcript filtration

269 The set of sexually dimorphic transcripts (SDTs) was defined by filtering transfrags that exhibited a
270 ≥ 2 -fold difference between the male and female gonads (using median expression values of sample
271 replicates) in at least one of the seven developmental stages. A Linear Models for Microarray Data
272 (LIMMA) statistical test was then used to identify transcripts displaying significant changes between
273 male and female gonads (F-value adjusted with a FDR $\leq 5\%$) (Smyth, 2004). Among SDTs, those
274 showing a differential expression as early as 6 PCW were designated as the set of early SDTs (or
275 early-SDTs). In addition, the set of developmental regulated transcripts (DRTs) were also selected by
276 isolating transfrags that exhibited a ≥ 2 -fold difference between two developmental stages during
277 either male or female gonad development. Similar to the selection of SDTs, a F-value adjusted with a
278 FDR $\leq 5\%$ was then used to identify candidates showing significant variations across
279 developmental stages. Finally, the difference between the sets of DRTs and SDTs allowed us to
280 discriminate non-sexually dimorphic transcripts (NSDTs) corresponding to the set of transfrags
281 showing a developmentally regulated expression pattern across fetal gonad development, but no
282 significant differential expression between male and female gonads.

283

284 Cluster and functional analyses

285 The resulting SDTs and NSDTs were then clustered into fourteen (named P1-14) and six (Q1-6)
286 expression patterns with the unsupervised hierarchical clustering on principle components (HCPC)
287 algorithm, respectively (Lê *et al.*, 2008). These clusters were then ordered according to peak
288 expression levels across developmental stages in testes first and then ovaries. Gene Ontology (GO)
289 term enrichments were estimated with the Fisher exact probability, using a Gaussian hypergeometric
290 test implemented in the Annotation Mappin Expression and Network suite (AMEN) (Chalmel and
291 Primig, 2008). A GO term was considered significantly associated with a given expression pattern if
292 the FDR-corrected P value was $\leq 5\%$ and the number of genes bearing this annotation was ≥ 5 .

293

294 Transcription factors and their related target genes

295 To get an insight into potentially important regulators that might be involved in early human gonad
296 development or sex determination, transcriptional factors and their targets were extracted from public
297 databases, the Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining
298 database (TRRUST) (Han *et al.*, 2015) and Transcription Factor encyclopedia (Yusuf *et al.*, 2012).

299

300 FACS sorting and quantitative PCR validation

301 *Single cell dissociation and cell sorting*

302 Single cell suspensions were obtained from gonads by a standard enzymatic and mechanic digestion
303 procedure. Gonads were cut into small pieces and digested in 0.25% Trypsin-0.02% EDTA (#T4049,
304 Sigma-Aldrich) and 0.05 mg/ml DNase (#DN25, Sigma-Aldrich, Missouri, USA) for 10 min at
305 37 °C. Trypsin digestion was stopped by adding 10% fetal bovine serum in M199 media and samples

306 were centrifuged at 350 g for 5 min at 37°C. Dispersed cells were resuspended in PBS and counted
307 on a Malassez hemocytometer after labeling of dead cells with Trypan blue. For testes, a plasma
308 membrane labelling of cord cells was performed using a mouse FITC-conjugated anti-human
309 epithelial antigen (clone Ber EP4; Dako # F0860, diluted 1:100) associated with a labelling of germ
310 cells with a mouse R-Phycoerythrin-coupled anti-human KIT proto-oncogene, receptor tyrosine
311 kinase (KIT/CD117) (clone 104D2, BioLegend, San Diego, CA, USA, # 313204, diluted 1:100) for
312 30 min at room temperature. For ovaries, germ cells were labeled with the mouse R-Phycoerythrin-
313 coupled anti-human KIT/CD117 as described above. Cells were sorted by a flow cytometer cell
314 sorter FACSARIAII (BD Biosciences, New Jersey, USA) equipped with Diva software. Cells were
315 collected in PBS, centrifuged at 500g for 45 min at 4°C and pelleted cells were stored at -80°C until
316 RNA extraction.

317

318 *Quantitative RT-PCR*

319 RNA was extracted from cell pellets with PicoPure RNA Isolation Kit (Thermo Fisher Scientific)
320 according to manufacturer's instructions. Total RNAs (100 ng) were reverse transcribed with iScript
321 cDNA synthesis kit (Biorad, Hercules, CA, USA) and quantitative PCR was performed using the
322 iTaq© universal SYBR green supermix (Biorad) according to manufacturer's instructions in a
323 Cfx384 OneTouch Real-Time PCR system (Biorad). The following amplification program was used:
324 an initial denaturation of 3 min at 95°C, 40 cycles of 10 sec denaturation at 95°C and 30 sec at 62°C
325 for annealing and extension. Dissociation curves were produced using a thermal melting profile
326 performed after the last PCR cycle. Primer pairs flanking introns were designed in order to avoid
327 amplification of contaminating genomic DNA whenever possible and they were aligned onto human
328 RefSeq transcripts using primer-BLAST to check for their specificity. Furthermore, only those
329 primers that produced a single peak during the melting curve step (i.e. with peak temperature

330 variance of no more than 0.5°C) were considered and their efficiency was evaluated using serial
331 dilutions of cDNA templates (Table I). Ribosomal protein lateral stalk subunit P0 (RPLP0) and
332 ribosomal protein S20 (RPS20) mRNA, used as internal controls for normalization purposes, were
333 initially validated in human adult testis gonad (Svingen *et al.*, 2014) and further used for human fetal
334 gonads (Jørgensen *et al.*, 2018). Results were calculated with Bio-Rad CFX Manager 3.1 using the
335 $\Delta\Delta$ CT method as n-fold differences in target gene expression, relative to the reference gene and
336 calibrator sample, which comprises an equal mixture of all the tested samples for a given organ.

337

338 Immunohistochemistry and immunofluorescence

339 Upon collection, additional gonads (seven testes and five ovaries) were fixed either in Bouins fluid
340 fixative or paraformaldehyde 4% (w/v) for 1 to 2 hours, embedded in paraffin using standard
341 procedures and cut into 5 μ m-thick sections. After dewaxing and rehydration, slides were treated for
342 antigen retrieval with pre-heated 10 mM citrate buffer, pH 6.0 at 80°C for 40 min before cooling at
343 room temperature . Sections were blocked for 1 h at room temperature with 4% bovine serum
344 albumin in PBS before the overnight incubation at 4°C with the primary antibody diluted in Dako
345 antibody diluent (Dako Cytomation, Trappes, France). Antibodies and conditions are described in
346 Table II. Secondary antibodies were goat anti-rabbit biotinylated antibody (E0432, Dako, diluted
347 1:500); or rabbit anti-mouse biotinylated antibody (E0464, Dako, diluted 1:500). Sections were
348 developed with streptavidin-horseradish peroxidase (Vectastain ABC kit, Vector Laboratories,
349 Burlingame, CA, USA) and 3,3'-diaminobenzidine tetrahydrochloride (Vector Laboratories Inc.) and
350 counterstained with hematoxylin. Stained sections were examined and photographed under light
351 microscopy (Olympus BX51). For cryosectioning, additional paraformaldehyde-fixed gonads were
352 cryopreserved in PBS-sucrose 20%, embedded in NEG50TM (Allan-Richard Scientific, Thermo

353 Fisher Scientific) and cut in 8 μm -thick sections. Thawed sections were treated for antigen retrieval
354 with citrate buffer, as described, when necessary (for WT1 transcription factor, WT1; neurexin 3,
355 NRXN3; contactin 1, CNTN1; and lin-28 homolog A, LIN28), rinsed in PBS and incubated
356 overnight at 4°C with primary antibody. Rinsed sections were incubated with the ad hoc fluorescent
357 secondary antibodies (1:500). The second primary antibody was subsequently incubated overnight at
358 4°C followed by the corresponding secondary antibody. Secondary antibodies were either 488 or 594
359 Alexa Fluor conjugated antibodies made in chicken for rabbit- and mouse-hosted primary antibodies,
360 and in donkey for Contactin 1 (CNTN1) primary antibody (Invitrogen). Sections were mounted in
361 prolong Gold anti-fade reagent with DAPI (Invitrogen, Carlsbad, CA, USA; Thermo Fisher
362 Scientific). Slides were examined and photographed with an AxioImager microscope equipped with
363 an AxioCam MRc5 camera and the ZEN software (Zeiss, Le Pecq, France).

364

365 **Results**

366 Expression profiling of fetal gonads identifies more than 300 new genes in the human genome

367 To investigate the expression program governing gonad differentiation in humans, we performed a
368 RNA-seq analysis on fetal testes ($n=24$) and ovaries ($n=24$), covering seven developmental stages
369 from 6 to 17 PCW (i.e. from 8 to 19 gestational weeks) (Fig. 1A). Following read mapping and
370 transcript reconstruction (Supplementary Table SI), a stringent refinement strategy selected a “high-
371 confidence” set of 35,194 transcripts expressed in human fetal gonads (Supplementary Fig. S1). A
372 PCA of these expression data provided a first hint on the biological relevance of our dataset: The first
373 three components indeed appeared strongly correlated with the developmental stage (dimensions 1
374 and 3) and the genetic sex (dimension 2) of samples (19.7% and 15.3% of variance, respectively)
375 (Fig. 1B and C). A hierarchical clustering based on the 35 first PCA dimensions explaining 90% of
376 the total variance of the data further confirms the reliable distribution of the samples according to the
377 sex and the developmental stage (Fig. 1D).

378 The comparison of reconstructed transcripts with RefSeq and Ensembl reference annotations
379 identified known (13,673; 38.9%) and novel (18,718; 53.2%) isoforms of annotated protein-coding
380 genes as well as known (529, 1.5%) and novel (680, 1.9%) isoforms of annotated long noncoding
381 RNAs (lncRNAs). Importantly this comparison also identified 318 novel unannotated transcripts
382 (NUTs) corresponding to new intronic (81, 0.2%), intergenic (164, 0.5%) or antisense (73, 0.2%) as
383 yet uncharacterized genes in the human genome. The comparison of genomic and expression features
384 highlighted significant differences between mRNAs, lncRNAs and NUTs (Supplementary Fig. S2).
385 As expected, lncRNAs are expressed at lower levels than mRNAs, are more specifically expressed
386 during single stages of development, and have lower sequence conservation, length, number of exons
387 and GC content. Interestingly, this “non-coding” trend is exacerbated by NUTs, as they have a lower

388 abundance, conservation, length, number of exons and GC content as well as a more restricted
389 expression than lncRNAs and mRNAs (Supplementary Fig. S2).

390 We next combined results from a coding potential analysis and from a proteomics informed by
391 transcriptomics approach (PIT) (Evans *et al.*, 2012) to characterize the protein-encoding potential of
392 novel isoforms and loci identified above. As expected, 98.5% of mRNAs (known and novel
393 isoforms) displayed a high protein-encoding potential, and 56% of them were supported at the
394 protein level thanks to at least one identified peptide (Fig. 1E). On the other hand, only 48.1% of
395 lncRNAs (known and novel isoforms) were predicted to have a high protein-encoding potential
396 (PEP), and as little as 8.5% were identified during the PIT analysis, usually through a single peptide
397 identification (Fig. 1E). When NUTs were finally evaluated, they mostly displayed features similar
398 to lncRNAs: 76.1% of them were predicted to have a low protein-encoding potential while only 0.4%
399 were confirmed at the protein level (Fig. 1E). Besides, 94 lncRNAs (including 53 with two or more
400 peptides identified) and one NUT were both predicted to have a high protein-encoding potential and
401 demonstrated at the protein level.

402

403 Transcriptome dynamics during human gonad development define sexually and non-sexually
404 dimorphic expression programs

405 We next focused on genes with dynamic expression patterns during gonad development. Several
406 steps of statistical filtration led to the selection of 13,145 transcripts (7,633 genes) that were
407 differentially-expressed between testes and ovaries (called SDT), most of which (10,521 transcripts,
408 6,587 genes) were also developmentally-regulated (Fig. 1F). Notably, 8,935 transcripts (5,961 genes)
409 were differentially-expressed during gonad development but did not show significant difference
410 between sexes (NSDT).

411 We further classified SDT into 14 clusters (termed P1-P14) according to their preferential expression
412 pattern (Fig. 2A and B). Patterns P1-P7 include transcripts showing peak expression in fetal testes at
413 6 to 7 PCW (P1; *SRY*; doublesex and mab-3 related transcription factor 1, *DMRT1*), at 7 PCW (P2;
414 *SOX9*, desert hedgehog signalling molecule, *DHH*; inhibin subunit beta B, *INHBB*), at 9 to 12 PCW
415 (P4; cytochrome P450 family 11 subfamily A member 1, *CYP11A1*; cytochrome P450 family 17
416 subfamily A member 1, *CYP17A1*; steroidogenic acute regulatory protein, *STAR*; luteinizing
417 hormone/choriogonadotropin receptor, *LHCGR*; insulin like 3, *INSL3*), at 13 to 17 PCW (P6; actin
418 alpha 2, *ACTA2*; prostaglandin D2 synthase, *PTGDS*) or with a broader expression pattern
419 throughout testis development (P3; nuclear receptor subfamily 5 group A member 1, *NR5A1*; nuclear
420 receptor subfamily 0 group B member 1, *NR0B1*; and P5; WT1 transcription factor, *WT1*; claudin 11
421 *CLDN11*). Similarly, transcripts belonging to patterns P8-P14 are preferentially expressed in fetal
422 ovaries and display peak expression at 7 PCW (P8; *RPSO1*, anti-Mullerian hormone receptor type 2,
423 *AMHR2*; nanog homeobox, *NANOG*), at 7-9 PCW (P9; POU class 5 homeobox 1, *POU5F1*;
424 developmental pluripotency associated 2 and 4, *DPPA2/4*; lin-29 homolog A and B, *LIN28A/B*), at
425 12 PCW (P11; *FOXL2*, deleted in azoospermia like, *DAZL*; piwi like RNA-mediated gene silencing
426 2, *PIWIL2*), at 13-14 PCW (P12; MET proto-oncogene, receptor tyrosine kinase, *MET*; delta like
427 canonical Notch ligand 4, *DLL4*; X inactive specific transcript, *XIST*; vascular endothelial growth
428 factor A, *VEGFA*), at 17 PCW (P13; meiotic double-stranded break formation protein 1, *MEI1*;
429 meiosis specific with OB-fold, *MEIOB*; SPO11 initiator of meiotic double stranded breaks, *SPO11*;
430 synaptonemal complex protein 1 to 3, *SYCP1-3*; and P14; folliculogenesis specific bHLH
431 transcription factor, *FIGLA*; NOBOX oogenesis homeobox, *NOBOX*; spermatogenesis and
432 oogenesis specific basic helix-loop-helix 1, *SOHLH1*) or with a broader expression throughout
433 ovarian development (P10; DNA meiotic recombinase 1, *DMC1*; empty spiracles homeobox 2,
434 *EMX2*; lymphoid enhancer binding factor 1 *LEF1*; Wnt family member 2B, *WNT2B*). We also

435 evaluated the functional relevance of expression patterns by a GO term enrichment analysis (Fig.
436 2C). Several broad biological processes related to organogenesis and/or cell differentiation were
437 enriched in various testis-associated (P1, P2, P5, P6) and ovary-associated (P8) patterns. More
438 precisely, expression pattern P4 was found to be enriched in genes involved in steroidogenesis while
439 several processes associated with meiosis and female germ cell development were found to be
440 enriched in P13 and P14, which is consistent with the differentiation and development of Leydig
441 cells in fetal testes from 7 PCW onwards and with the commitment of ovarian germ cells into
442 meiosis from 12 PCW onwards, respectively. Finally, we investigated the distribution of RNA
443 biotypes and found that lncRNAs and NUTs were significantly enriched in expression patterns P12
444 to P14 (Fig. 2D).

445 The 8,935 NSDTs were also classified according to peak expression into six expression patterns
446 (termed Q1-Q6) and include transcripts expressed at early stages of gonad differentiation (6-7 PCW;
447 Q1 and Q2; including *WT1*, GATA binding protein 4, *GATA4*), following sexual differentiation (7-9
448 PCW; Q3 and Q4; *DMRT1*, SRY-box transcription factor 8, *SOX8*) or at later stages of gonad
449 development (12-17 PCW; Q5 and Q6; nuclear receptor subfamily 6 group A member 1, *NR6A1*)
450 (Fig. 3A-C). Finally, when investigating the distribution of RNA biotypes within co-expression
451 groups, we found that lncRNAs and NUTs were enriched in expression patterns Q1 and Q6 (Fig.
452 3D). While these transcripts are also likely to include factors with important roles during gonadal
453 differentiation and development, they were not further investigated in this study. All data, however,
454 are available through the ReproGenomics Viewer (RGV) genome browser (<http://rgv.genouest.org/>)
455 (Darde *et al.*, 2015, 2019) and are also available as a searchable table (.xlsx) containing information
456 on genomic features and expression data for all refined transcripts (submitted to the NCBI GEO
457 under accession number GSE116278).

458

459 A complex transcriptional program governing early gonadal differentiation

460 In order to highlight new candidate genes that could be involved in early gonadal differentiation, we
461 focused our analysis on 1,479 SDTs showing a significant differential expression in fetal gonads as
462 early as 6 PCW (Supplementary Fig. S3). Most of these early-SDTs (61.7%) logically belong to
463 early expression patterns P1, P2, P8 and P9. This set of genes is composed of important actors
464 including *SRY* (P1), *SOX9*, *DHH*, patched 1 (*PTCH1*) and cytochrome P450, family 26 subfamily b
465 polypeptide 1 (*CYP26B1*) (P2), LIM homeobox 9 (*LHX9*) (P6), activin A receptor type 1B
466 (*ACVR1B*) (P8), *AMHR2* (P9) or *FOXL2* (P11) which demonstrates the relevance of this filtration for
467 selecting important factors in sex differentiation (Supplementary Fig. S3B and S3C). In addition to
468 well-known transcription factors, such as *SRY*, *SOX9*, *LHX9*, or *FOXL2*, 174 early SDTs
469 correspond to 131 genes encoding transcriptional regulators that should also play a critical role in the
470 establishment of this complex sexually dimorphic expression program (Supplementary Table SII).
471 Although the proportion of early-SDTs in the distinct expression patterns P1-P14 according to their
472 coding status are generally similar to those of SDT (Supplementary Fig. S3D), it is important to note
473 that this set of candidates includes 40 lncRNAs and 20 NUTs which may be involved in early steps
474 of gonad differentiation.

475

476 Distinct cellular expression patterns of newly identified genes involved in human sex determination

477

478 To investigate further the cellular origin of selected candidates, we performed immunohistochemistry
479 experiments as well as quantitative PCR (qPCR) on FACS-sorted cells. The successful enrichment of
480 Sertoli cells (hEpA+/KIT-), germ cells (KIT+) and interstitial cells (hEpA-/KIT-) from 6–7 PCW
481 testes was notably validated by the expression of *SOX9*, *KIT* proto-oncogene, receptor tyrosine

482 kinase (*KIT*) and nuclear receptor subfamily 2 group F member 2 (*NR2F2*), respectively
483 (Supplementary Fig. S4A-C), while that of germ cells (*KIT*⁺) and somatic cells (*KIT*⁻) from 6–8
484 PCW and 10–12 PCW ovaries was confirmed by the high expression levels of *KIT*, *FOXL2* or
485 *NR2F2*, respectively (Supplementary Fig. S4D-F). We first investigated genes that display
486 expression profiles similar to that of *SRY* (P1; high expression and clear sexual dimorphism in 6
487 PCW testes), such as Wnt ligand secretion mediator (*WLS*), C-X-C motif chemokine ligand 14
488 (*CXCL14*) and C-C motif chemokine receptor 1 (*CCR1*) (Fig. 4A). While *WLS* was mainly expressed
489 in Sertoli cells it was also substantially expressed in interstitial cells, whereas *CXCL14* was only
490 expressed in Sertoli cells (Fig. 4A). In contrast, the expression of *CCR1* was mainly detected in germ
491 cells. We also analysed genes with an expression profile similar to *SOX9* (P2; clear sexual
492 dimorphism and peak of expression in 7 PCW testes), such as EPH receptor B1 (*EPHB1*), fetal and
493 adult testis expressed 1 (*FATE1*), MAGE family member B1 (*MAGEB1*), erb-b2 receptor tyrosine
494 kinase 3 (*ERBB3*), Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 1
495 (*CITED1*) and a NUT antisense to *CITED1* (TCONS_00249587) (Fig. 4B). We found that *EPHB1*,
496 *FATE1*, *MAGEB1* and *ERBB3* were indeed expressed in Sertoli cells and to a lesser extent in
497 interstitial cells, while *SRY*-box transcription factor 10 (*SOX10*) was expressed at similar levels in
498 Sertoli cells and interstitial cells (Fig. 4B). Interestingly, we found that both *CITED1* and its potential
499 antisense RNA were specifically and simultaneously expressed in Sertoli cells. Consistently with
500 qPCR results, *SOX10*, *EPHB1*, *MAGEB1*, and *FATE1* proteins were indeed all found in cord cells at
501 the histological level, but exhibited varying ratios of expression in Sertoli and germ cells (Fig. 4C).
502 For instance, *SOX9* was expressed only in Sertoli cells, whereas *MAGEB1* was clearly expressed in
503 germ cells as well.

504 We also investigated the cell distribution of genes preferentially expressed in ovary (P8-P9) (Fig.
505 5A-D). First, several of them show a higher differential expression as early as 6 PCW, and were

506 preferentially expressed in somatic cells, including neurexin 3 (*NRXN3*), contactin 1 (*CNTN1*) and
507 SET nuclear proto-oncogene (*SET*), or specifically in somatic cells such as the NUT
508 *TCONS_00153406* (Fig. 5A). In agreement, immunolabeling showed NRXN3 protein in the nucleus
509 of cells surrounding KIT⁺ germ cells, with a pattern very similar to that of WT1 (Fig. 5D).
510 Interestingly, CNTN1 protein was found in a subset of epithelial cells surrounding LIN28⁺ germ
511 cells in 6 PCW ovaries, or in ovarian cords adjacent to the mesonephric-gonadal junction (Fig. 5D)
512 but not in the surface epithelium. Several genes show higher levels of differential expression at later
513 stages in gonad development (Fig. 5B-C). Some of those genes, such as neuropeptide Y (*NPY*), SRY-
514 box transcription factor 4 (*SOX4*) and the novel transcript *TCONS_00224470*, were preferentially
515 expressed in somatic cells, as was *RSPO1* (Fig. 5B). In contrast, others displayed patterns typical of
516 germline-associated expression patterns, including *POU5F1* (a well-known germ cell marker) and
517 three NUTs, *TCONS_00113718*, *TCONS_00055038* and *TCONS_00042565*, which were highly
518 expressed in female germ cells from 7 to 12 PCW (Fig. 5C).

519 **Discussion**

520 Unravelling the molecular sequence of events involved in gonadogenesis and sex determination is
521 essential in order to understand DSDs. Although a significant number of studies have already
522 examined the sexually dimorphic expression program driving gonad development in animal models
523 (Beverdam and Koopman, 2006), its characterization remains elusive in humans. Three studies have
524 investigated the transcriptome of the developing gonads from 5.7 to 10 PCW (Mamsen *et al.*, 2017;
525 del Valle *et al.*, 2017) and of ovarian primordial follicle formation from 13 to 18 PCW (Fowler *et al.*,
526 2009) in humans. However, they were based on microarray technologies, thus restricting the gene set
527 studied and limiting the characterization of non-coding transcripts and the identification of new
528 genes. Our study is among the first to capitalize on the power of the “bulk” RNA-seq technology to
529 perform an in-depth characterization of the dynamic transcriptional landscape of whole human fetal
530 gonads, from early differentiation (i.e. 6 PCW) up to Leydig cell transition in the testis and primary
531 follicle formation in the ovary (i.e. 17 PCW), at both the protein-coding and non-coding levels. In
532 particular our results identify transcriptional regulators, lncRNAs and novel genes (NUTs) that show
533 an early sexually dimorphic expression pattern and could therefore play important regulatory roles
534 from sex determination onwards. Nevertheless, as in any model, including animals that are sacrificed
535 with anesthesia or CO₂ or in the case of spontaneous abortions where the development of the embryo
536 or foetus can be disturbed, it should be borne in mind that there is a small chance that some
537 transcriptional alterations might result from exposure to pre-abortive drugs. Collectively this work
538 constitutes a rich resource for the community by providing new information regarding the early
539 molecular events that could be involved in both normal sex differentiation and DSDs.

540 Our study confirms and complements previous findings accumulated in humans and other species
541 (Nef *et al.*, 2005; Beverdam and Koopman, 2006; Jameson *et al.*, 2012b; Zhao *et al.*, 2018; Planells
542 *et al.*, 2019). For instance our dataset validates the onset of *SRY* transcription prior to 6 PCW

543 (detected at 5.5 PCW in (Mamsen *et al.*, 2017)) but also demonstrates the over-expression of *SOX9*
544 in the testis as early as 6 PCW (previously reported at only 6.8 PCW in (Mamsen *et al.*, 2017)),
545 which may reflect the higher sensitivity of RNA-seq as compared to microarrays (Mantione *et al.*,
546 2014). Altogether our transcriptional profiling allowed us to identify over 33,000 transcripts
547 expressed in human developing fetal gonads, including mRNAs and lncRNAs as well as unknown
548 genes. Although our analysis was mostly focused on SDTs, a set of almost 9,000 transcripts showing
549 similar expression profiles in testes and ovaries (NSDTs) was also identified despite major cell
550 composition differences between the two gonad types, especially at later developmental stages.
551 These transcripts therefore represent valuable information on critical molecular factors underlying or
552 required for the development of both XX and XY gonads, including for instance WT1 and its
553 multiple isoforms required at different stages (Hastie, 2017).

554 We then focused our analysis on 1,479 SDTs showing sexual dimorphism as early as 6 PCW,
555 including more than 1,000 candidate genes that have not previously been associated with sex
556 differentiation. It is noteworthy that most of these early dimorphic profiles are likely to result from
557 true differential transcriptional regulation rather than from dilution effects, as the cell composition of
558 XX and XY fetal gonads at this stage remains highly analogous. To further highlight new promising
559 candidates that might be involved in the regulation of this complex expression program we next
560 focused on the 131 genes encoding transcription factors and showing an early SDT pattern
561 (Supplementary Table SII). Among them, the cAMP responsive element modulator (CREM), a well-
562 known regulator of gene expression programming of post-meiotic germ cells in the adult testis
563 (Hogeveen and Sassone-Corsi, 2006), is preferentially expressed in XY gonads at 6 PCW. We also
564 found that one of its target genes, the tachykinin precursor 1 (*TAC1*) (Qian *et al.*, 2001), is over-
565 expressed in the fetal ovary at this early developmental stage suggesting that CREM might
566 negatively regulate *TAC1* in the human fetal testis. While its role during sex determination remains

567 unknown, TAC1 encodes several neuropeptides belonging to the tachykinin family that are critical
568 for many biological processes (Dehlin and Levick, 2014; Sun and Bhatia, 2014; Sorby-Adams *et al.*,
569 2017). GLI family zing finger 1 (*GLII*) is also an interesting candidate as it is preferentially
570 expressed in the fetal testis as early as 6 PCW (Mamsen *et al.*, 2017) and encodes a transcription
571 factor known to regulate the expression of the secreted frizzled related protein 1 (*SFRP1*) (Kim *et al.*,
572 2010). Since *SFRP1* is critical for fetal testis development in the mouse (Warr *et al.*, 2009), acting
573 through its suppression of Wnt signalling (Kim *et al.*, 2010), this could suggest a potential important
574 role for both *GLII* and *SFRP1* during sex differentiation in humans.

575 Mamsen and collaborators deduced from their microarray experiment that the onset of
576 steroidogenesis in male gonads occurred at 7.5 PCW (Mamsen *et al.*, 2017). We also consistently
577 found that the expression of genes involved in steroidogenesis increased drastically in testes from 7
578 PCW onwards. However, several genes, such as *CYP17A1*, *CYP11A1*, hydroxy-delta-5-steroid
579 dehydrogenase, 3 beta- and steroid delta-isomerase 2 (*HSD3B2*) and hydroxysteroid 17-beta
580 dehydrogenase 3 (*HSD17B3*), actually exhibited sexual dimorphism as early as 6 PCW, suggestive of
581 an earlier induction of the molecular networks underlying Leydig cell differentiation. A more likely
582 explanation could be that even if they are expressed only in Leydig cells later in development, (pre-
583)Sertoli cells may also express such factors in early stages. This would be in line with the described
584 co-operation between these two cell types for the synthesis of androgens in the mouse fetal testis
585 (O'Shaughnessy *et al.*, 2000; Shima *et al.*, 2013) at 6 PCW. Among male-biased early-SDTs we also
586 identified expression patterns similar to that of *SRY*, such as for the Wnt ligand secretion mediator
587 (*WLS*) and for the C-C motif chemokine receptor 1 (*CCR1*). *WLS*, which we found to be
588 preferentially expressed in fetal Sertoli cells, is an important mediator of Wnt secretion (Bänziger *et al.*
589 *et al.*, 2006; Das *et al.*, 2012), suggesting a potential role in promoting sex determination. We found
590 *CCR1* to be preferentially expressed in germ cells. This was rather surprising since a sexually-

591 dimorphic expression pattern at such an early developmental stage (i.e. as early as 6PCW) is
592 expected to result from primary changes in expression in somatic cells as they commit to their male
593 or female fates. *CCR1* encodes a chemokine receptor thought to be implicated in stem cell niche
594 establishment and maintenance and it has already been shown to be expressed in postnatal gonocytes
595 in the mouse (Simon *et al.*, 2010). Although we demonstrated the high quality of the sorted cell
596 populations used in the current study, it cannot be totally excluded that the germline expression of
597 *CCR1* may indeed correspond to a contamination by KIT-expressing somatic cells such as
598 macrophages. Many other candidate genes display a *SOX9*-like expression pattern suggesting that
599 some of them could also be important for Sertoli cell differentiation. Among these candidate genes,
600 several are already known to be important for gonad development or fertility in humans and/or mice,
601 such as such as *FATE1*, *MAGEB1* or *SOX10*. Interestingly, while we found the expression pattern
602 of *SOX10* to be conserved between human and mouse (i.e. with strong preferential expression in
603 young fetal testes), that of *SOX8* was not. Instead, we found *SOX8* to be expressed in a similar
604 manner in human fetal testes and ovaries, with peak expression between 6 and 7 PWC followed by
605 subsequent downregulation. While this expression profile is clearly not incompatible with a role
606 during early testis differentiation as in the mouse (Schepers *et al.*, 2003), it also suggests a potential
607 broader involvement in development of both XX and XY gonads in humans. The role of other
608 candidates, such as the Erb-b2 receptor tyrosine kinase 3 (*ERBB3*) and the EPH receptor B1
609 (*EPHB1*), remains unknown during early gonad development. The expression of *ERBB3* has been
610 reported in mouse PGCs in the genital ridge suggesting that the ErbB signalling might contribute to
611 control of growth and survival of PGCs (Toyoda-Ohno *et al.*, 1999). The expression of *EPHB1* has
612 never been reported in the fetal testis but is involved in angiogenesis and neural development
613 (Pasquale, 2005).

614 Due to the limited number of known markers for distinct fetal ovarian somatic cells, the association
615 of female-biased expression patterns (P8-P14) with specific cell populations remains challenging at
616 the whole gonad level. We found that several PGC markers, such as *KIT*, *POU5F1*, *NANOG* or
617 *LIN28A*, are over-expressed in the ovary, compared with the testis, as early as 6 PCW. This is in line
618 with the fact that PGCs proliferate at a higher rate than somatic cells in the human fetal ovary
619 (Bendsen *et al.*, 2003; Lutterodt *et al.*, 2009; Mamsen *et al.*, 2010). Experimental investigations
620 allowed us to identify candidate genes associated with the ovarian somatic cell lineages, such as
621 SRY-box 4 (*SOX4*), SET nuclear proto-oncogene (*SET*), contactin 1 (*CNTN1*), neurexin 3 (*NRXN3*)
622 and neuropeptide Y (*NPY*). This set of genes holds great promise as potential key factors for female
623 sex determination and ovary differentiation. *SOX4* encodes a transcription factor with a high mobility
624 group box domain and its expression has already been described in supporting cells of the mouse
625 gonads, although without evident sexual dimorphism (Zhao *et al.*, 2017). *SET*, for which we
626 demonstrate a highly sexually dimorphic expression as early as 6 PCW, is implicated in
627 transcriptional regulation through epigenomic modifications and has been associated with polycystic
628 ovary syndrome (Jiang *et al.*, 2017). Other candidate genes expressed in somatic cells appear to be
629 implicated in neurogenesis, such as *NRXN3*, *NPY* and *CNTN1* (Sutton *et al.*, 1988; Markiewicz *et al.*,
630 2003; Bizzoca *et al.*, 2012; Harkin *et al.*, 2016). *NRXN3* is expressed at a very weak level in the
631 human fetal brain between 8 and 12 PCW (Harkin *et al.*, 2016) , i.e. 2 weeks after a high
632 transcriptional induction in the fetal ovary at 6 PCW, which may indicate an independent role of the
633 gene in both processes. *NPY* is already known to control female reproductive processes at the
634 hypothalamus level, and to have a direct action on ovarian cell proliferation and apoptosis in
635 prepubertal gilts (Sirotkin *et al.*, 2015). In contrast, *CNTN1* encodes a neuronal cell adhesion
636 molecule that has never been described in reproductive-related processes, but seems to be a key
637 factor in the development of many cancers (Chen *et al.*, 2018). The role of these three candidates in

638 female developing gonads remains unknown. All of the above mentioned male-biased (*WLS*, *CCR1*,
639 *ERBB3*, *EPHB1*, *CITED1* and *asCITED1*) and female-biased (*SOX4*, *SET*, *NRXN3* and *NPY*)
640 candidate genes would require further functional experiments to untangle their role during gonad
641 development.

642 One of the most original contributions of our study is to unravel the non-coding counterpart of the
643 fetal gonadal transcriptome. To the best of our knowledge, this is the first study to address this issue
644 in humans. As mentioned before, we assembled 1,209 lncRNAs and 318 NUTs expressed in
645 developing fetal gonads. The statistical comparison of their genomic features and a protein-encoding
646 analysis strongly suggest that the vast majority of NUTs corresponds to newly identified lncRNAs.
647 However, based on the PIT approach and the protein-encoding analyses, a small fraction (6.2%) of
648 the identified noncoding transcripts are good candidates for novel protein-coding genes as they were
649 confirmed at the protein level. Our RNA-seq analysis also contributed to the identification of 680
650 antisense lncRNAs, including one located on the opposite strand of the Cbp/p300 interacting
651 transactivator with Glu/Asp rich carboxy-terminal domain 1 (*CITED1*). Both sense and antisense
652 (*asCITED1*, *TCONS_00249587*) transcripts showed a preferential, highly correlated expression in
653 fetal Sertoli cells as early as 6 PCW. In the mouse, *Cited1* has been reported to be a potential target
654 of *Sry* (Li *et al.*, 2014) and is specifically expressed in the adult testis (Fagerberg *et al.*, 2014). Our
655 results suggest that *asCITED1* might contribute to the regulation of *CITED1*, and could therefore be
656 implicated in early testis development in humans. We also report an accumulation of lncRNAs in
657 expression patterns associated with female meiosis (P13-P14). This result is line with similar
658 observations that have been made in adult germ cells from meiosis onwards (Cabili *et al.*, 2011;
659 Laiho *et al.*, 2013; Chalmel *et al.*, 2014; Rolland *et al.*, 2019). This phenomenon thus seems to be
660 conserved in both male and female germ cells, which suggests that lncRNAs might also play critical
661 roles in human fetal oocytes. Furthermore, we observed that the vast majority of noncoding early-

662 SDTs (16/20 NUTs, and 28/40 lncRNAs) were preferentially expressed in fetal ovaries, which may
663 reflect a specific non-coding transcriptional program at play during early ovary development. Further
664 investigation allowed us to identify that early, female-biased NUTs were preferentially expressed in
665 germ cells (*TCONS_00042656*, *TCONS_00055038*, *TCONS_00113718*) although some were also
666 expressed in somatic cells (*TCONS_00153406* and *TCONS_00224470*). Additional functional
667 analysis will be essential to elucidate the role of these germline and somatic candidates in the
668 physiology of the fetal developing gonads and in the aetiology of DSDs. DSDs indeed comprise
669 heterogeneous conditions affecting the genital system, with a wide range of phenotypes. The
670 management of these disorders is globally improved by genetic diagnosis, as it leads to a more
671 accurate prognosis and prediction of the long-term outcome. Recently, recommendations from the
672 European Cooperation in Science and Technology state that genetic diagnosis should preferentially
673 use whole exome sequencing of a panel of candidate genes, while whole genome sequencing should
674 be restricted to suspected oligo- or poly-genic DSDs (Audí *et al.*, 2018). Nevertheless, the majority
675 of genetic testing remains inconclusive as most causative genes involved in DSDs have not yet been
676 identified (Alhomaidah *et al.*, 2017). Although many challenges remain to understand the
677 implications of lncRNAs during gonad development in humans, their functional roles in almost all
678 investigated biological systems are now supported by several studies (Cheng *et al.*, 2016; Tao *et al.*,
679 2016), including during gonad development (Rastetter *et al.*, 2015; Taylor *et al.*, 2015; Winge *et al.*,
680 2017) and for gonadal functions (Ohhata *et al.*, 2011; Bao *et al.*, 2013; Taylor *et al.*, 2015; Watanabe
681 *et al.*, 2015; Wen *et al.*, 2016; Hosono *et al.*, 2017; Wichman *et al.*, 2017; Jégu *et al.*, 2019).
682 Genome-wide association studies of patients with DSD would therefore greatly benefit from
683 screening for new causal genetic variants in lncRNAs expressed early in sex determination. In this
684 context our resource will assist geneticists to refine and complete the required panel of disease

685 candidate genes by including non-coding genes involved in testicular and ovarian dysgenesis
686 syndromes with a fetal origin, including cryptorchidism and testicular cancers.

687 Single-cell technologies now open new avenues for the genomic characterization of biological
688 systems, including the study of cellular heterogeneity. When compared to bulk approaches, single-
689 cell transcriptomics allows transcriptional signatures to be robustly assigned to specific cell types. In
690 this fast-evolving field, however, distinct available technologies have specific advantages and
691 limitations, and may be favoured depending on the scientific question (Baran-Gale *et al.*, 2018). For
692 instance, high-throughput systems that enable the analysis of several thousands of cells, including the
693 mature and popular droplet-based high-throughput system from 10x Genomics, are needed in order
694 to study discrete cell populations and/or to accurately reconstruct cell differentiation processes. On
695 the other hand these systems suffer from a relatively low sensitivity and specificity: they only capture
696 a partial fraction of the transcriptome of each individual cell (~2-4,000 genes per cell). Furthermore,
697 by focusing on either the 3' or the 5' extremity of RNA molecules, they do not allow the
698 reconstruction of transcript isoforms or the discovery of new genes. In the near future, increased
699 sensitivity of droplet-based methods, combined with long-read sequencing technologies, will provide
700 accurate transcriptome information at the isoform level and at a single-cell resolution (Byrne *et al.*,
701 2019), hopefully at an affordable price. In the meantime, bulk RNA-seq and current single-cell
702 technologies remain highly complementary. The current study will support and complement future
703 single-cell experiments aimed at reconstructing cell lineage progression in fetal gonads.

704 Overall, our study comprehensively describes the dynamic transcriptional landscape of the fetal
705 gonads at seven key developmental stages in humans. This work discovered extensive sexually and
706 non-sexually dimorphic expression changes, not only of protein-coding genes but also of lncRNAs
707 and novel genes that are triggered early during gonad differentiation. This rich resource significantly

708 extends existing state of the art knowledge and constitutes an invaluable reference atlas for the field
709 of reproductive sciences and sex determination in particular.

710

711

712 **Acknowledgments**

713 We thank all members of the SEQanswers forums for helpful advice; Steven Salzberg and Cole
714 Trapnell for continuous support with the “Tuxedo” suite; and the UCSC Genome team members.
715 Sequencing was performed by the GenomEast platform, a member of the ‘France Génomique’
716 consortium (ANR-10-INBS-0009). We thank Ms Linda Robertson, Ms Margaret Fraser, Ms
717 Samantha Flannigan (University of Aberdeen) and the staff at Grampian NHS Pregnancy
718 Counselling Service, and all the staff of the Department of Obstetrics and Gynecology of the Rennes
719 Sud Hospital for their expert assistance and help, and the participating women, without whom this
720 study would not have been possible. The authors are grateful for Ms Gersende Lacombe and Mr
721 Laurent Deleurme from the Biosit CytomeTri cytometry core facility of Rennes 1 University.

722 **Authors’ roles**

723 FC, ADR, SMG and BJ designed the study. FC, ADR and EL wrote the manuscript. FC and ADR
724 supervised the research. EL and FC prepared, analysed, and interpreted data. ADR, SMG, IC, MBM,
725 PF, PAF, SLP, and BJ prepared the samples and interpreted sequencing data. BE, ADR and SMG
726 validated expression data. SMG, PF, PAF and BJ contributed to the manuscript. All authors
727 approved the final version of the manuscript.

728

729 **Funding**

730 This work was supported by the French National Institute of Health and Medical Research (Inserm),
731 the University of Rennes 1, the French School of Public Health (EHESP), the Swiss National Science
732 Foundation [SNF n° CRS115_171007 to B.J.], the French National Research Agency [ANR n° 16-
733 CE14-0017-02 and n°18-CE14-0038-02 to F.C], the Medical Research Council [MR/L010011/1 to
734 PAF] and the European Community's Seventh Framework Programme (FP7/2007-2013) [under grant
735 agreement no 212885 to PAF] and from the European Union's Horizon 2020 Research and
736 Innovation Programme [under grant agreement no 825100 to PAF and SMG]. The authors have no
737 competing financial interests.

738

739 **Conflict of interest**

740 There are no competing interests related to this study.

741

742 **References**

- 743 Adusumilli R, Mallick P. Data Conversion with ProteoWizard msConvert. *Methods Mol Biol*
744 [Internet] 2017;**1550**:, p. 339–368.
- 745 Akane A, Seki S, Shiono H, Nakamura H, Hasegawa M, Kagawa M, Matsubara K, Nakahori Y,
746 Nagafuchi S, Nakagome Y. Sex determination of forensic samples by dual PCR amplification of
747 an X-Y homologous gene. *Forensic Sci Int* [Internet] 1992;**52**:143–148.
- 748 Alhomaidah D, McGowan R, Ahmed SF. The current state of diagnostic genetics for conditions
749 affecting sex development. *Clin Genet* [Internet] 2017;**91**:157–162. Blackwell Publishing Ltd.
- 750 Audí L, Ahmed SF, Krone N, Cools M, McElreavey K, Holterhus PM, Greenfield A, Bashamboo A,
751 Hiort O, Wudy SA, *et al.* GENETICS IN ENDOCRINOLOGY: Approaches to molecular
752 genetic diagnosis in the management of differences/disorders of sex development (DSD):
753 position paper of EU COST Action BM 1303 ‘DSDnet.’ *Eur J Endocrinol* [Internet]
754 2018;**179**:R197–R206.
- 755 Bagheri-Fam S, Bird AD, Zhao L, Ryan JM, Yong M, Wilhelm D, Koopman P, Eswarakumar VP,
756 Harley VR. Testis Determination Requires a Specific FGFR2 Isoform to Repress FOXL2.
757 *Endocrinology* [Internet] 2017;**158**:3832–3843. Oxford University Press.
- 758 Bänziger C, Soldini D, Schütt C, Zipperlen P, Hausmann G, Basler K. Wntless, a Conserved
759 Membrane Protein Dedicated to the Secretion of Wnt Proteins from Signaling Cells. *Cell*
760 [Internet] 2006;**125**:509–522.
- 761 Bao J, Wu J, Schuster AS, Hennig GW, Yan W. Expression profiling reveals developmentally
762 regulated lncRNA repertoire in the mouse male germline. *Biol Reprod* [Internet] 2013;**89**:107.
763 Society for the Study of Reproduction.
- 764 Baran-Gale J, Chandra T, Kirschner K. Experimental design for single-cell RNA sequencing. *Brief*
765 *Funct Genomics* [Internet] 2018;**17**:233–239. Oxford University Press.

- 766 Bendsen E, Byskov AG, Laursen SB, Larsen H-PE, Andersen CY, Westergaard LG. Number of
767 germ cells and somatic cells in human fetal testes during the first weeks after sex differentiation.
768 *Hum Reprod* [Internet] 2003;**18**:13–18.
- 769 Beverdam A, Koopman P. Expression profiling of purified mouse gonadal somatic cells during the
770 critical time window of sex determination reveals novel candidate genes for human sexual
771 dysgenesis syndromes. *Hum Mol Genet* [Internet] 2006;**15**:417–431. Oxford University Press.
- 772 Bizzoca A, Corsi P, Polizzi A, Pinto MF, Xenaki D, Furley AJW, Gennarini G. F3/Contactin acts as
773 a modulator of neurogenesis during cerebral cortex development. *Dev Biol* [Internet]
774 2012;**365**:133–151.
- 775 Bouffant R Le, Guerquin MJ, Duquenne C, Frydman N, Coffigny H, Rouiller-Fabre V, Frydman R,
776 Habert R, Livera G. Meiosis initiation in the human ovary requires intrinsic retinoic acid
777 synthesis. *Hum Reprod* [Internet] 2010;**25**:2579–2590.
- 778 Bouma GJ, Affourtit JJP, Bult CJ, Eicher EM. *Transcriptional profile of mouse pre-granulosa and*
779 *Sertoli cells isolated from early-differentiated fetal gonads* [Internet]. *Gene Expr Patterns*
780 [Internet] 2007;**7**:113–123.
- 781 Bouma GJ, Hudson QJ, Washburn LL, Eicher EM. New Candidate Genes Identified for Controlling
782 Mouse Gonadal Sex Determination and the Early Stages of Granulosa and Sertoli Cell
783 Differentiation1. *Biol Reprod* [Internet] 2010;**82**:380–389.
- 784 Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, Tolstoy I, Tatusova T, Pruitt KD,
785 Maglott DR, *et al.* Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*
786 [Internet] 2015;**43**:D36–D42.
- 787 Byrne A, Cole C, Volden R, Vollmers C. Realizing the potential of full-length transcriptome
788 sequencing. *Philos Trans R Soc Lond B Biol Sci* [Internet] 2019;**374**:20190097. Royal Society
789 Publishing.

- 790 Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation
791 of human large intergenic noncoding RNAs reveals global properties and specific subclasses.
792 *Genes Dev* [Internet] 2011;**25**:1915–1927. Cold Spring Harbor Laboratory Press.
- 793 Chalmel F, Lardenois a., Evrard B, Rolland a. D, Sallou O, Dumargne M-C, Coiffec I, Collin O,
794 Primig M, Jegou B. High-Resolution Profiling of Novel Transcribed Regions During Rat
795 Spermatogenesis. *Biol Reprod* [Internet] 2014;**91**:5–5.
- 796 Chalmel F, Primig M. The Annotation, Mapping, Expression and Network (AMEN) suite of tools for
797 molecular systems biology. *BMC Bioinformatics* [Internet] 2008;**9**:86. BioMed Central.
- 798 Chang H, Gao F, Guillou F, Taketo MM, Huff V, Behringer RR. Wt1 negatively regulates beta-
799 catenin signaling during testis development. *Development* [Internet] 2008;**135**:1875–1885.
- 800 Chassot A-A, Ranc F, Gregoire EP, Roepers-Gajadien HL, Taketo MM, Camerino G, Rooij DG de,
801 Schedl A, Chaboissier M-C. Activation of β -catenin signaling by Rspo1 controls differentiation
802 of the mammalian ovary. *Hum Mol Genet* [Internet] 2008;**17**:1264–1277.
- 803 Chen N, He S, Geng J, Song Z-J, Han P-H, Qin J, Zhao Z, Song Y-C, Wang H-X, Dang C-X.
804 Overexpression of Contactin 1 promotes growth, migration and invasion in Hs578T breast
805 cancer cells. *BMC Cell Biol* [Internet] 2018;**19**:5.
- 806 Cheng L, Ming H, Zhu M, Wen B. Long noncoding RNAs as Organizers of Nuclear Architecture.
807 *Sci China Life Sci* [Internet] 2016;**59**:236–244.
- 808 Childs AJ, Cowan G, Kinnell HL, Anderson RA, Saunders PTK. Retinoic Acid Signalling and the
809 Control of Meiotic Entry in the Human Fetal Gonad. In Clarke H, editor. *PLoS One* [Internet]
810 2011;**6**:e20249.
- 811 Chocu S, Evrard B, Lavigne R, Rolland AD, Aubry F, Jégou B, Chalmel F, Pineau C. Forty-Four
812 Novel Protein-Coding Loci Discovered Using a Proteomics Informed by Transcriptomics (PIT)
813 Approach in Rat Male Germ Cells¹. *Biol Reprod* [Internet] 2014;**91**:123–123.

- 814 Darde TA, Lecluze E, Lardenois A, Stévant I, Alary N, Tüttelmann F, Collin O, Nef S, Jégou B,
815 Rolland AD, *et al.* The ReproGenomics Viewer: a multi-omics and cross-species resource
816 compatible with single-cell studies for the reproductive science community. *Bioinformatics*
817 [Internet] 2019; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30668675>.
- 818 Darde TA, Sallou O, Becker E, Evrard B, Monjeaud C, Bras Y Le, Jégou B, Collin O, Rolland AD,
819 Chalmel F. The ReproGenomics Viewer: an integrative cross-species toolbox for the
820 reproductive science community. *Nucleic Acids Res* [Internet] 2015;**43**:W109-16. Oxford
821 University Press.
- 822 Das S, Yu S, Sakamori R, Stypulkowski E, Gao N. Wntless in Wnt secretion: molecular, cellular and
823 genetic aspects. *Front Biol (Beijing)* [Internet] 2012;**7**:587–593. NIH Public Access.
- 824 Dehlin HM, Levick SP. Substance P in heart failure: The good and the bad. *Int J Cardiol* [Internet]
825 2014;**170**:270–277.
- 826 Eggers S, Sadedin S, Bergen JA van den, Robevska G, Ohnesorg T, Hewitt J, Lambeth L, Bouty A,
827 Knarston IM, Tan TY, *et al.* Disorders of sex development: insights from targeted gene
828 sequencing of a large international patient cohort. *Genome Biol* [Internet] 2016;**17**:243.
- 829 Evans VC, Barker G, Heesom KJ, Fan J, Bessant C, Matthews DA. De novo derivation of proteomes
830 from transcriptomes for transcript and protein identification. *Nat Methods* [Internet]
831 2012;**9**:1207–1211.
- 832 Evtouchenko L, Studer L, Spencer C, Dreher E, Seiler RW. A mathematical model for the estimation
833 of human embryonic and fetal age. *Cell Transplant* [Internet] 1996;**5**:453–464.
- 834 Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M,
835 Tahmasebpour S, Danielsson A, Edlund K, *et al.* Analysis of the human tissue-specific
836 expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol*
837 *Cell Proteomics* [Internet] 2014;**13**:397–406.

- 838 Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching.
839 *Nucleic Acids Res* [Internet] 2011;**39**:W29-37.
- 840 Fowler PA, Cassie S, Rhind SM, Brewer MJ, Collinson JM, Lea RG, Baker PJ, Bhattacharya S,
841 O'Shaughnessy PJ. Maternal Smoking during Pregnancy Specifically Reduces Human Fetal
842 Desert Hedgehog Gene Expression during Testis Development. *J Clin Endocrinol Metab*
843 [Internet] 2008;**93**:619–626.
- 844 Fowler PA, Flannigan S, Mathers A, Gillanders K, Lea RG, Wood MJ, Maheshwari A, Bhattacharya
845 S, Collie-Duguid ESR, Baker PJPJ, *et al.* Gene expression analysis of human fetal ovarian
846 primordial follicle formation. *J Clin Endocrinol Metab* [Internet] 2009;**94**:1427–1435.
- 847 Friel A, Houghton JA, Glennon M, Lavery R, Smith T, Nolan A, Maher M. A preliminary report on
848 the implication of RT-PCR detection of DAZ, RBMY1, USP9Y and Protamine-2 mRNA in
849 testicular biopsy samples from azoospermic men. *Int J Androl* [Internet] 2002;**25**:59–64.
- 850 Gkountela S, Zhang KXX, Shafiq TAA, Liao W-WW, Hargan-Calvopiña J, Chen P-YY, Clark
851 ATT, Hargan-Calvopiña J, Chen P-YY, Clark ATT, *et al.* DNA demethylation dynamics in the
852 human prenatal germline. *Cell* [Internet] 2015;**161**:1425–1436.
- 853 Greenfield A. Understanding sex determination in the mouse: genetics, epigenetics and the story of
854 mutual antagonisms. *J Genet* [Internet] 2015;**94**:585–590.
- 855 Guo F, Yan L, Guo H, Li L, Hu B, Zhao Y, Yong J, Hu Y, Wang X, Wei Y, *et al.* The transcriptome
856 and DNA methylome landscapes of human primordial germ cells. *Cell* [Internet]
857 2015;**161**:1437–1452. Elsevier Inc.
- 858 Guo H, Hu B, Yan L, Yong J, Wu Y, Gao Y, Guo F, Hou Y, Fan X, Dong J, *et al.* DNA methylation
859 and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res* [Internet]
860 2017;**27**:165–183. Nature Publishing Group.
- 861 Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim HH, Cho A, Kim E, Lee T, *et al.* TRRUST: a

- 862 reference database of human transcriptional regulatory interactions. *Sci Rep* [Internet]
863 2015;**5**:11432. Nature Publishing Group.
- 864 Hanley N., Hagan D., Clement-Jones M, Ball S. S, Strachan T, Salas-Cortés L, McElreavey K,
865 Lindsay S, Robson S, Bullen P, *et al.* SRY, SOX9, and DAX1 expression patterns during
866 human sex determination and gonadal development. *Mech Dev* [Internet] 2000;**91**:403–407.
- 867 Hanley NA, Ball SG, Clement-Jones M, Hagan DM, Strachan T, Lindsay S, Robson S, Ostrer H,
868 Parker KL, Wilson DI. Expression of steroidogenic factor 1 and Wilms' tumour 1 during early
869 human gonadal development and sex determination. *Mech Dev* 1999;**87**:175–180.
- 870 Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for
871 biomedical research and clinical applications. *Genome Med* [Internet] 2017;**9**:75.
- 872 Harkin LF, Lindsay SJ, Xu Y, Alzu'bi A, Ferrara A, Gullon EA, James OG, Clowry GJ. Neurexins
873 1–3 Each Have a Distinct Pattern of Expression in the Early Developing Human Cerebral
874 Cortex. *Cereb Cortex* [Internet] 2016;**27**:4497–4505. Oxford University Press.
- 875 Hastie ND. Wilms' tumour 1 (WT1) in development, homeostasis and disease. *Development*
876 [Internet] 2017;**144**:2862–2872.
- 877 Hogeveen KN, Sassone-Corsi P. Regulation of gene expression in post-meiotic male germ cells:
878 CREM-signalling pathways and male fertility. *Hum Fertil (Camb)* [Internet] 2006;**9**:73–79.
- 879 Hosono Y, Niknafs YS, Prensner JR, Iyer MK, Dhanasekaran SM, Mehra R, Pitchiaya S, Tien J,
880 Escara-Wilke J, Poliakov A, *et al.* Oncogenic Role of THOR, a Conserved Cancer/Testis Long
881 Non-coding RNA. *Cell* [Internet] 2017;**171**:1559-1572.e20.
- 882 Houmard B, Small C, Yang L, Naluai-Cecchini T, Cheng E, Hassold T, Griswold M. Global Gene
883 Expression in the Human Fetal Testis and Ovary. *Biol Reprod* [Internet]
884 2009;**443**:biolreprod.108.075747.
- 885 Inoue M, Shima Y, Miyabayashi K, Tokunaga K, Sato T, Baba T, Ohkawa Y, Akiyama H, Suyama

- 886 M, Morohashi K. Isolation and Characterization of Fetal Leydig Progenitor Cells of Male Mice.
887 *Endocrinology* [Internet] 2016;**157**:1222–1233.
- 888 Jameson SA, Lin Y-T, Capel B. Testis development requires the repression of Wnt4 by Fgf
889 signaling. *Dev Biol* [Internet] 2012a;**370**:24–32.
- 890 Jameson SA, Natarajan A, Cool J, DeFalco T, Maatouk DM, Mork L, Munger SC, Capel B.
891 Temporal Transcriptional Profiling of Somatic and Germ Cells Reveals Biased Lineage Priming
892 of Sexual Fate in the Fetal Mouse Gonad. In Barsh GS, editor. *PLoS Genet* [Internet]
893 2012b;**8**:e1002575. Public Library of Science.
- 894 Jégou B, Sankararaman S, Rolland AD, Reich D, Chalmel F. Meiotic Genes Are Enriched in
895 Regions of Reduced Archaic Ancestry. *Mol Biol Evol* [Internet] 2017;**34**:1974–1980.
- 896 Jégu T, Blum R, Cochrane JC, Yang L, Wang C-Y, Gilles M-E, Colognori D, Szanto A, Marr SK,
897 Kingston RE, *et al.* Xist RNA antagonizes the SWI/SNF chromatin remodeler BRG1 on the
898 inactive X chromosome. *Nat Struct Mol Biol* [Internet] 2019;**26**:96.
- 899 Jiang S-W, Xu S, Chen H, Liu X, Tang Z, Cui Y, Liu J. Pathologic significance of SET/I2PP2A-
900 mediated PP2A and non-PP2A pathways in polycystic ovary syndrome (PCOS). *Clin Chim Acta*
901 [Internet] 2017;**464**:155–159.
- 902 Jørgensen A, Macdonald J, Nielsen JE, Kilcoyne KR, Perlman S, Lundvall L, Langhoff Thuesen L,
903 Juul Hare K, Frederiksen H, Andersson AM, *et al.* Nodal Signaling Regulates Germ Cell
904 Development and Establishment of Seminiferous Cords in the Human Fetal Testis. *Cell Rep*
905 2018;**25**:1924-1937.e4. Elsevier B.V.
- 906 Kashimada K, Pelosi E, Chen H, Schlessinger D, Wilhelm D, Koopman P. FOXL2 and BMP2 act
907 cooperatively to regulate follistatin gene expression during ovarian development.
908 *Endocrinology* [Internet] 2011;**152**:272–280.
- 909 Kim J-H, Shin HS, Lee SH, Lee I, Lee YSYC, Park JC, Kim YJ, Chung JB, Lee YSYC. Contrasting

- 910 activity of Hedgehog and Wnt pathways according to gastric cancer cell differentiation:
911 relevance of crosstalk mechanisms. *Cancer Sci* [Internet] 2010;**101**:328–335.
- 912 Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS,
913 Isserlin R, Jain S, *et al.* A draft map of the human proteome. *Nature* [Internet] 2014;**509**:575–
914 581.
- 915 Kim Y, Capel B. Balancing the bipotential gonad between alternative organ fates: A new perspective
916 on an old problem. *Dev Dyn* [Internet] 2006;**235**:2292–2300.
- 917 Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G. CPC: assess the protein-coding
918 potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*
919 [Internet] 2007;**35**:W345-9.
- 920 Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. Male development of
921 chromosomally female mice transgenic for Sry. *Nature* [Internet] 1991;**351**:117–121.
- 922 Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform*
923 [Internet] 2013;**14**:144–161.
- 924 Laiho A, Kotaja N, Gyenesei A, Sironen A. Transcriptome profiling of the murine testis during the
925 first wave of spermatogenesis. *PLoS One* [Internet] 2013;**8**:e61558. Public Library of Science.
- 926 Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. *J Stat Softw*
927 [Internet] 2008;**25**:1–18.
- 928 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000
929 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and
930 SAMtools. *Bioinformatics* [Internet] 2009;**25**:2078–2079.
- 931 Li L, Dong J, Yan L, Yong J, Liu X, Hu Y, Fan X, Wu X, Guo H, Wang X, *et al.* Single-Cell RNA-
932 Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions.
933 *Cell Stem Cell* [Internet] 2017;**20**:891–892.

- 934 Li Y, Zheng M, Lau Y-FC. *The Sex-Determining Factors SRY and SOX9 Regulate Similar Target*
935 *Genes and Promote Testis Cord Formation during Testicular Differentiation* [Internet]. *Cell*
936 *Rep* [Internet] 2014;**8**:723–733.
- 937 Liu C-F, Bingham N, Parker K, Yao HH-C. Sex-specific roles of β -catenin in mouse gonadal
938 development. *Hum Mol Genet* [Internet] 2009;**18**:405–417.
- 939 Lutterodt MC, Sørensen KP, Larsen KB, Skouby SO, Andersen CY, Byskov AG. The number of
940 oogonia and somatic cells in the human female embryo and fetus in relation to whether or not
941 exposed to maternal cigarette smoking. *Hum Reprod* [Internet] 2009;**24**:2558–2566.
- 942 Maatouk DM, DiNapoli L, Alvers A, Parker KL, Taketo MM, Capel B. Stabilization of β -catenin in
943 XY gonads causes male-to-female sex-reversal. *Hum Mol Genet* [Internet] 2008;**17**:2949–2955.
- 944 Mamsen LS, Ernst EHE, Borup R, Larsen A, Olesen RH, Ernst EHE, Anderson RA, Kristensen SG,
945 Andersen CY. Temporal expression pattern of genes during the period of sex differentiation in
946 human embryonic gonads. *Sci Rep* [Internet] 2017;**7**:15961. Nature Publishing Group.
- 947 Mamsen LS, Lutterodt MC, Andersen EW, Skouby SO, Sørensen KP, Andersen CY, Byskov AG.
948 Cigarette smoking during early pregnancy reduces the number of embryonic germ and somatic
949 cells. *Hum Reprod* [Internet] 2010;**25**:2755–2761.
- 950 Mantione KJ, Kream RM, Kuzelova H, Ptacek R, Raboch J, Samuel JM, Stefano GB. Comparing
951 bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med Sci Monit*
952 *Basic Res* [Internet] 2014;**20**:138–142. International Scientific Literature, Inc.
- 953 Markiewicz W, Jaroszewski JJ, Bossowska A, Majewski M. NPY: its occurrence and relevance in
954 the female reproductive system. *Folia Histochem Cytobiol* [Internet] 2003;**41**:183–192.
- 955 McClelland KS, Bell K, Larney C, Harley VR, Sinclair AH, Oshlack A, Koopman P, Bowles J.
956 Purification and Transcriptomic Analysis of Mouse Fetal Leydig Cells Reveals Candidate
957 Genes for Specification of Gonadal Steroidogenic Cells¹. *Biol Reprod* [Internet] 2015;**92**:1–12.

- 958 Oxford University Press.
- 959 Munger SCS, Natarajan A, Looger LL, Ohler U, Capel B, Munger SCS, Aylor D, Syed H, Magwene
960 P, Threadgill D, *et al.* Fine Time Course Expression Analysis Identifies Cascades of Activation
961 and Repression and Maps a Putative Regulator of Mammalian Sex Determination. In Beier DR,
962 editor. *PLoS Genet* [Internet] 2013;**9**:e1003630. Public Library of Science.
- 963 Nef S, Schaad O, Stallings NR, Cederroth CR, Pitetti J-L, Schaer G, Malki S, Dubois-Dauphin M,
964 Boizet-Bonhoure B, Descombes P, *et al.* Gene expression during sex determination reveals a
965 robust female genetic program at the onset of ovarian development. *Dev Biol* [Internet]
966 2005;**287**:361–377.
- 967 O’Shaughnessy PJ, Antignac JP, Bizec B Le, Morvan ML, Svechnikov K, Söder O, Savchuk I,
968 Monteiro A, Soffientini U, Johnstonid ZC, *et al.* Alternative (Backdoor) androgen production
969 and masculinization in the human fetus. *PLoS Biol* 2019;**17**:. Public Library of Science.
- 970 O’Shaughnessy PJ, Baker PJ, Heikkilä M, Vainio S, McMahon AP. Localization of 17 β -
971 Hydroxysteroid Dehydrogenase/17-Ketosteroid Reductase Isoform Expression in the
972 Developing Mouse Testis—Androstenedione Is the Major Androgen Secreted by Fetal/Neonatal
973 Leydig Cells¹. *Endocrinology* [Internet] 2000;**141**:2631–2637.
- 974 O’Shaughnessy PJ, Baker PJJ, Monteiro A, Cassie S, Bhattacharya S, Fowler PA, O’Shaughnessy
975 PJ, Baker PJJ, Monteiro A, Cassie S, *et al.* Developmental changes in human fetal testicular cell
976 numbers and messenger ribonucleic acid levels during the second trimester. *J Clin Endocrinol*
977 *Metab* [Internet] 2007;**92**:4792–4801. Endocrine Society.
- 978 Ohhata T, Senner CE, Hemberger M, Wutz A. Lineage-specific function of the noncoding Tsix RNA
979 for Xist repression and Xi reactivation in mice. *Genes Dev* [Internet] 2011;**25**:1702–1715. Cold
980 Spring Harbor Laboratory Press.
- 981 Ostrer H, Huang HY, Masch RJ, Shapiro E. A cellular study of human testis development. *Sex Dev*

- 982 2007;**1**:286–292.
- 983 Ottolenghi C, Pelosi E, Tran J, Colombino M, Douglass E, Nedorezov T, Cao A, Forabosco A,
984 Schlessinger D. Loss of Wnt4 and Foxl2 leads to female-to-male sex reversal extending to germ
985 cells. *Hum Mol Genet* [Internet] 2007;**16**:2795–2804.
- 986 Pasquale EB. Developmental cell biology: Eph receptor signalling casts a wide net on cell behaviour.
987 *Nat Rev Mol Cell Biol* [Internet] 2005;**6**:462–475.
- 988 Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, Fan L, Sandelin A, Rinn JL, Regev
989 A, *et al.* Systematic identification of long noncoding RNAs expressed during zebrafish
990 embryogenesis. *Genome Res* [Internet] 2012;**22**:577–591.
- 991 Planells B, Gómez-Redondo I, Pericuesta E, Lonergan P, Gutiérrez-Adán A. Differential isoform
992 expression and alternative splicing in sex determination in mice. *BMC Genomics* [Internet]
993 2019;**20**..
- 994 Pollier J, Rombauts S, Goossens A. Analysis of RNA-Seq data with TopHat and Cufflinks for
995 genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol*
996 *Biol* [Internet] 2013;**1011**:305–315.
- 997 Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA,
998 Grasso CS, Kominsky HD, *et al.* Transcriptome sequencing across a prostate cancer cohort
999 identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol*
1000 [Internet] 2011;**29**:742–749.
- 1001 Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J,
1002 Landrum MJ, McGarvey KM, *et al.* RefSeq: an update on mammalian reference sequences.
1003 *Nucleic Acids Res* [Internet] 2014;**42**:D756–D763.
- 1004 Pundir S, Magrane M, Martin MJ, O'Donovan C, UniProt Consortium. Searching and Navigating
1005 UniProt Databases. *Curr Protoc Bioinforma* [Internet] 2015;**50**:1.27.1-10. John Wiley & Sons,

1006 Inc.: Hoboken, NJ, USA.

1007 Qian J, Yehia G, Molina C, Fernandes A, Donnelly R, Anjaria D, Gascon P, Rameshwar P. Cloning
1008 of human preprotachykinin-I promoter and the role of cyclic adenosine 5'-monophosphate
1009 response elements in its expression by IL-1 and stem cell factor. *J Immunol (Baltimore, Md*
1010 *1950)* [Internet] 2001;**166**:2553–2561.

1011 Rahmoun M, Lavery R, Laurent-Chaballier S, Bellora N, Philip GK, Rossitto M, Symon A, Pailhoux
1012 E, Cammas F, Chung J, *et al.* In mammalian foetal testes, SOX9 regulates expression of its
1013 target genes by binding to genomic regions with conserved signatures. *Nucleic Acids Res*
1014 [Internet] 2017;**45**:7191–7211.

1015 Rastetter RH, Smith CA, Wilhelm D. The role of non-coding RNAs in male sex determination and
1016 differentiation. *Reproduction* [Internet] 2015;**150**:R93-107. Society for Reproduction and
1017 Fertility.

1018 Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite.
1019 *Trends Genet* [Internet] 2000;**16**:276–277.

1020 Rolland AD, Evrard B, Darde TA, Béguec C Le, Bras Y Le, Bensalah K, Lavoué S, Jost B, Primig
1021 M, Dejuq-Rainsford N, *et al.* RNA profiling of human testicular cells identifies syntenic
1022 lncRNAs associated with spermatogenesis. *Hum Reprod* [Internet] 2019;Available from:
1023 <http://www.ncbi.nlm.nih.gov/pubmed/31247106>.

1024 Rolland AD, Lehmann KP, Johnson KJ, Gaido KW, Koopman P. Uncovering gene regulatory
1025 networks during mouse fetal germ cell development. *Biol Reprod* [Internet] 2011;**84**:790–800.
1026 Society for the Study of Reproduction.

1027 Santa Barbara P de, Méjean C, Moniot B, Malclès MH, Berta P, Boizet-Bonhoure B. Steroidogenic
1028 factor-1 contributes to the cyclic-adenosine monophosphate down-regulation of human SRY
1029 gene expression. *Biol Reprod* [Internet] 2001;**64**:775–783.

- 1030 Schepers G, Wilson M, Wilhelm D, Koopman P. SOX8 Is Expressed during Testis Differentiation in
1031 Mice and Synergizes with SF1 to Activate the *Amh* Promoter *in Vitro*. *J Biol Chem* [Internet]
1032 2003;**278**:28101–28108.
- 1033 Schmidt D. The murine winged-helix transcription factor Foxl2 is required for granulosa cell
1034 differentiation and ovary maintenance. *Development* [Internet] 2004;**131**:933–942.
- 1035 Sekido R, Lovell-Badge R. Sex determination involves synergistic action of SRY and SF1 on a
1036 specific Sox9 enhancer. *Nature* [Internet] 2008;**453**:930–934. Nature Publishing Group.
- 1037 Shima Y, Miyabayashi K, Haraguchi S, Arakawa T, Otake H, Baba T, Matsuzaki S, Shishido Y,
1038 Akiyama H, Tachibana T, *et al*. Contribution of Leydig and Sertoli Cells to Testosterone
1039 Production in Mouse Fetal Testes. *Mol Endocrinol* [Internet] 2013;**27**:63–73.
- 1040 Simon L, Ekman GC, Garcia T, Carnes K, Zhang Z, Murphy T, Murphy KM, Hess RA, Cooke PS,
1041 Hofmann M. ETV5 Regulates Sertoli Cell Chemokines Involved in Mouse Stem/Progenitor
1042 Spermatogonia Maintenance. *Stem Cells* [Internet] 2010;**28**:1882–1892.
- 1043 Sirotkin A V, Kardošová D, Alwasel SH, Harrath AH. Neuropeptide Y directly affects ovarian cell
1044 proliferation and apoptosis. *Reprod Biol* [Internet] 2015;**15**:257–260.
- 1045 Small CL, Shima JE, Uzumcu M, Skinner MK, Griswold MD. Profiling Gene Expression During the
1046 Differentiation and Development of the Murine Embryonic Gonad. *Biol Reprod* [Internet]
1047 2005;**72**:492–501. NIH Public Access.
- 1048 Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in
1049 Microarray Experiments. *Stat Appl Genet Mol Biol* [Internet] 2004;**3**:1–25.
- 1050 Sorby-Adams AJ, Marcoionni AM, Dempsey ER, Woenig JA, Turner RJ. The Role of Neurogenic
1051 Inflammation in Blood-Brain Barrier Disruption and Development of Cerebral Oedema
1052 Following Acute Central Nervous System (CNS) Injury. *Int J Mol Sci* [Internet] 2017;**18**..
- 1053 Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, Nejad P, Lee BT, Learned K, Karolchik

- 1054 D, Hinrichs AS, *et al.* The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res*
1055 [Internet] 2016;**44**:D717–D725.
- 1056 Stévant I, Kühne F, Greenfield A, Chaboissier M-C, Dermitzakis ET, Nef S. Dissecting Cell Lineage
1057 Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell
1058 Transcriptomics. *Cell Rep* [Internet] 2019;**26**:3272-3283.e3.
- 1059 Stévant I, Neirijnck Y, Borel C, Escoffier J, Smith LB, Antonarakis SE, Dermitzakis ET, Nef S.
1060 Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA
1061 Sequencing. *Cell Rep* [Internet] 2018;**22**:1589–1599.
- 1062 Sun J, Bhatia M. Substance P at the neuro-immune crosstalk in the modulation of inflammation,
1063 asthma and antimicrobial host defense. *Inflamm Allergy Drug Targets* [Internet] 2014;**13**:112–
1064 120.
- 1065 Sutton SW, Toyama TT, Otto S, Plotsky PM. Evidence that neuropeptide Y (NPY) released into the
1066 hypophysial-portal circulation participates in priming gonadotropes to the effects of
1067 gonadotropin releasing hormone (GnRH). *Endocrinology* [Internet] 1988;**123**:1208–1210.
- 1068 Svingen T, Jørgensen A, Rajpert-De Meyts E. Validation of endogenous normalizing genes for
1069 expression analyses in adult human testis and germ cell neoplasms. *Mol Hum Reprod*
1070 2014;**20**:709–718.
- 1071 Tao S, Xiu-Lei Z, Xiao-Lin L, Sai-Nan M, Yu-Zhu G, Xiang-Ting W. Recent Progresses of Long
1072 Noncoding RNA. *Biomed Sci* [Internet] 2016;**1**:34.
- 1073 Taylor DH, Chu ET-J, Spektor R, Soloway PD. Long non-coding RNA regulation of reproduction
1074 and development. *Mol Reprod Dev* [Internet] 2015;**82**:932–956. NIH Public Access.
- 1075 Toyoda-Ohno H, Obinata M, Matsui Y. Members of the ErbB receptor tyrosine kinases are involved
1076 in germ cell development in fetal mouse gonads. *Dev Biol* [Internet] 1999;**215**:399–406.
- 1077 Trapnell C, Roberts A, Goff L, Petrea G, Kim D, Kelley DR, Pimentel H, Salzberg S, Rinn JL,

- 1078 Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with
1079 TopHat and Cufflinks. *Natures Protoc* 2012;**7**:562–578.
- 1080 Uda M, Ottolenghi C, Crisponi L, Garcia JE, Deiana M, Kimber W, Forabosco A, Cao A,
1081 Schlessinger D, Pilia G. Foxl2 disruption causes mouse ovarian failure by pervasive blockage of
1082 follicle development. *Hum Mol Genet* [Internet] 2004;**13**:1171–1181.
- 1083 Vainio S, Heikkilä M, Kispert A, Chin N, McMahon AP. Female development in mammals is
1084 regulated by Wnt-4 signalling. *Nature* [Internet] 1999;**397**:405–409.
- 1085 Valle I del, Buonocore F, Duncan AJ, Lin L, Barenco M, Parnaik R, Shah S, Hubank M, Gerrelli D,
1086 Achermann JC. A genomic atlas of human adrenal and gonad development. *Wellcome Open Res*
1087 [Internet] 2017;**2**:25.
- 1088 Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: An open-source graphical
1089 user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics* 2011;**11**:996–999.
- 1090 Vaudel M, Burkhardt JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, Martens L, Barsnes H.
1091 PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat Biotechnol* [Internet]
1092 2015;**33**:22–24.
- 1093 Vidal VPI, Chaboissier M-C, Rooij DG de, Schedl A. Sox9 induces testis development in XX
1094 transgenic mice. *Nat Genet* [Internet] 2001;**28**:216–217.
- 1095 Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y,
1096 Reisinger F, Ternent T, *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic*
1097 *Acids Res* [Internet] 2016;**44**:11033. Oxford University Press.
- 1098 Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool
1099 using an alignment-free logistic regression model. *Nucleic Acids Res* [Internet] 2013;**41**:e74.
- 1100 Warr N, Siggers P, Bogani D, Brixey R, Pastorelli L, Yates L, Dean CH, Wells S, Satoh W, Shimono
1101 A, *et al.* Sfrp1 and Sfrp2 are required for normal male sexual development in mice. *Dev Biol*

- 1102 [Internet] 2009;**326**:273–284.
- 1103 Watanabe T, Cheng E, Zhong M, Lin H. Retrotransposons and pseudogenes regulate mRNAs and
1104 lncRNAs via the piRNA pathway in the germline. *Genome Res* [Internet] 2015;**25**:368–380.
1105 Cold Spring Harbor Laboratory Press.
- 1106 Wen K, Yang L, Xiong T, Di C, Ma D, Wu M, Xue Z, Zhang X, Long L, Zhang W, *et al*. Critical
1107 roles of long noncoding RNAs in Drosophila spermatogenesis. *Genome Res* [Internet]
1108 2016;**26**:1233–1244.
- 1109 Wichman L, Somasundaram S, Breindel C, Valerio DM, McCarrey JR, Hodges CA, Khalil AM.
1110 Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis
1111 and fertility. *Biol Reprod* [Internet] 2017;**97**:313–323. Oxford University Press.
- 1112 Wilhelm D, Washburn LL, Truong V, Fellous M, Eicher EM, Koopman P. Antagonism of the testis-
1113 and ovary-determining pathways during ovotestis development in mice. *Mech Dev* [Internet]
1114 2009;**126**:324–336.
- 1115 Wilhelm D, Yang JX, Thomas P. Mammalian sex determination and gonad development. In Thomas
1116 P, editor. *Curr Top Dev Biol* [Internet] 2013;**106**., p. 89–121. Academic Press.
- 1117 Winge SB, Dalgaard MD, Jensen JM, Graem N, Schierup MH, Juul A, Rajpert-De Meyts E,
1118 Almstrup K. Transcriptome profiling of fetal Klinefelter testis tissue reveals a possible
1119 involvement of long non-coding RNAs in gonocyte maturation. *Hum Mol Genet* [Internet]
1120 2017;**27**:430–439.
- 1121 Wu R, Su Y, Wu H, Dai Y, Zhao M, Lu Q. Characters, functions and clinical perspectives of long
1122 non-coding RNAs. *Mol Genet Genomics* [Internet] 2016;**291**:1013–1033.
- 1123 Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P,
1124 Fitzgerald S, Gil L, *et al*. Ensembl 2016. *Nucleic Acids Res* [Internet] 2016;**44**:D710–D716.
- 1125 Yusuf D, Butland SL, Swanson MI, Bolotin E, Ticoll A, Cheung WA, Zhang XYC, Dickman CTD,

- 1126 Fulton DL, Lim JS, *et al.* The transcription factor encyclopedia. *Genome Biol* [Internet]
1127 2012;**13**:R24. BioMed Central.
- 1128 Zhao L, Arsenault M, Ng ET, Longmuss E, Chau TC-Y, Hartwig S, Koopman P. SOX4 regulates
1129 gonad morphogenesis and promotes male germ cell differentiation in mice. *Dev Biol* [Internet]
1130 2017;**423**:46–56.
- 1131 Zhao L, Wang C, Lehman ML, He M, An J, Svingen T, Spiller CM, Ng ET, Nelson CC, Koopman P.
1132 Transcriptomic analysis of mRNA expression and alternative splicing during mouse sex
1133 determination. *Mol Cell Endocrinol* [Internet] 2018;Available from:
1134 <http://www.sciencedirect.com/science/article/pii/S030372071830234X>.
- 1135 Zimmermann C, Stévant I, Borel C, Conne B, Pitetti J-L, Calvel P, Kaessmann H, Jégou B, Chalmel
1136 F, Nef S. Research Resource: The Dynamic Transcriptional Profile of Sertoli Cells During the
1137 Progression of Spermatogenesis. *Mol Endocrinol* [Internet] 2015;**29**:627–642.
- 1138

1139 Figure legends**1140 Figure 1** Sample collection and assessment of homogeneity.

1141 (A) Human fetal gonads used in this study were collected at seven developmental stages, i.e. at 6,
1142 early 7, late 7, 9, 12, 13–14 and 17 postconceptional week (PCW). The number of replicates is
1143 indicated for each stage and sex. An overview of the main differentiation processes within human
1144 fetal testes and ovaries during the studied time window is also provided. PGC = primordial germ
1145 cell; LC = Leydig cell. The panel (B) displays the correlation (R^2) of the first 10 dimensions of
1146 principal component analysis (PCA) with the development stage and the genetic sex. The PCA was
1147 performed on expression data from 35,194 refined transcripts across all 48 human fetal gonads. Red
1148 values represent significant correlations (p -value $\leq 1\%$). (C) A scatter plot represents the position of
1149 each sample along the first two dimensions. Smaller dots represent samples, and are linked to bigger
1150 dots that represent the average expression of transcripts across replicates. The histogram represents
1151 the percentage of information carried by each dimension of the PCA. Testis samples are colored in
1152 blue, while ovaries are in red. Time point of each condition is provided in PCW. e7 = early 7 PCW;
1153 17 = late 7 PCW. The variability between the samples is mainly explained by their age of
1154 development (dimension 1) and by their genetic sex (dimension 2). The two arrows highlight the
1155 divergence of gonads transcriptomes, from a common origin (at 6 PCW) to their distinct fate. (D) A
1156 dendrogram shows the hierarchical relationship between the 48 samples. The hierarchical clustering
1157 is based on the 35 first PCA dimensions explaining 90% of the total variance of the data. Male
1158 samples are colored in blue, female samples are in red. (E) Coding potential analysis of refined
1159 transcripts. The combined results of the Protein-Encoding Potential (PEP) and the Proteomics
1160 Informed by Transcriptomic (PIT) strategies, i.e. Low or High PEP transcripts with (PIT+) or
1161 without (PIT-) identified peptide(s) are represented for each transcript biotype. (F) Two statistical
1162 filtrations were used to select differentially-expressed transcripts. First, we performed an “intra-sex”

1163 comparison in which all developmental stages were compared to each other during testis
1164 development on the one hand, and during ovarian development on the other hand (Fold-change ≥ 2
1165 in at least one comparison). Second, we performed an “inter-sex” comparison in which testes and
1166 ovaries were compared at each developmental stage (Fold-change ≥ 2 in at least one comparison).
1167 Subsequently, a linear models for microarray data (LIMMA) statistical test was performed on both
1168 sets of transcripts to select those with significant expression variation across replicates [false
1169 discovery rate (FDR)-adjusted F-value of ≤ 0.05]. A total of 13,145 transcripts that display “inter-
1170 sex” expression variations were defined as “Sexually dimorphic transcripts” (SDT), while 8,935
1171 developmentally-regulated transcripts that do not exhibit sexual dimorphism were defined as “Non-
1172 sexually dimorphic transcripts” (NSDT).

1173

1174 **Figure 2** Sexually dimorphic expression patterns during human gonad development.

1175 **(A)** Heatmap representation of 13,145 SDTs, distributed into 14 expression patterns (P1 to P14),
1176 across seven developmental stages for both testes and ovaries. Each row corresponds to a transcript,
1177 and each column an experimental condition, *i.e.* the average of testes or ovaries from a given PCW.
1178 The standardized abundance of transcripts is color-coded according to the scale bar, red
1179 corresponding to the highest expression level, blue to the lowest. **(B)** Repartition of known markers
1180 involved in gonad differentiation and development within SDT expression patterns. Note that several
1181 isoforms of a given transcript can be assembled and display distinct expression. **(C)** Gene ontology
1182 (GO) terms found to be enriched (BH corrected p-value < 0.05) in each expression pattern. **(D)**
1183 Transcript biotypes and isoform status proportion in SDT (pie chart) and within each cluster of
1184 differentially expressed transcripts (barplot). The comparison of the 13,145 SDT with the human
1185 reference transcriptome by Cuffcompare (Pollier *et al.*, 2013) classified them as known isoform

1186 (class code “=”), novel isoforms (class code “j”), novel unannotated transcripts (NUTs) in intronic
1187 regions (class code “i”), intergenic regions (“u”), antisense of known transcripts (class code “x”) or
1188 other ambiguous biotypes. Proportion of mRNAs, long non-coding (lnc)RNAs and NUTS in SDT
1189 clusters is given. Total number of transcripts in each cluster is indicated on the right side of the plot.
1190 An enrichment analysis using a hypergeometric strategy highlighted a significant accumulation of
1191 lncRNAs and NUT in the P12, P13 and P14 cluster (p-value <0.05) compared to their distribution
1192 within the 14 SDT clusters.

1193

1194 **Figure 3** Non-sexually dimorphic expression patterns during human gonad development.

1195 **(A)** Heatmap representation of 8,935 NSDTs, distributed into six expression patterns (Q1 to Q6),
1196 across seven developmental stages for both testes and ovaries. Each row is a transcript, and each
1197 column is an experimental condition, i.e. the average of testes or ovaries from a given PCW. The
1198 standardized abundance of transcripts is color-coded according to the scale bar, red corresponding to
1199 the highest expression level, blue to the lowest. **(B)** Repartition of known markers involved in gonad
1200 differentiation and development within SDT expression patterns. Note that several isoforms of a
1201 given can be assembled and display distinct expression. **(C)** GO terms found to be enriched (BH
1202 corrected p-value <0.05) in each expression pattern. **(D)** The comparison of the 8,935 NSDT with the
1203 human reference transcriptome by Cuffcompare (Pollier *et al.*, 2013) classified them as known
1204 isoform (class code “=”), novel isoforms (class code “j”), novel unannotated transcripts (NUTs) in
1205 intronic regions (class code “i”), intergenic regions (“u”), antisense of known transcripts (class code
1206 “x”) or other ambiguous biotypes. Proportions of mRNAs, lncRNAs and NUTS in NSDT (pie chart)
1207 and within NSDT clusters are given (barplot). The total number of transcripts in each cluster is
1208 indicated on the right side of the barplot. An enrichment analysis using a hypergeometric strategy

1209 highlighted a significant accumulation of lncRNAs and NUTs in Q1 and Q6 clusters (p-value <0.05)
1210 compared to their distribution within the six NSDT clusters.

1211

1212 **Figure 4** Cellular investigation of early-SDTs that are over-expressed in fetal testis.

1213 Expression levels (line graphic) and quantitative RT-PCR (histograms) of genes from expression
1214 pattern P1 and P2, which exhibit a higher differential expression (A) at 6 PCW, such as sex
1215 determining region Y (*SRY*), wntless Wnt ligand secretion mediator (*WLS*), C-X-C motif chemokine
1216 ligand 14 (*CXCL14*) and C-C motif chemokine receptor 1 (*CCR1*), and (B) at 7 PCW, such as SRY-
1217 box 9 (*SOX9*), SRY-box 10 (*SOX10*), EPH receptor B1 (*EPHB1*), MAGE family member B1
1218 (*MAGEB1*), fetal and adult testis expressed 1 (*FATE1*), erb-b2 receptor tyrosine kinase 3 (*ERBB3*),
1219 Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 1 (*CITED1*) and
1220 novel unannotated transcript antisense of *CITED1* (*TCONS_00249587*). Expression levels from
1221 RNA-sequencing (RNA-seq) as a function of age are depicted as blue lines for the testis and pink
1222 lines for the ovaries. Each point represents the mean fragments per kilobase of exon model per
1223 million reads mapped (FPKM) \pm SEM of the levels measured in four (12 PCW and younger) and two
1224 different gonads (13-14 and 17 PCW). Quantitative PCR was performed on the different testicular
1225 sorted cell populations of germ cells (KIT+, red bars) Sertoli cells (hEpA+, green bars) and other
1226 cells (KIT-/ hEpA-, grey bars). Each column shows a pool of sorted cells from five fetal (6.9-7.3
1227 PCW) testes. Each bar represents the mean \pm SEM of the fold change in target gene expression
1228 relative to the reference genes *RPLP0* and *RPS20*. (D) Representative immunohistochemistry of
1229 *SOX9*, *SOX10*, *KIAA1210*, *EPHB1*, *MAGEB1* and *FATE1* on a 7.1 PCW testis. Arrows indicate
1230 germ cells (GC). Scale bar: 50 μ M.

1231

1232 **Figure 5** Cellular investigation of early-SDTs that are over-expressed in fetal ovary.

1233 Expression levels (line graphic) and quantitative RT-PCR (histograms) of genes from expression
1234 pattern P8 and P9, which exhibit a higher differential expression (**A**) at 6 PCW, such as Neurexin 3
1235 (*NRXN3*), contactin 1 (*CNTN1*) and SET nuclear proto-oncogene (*SET*), (**B**) at 7 PCW, such as R-
1236 spondin 1 (*RSPO1*), neuropeptide Y (*NPY*), SRY-box 4 (*SOX4*) and NUT *TCONS_00224470*, and
1237 (**C**) or later on, as POU class 5 homeobox 1 (*POU5F1*) and NUTs *TCONS_00113718*,
1238 *TCONS_00055038* and *TCONS_00042565*. Expression levels from RNA-seq as a function of age are
1239 depicted as blue lines for the testis and pink lines for the ovaries. Each point represents the mean
1240 FPKM \pm SEM of the levels measured in four (12 PCW and younger) and two different gonads (13-
1241 14 and 17 PCW). Quantitative RT-PCR was performed on the ovarian sorted cell populations of
1242 germ cells (KIT+, pink bars) and other cells (KIT-, grey bars). Each column shows a pool of sorted
1243 cells from seven early differentiating (6.7-8.7 PCW, 7–9 PCW) and three fetal (10.6-11.7 PCW, 10–
1244 12 PCW) ovaries. Each bar represents the mean \pm SEM of the fold change in target gene expression
1245 relative to the reference genes *RPLP0* and *RPS20*. (**D**) Immunofluorescence for NRXN3 (green) and
1246 KIT (red), LIN28 (green) and CNTN1 (red), and WT1 (green) and KIT (red), in early differentiating
1247 ovaries (6-6.6 PCW). Scale bar: 100 μ M.

1248

1249 **Supplementary Figure S1** Refinement strategy of assembled transcripts.

1250 Following transcript reconstruction with Cufflinks, a refinement strategy was performed to discard
1251 sequencing and assembly artefacts: only transcripts with an expression of ≥ 1 FPKM in at least one
1252 experimental condition (average value of biological replicates) were considered; transcripts with a
1253 length of less than 200 nucleotides were discarded; novel transcript isoforms (Cuffcompare class “j”)

1254 and genes (classes “i”, “u” and “x”) were required to harbor at least two exons. FPKM = fragments
1255 per kilobase of exon model per million reads mapped.

1256

1257 **Supplementary Figure S2** Genomic and expression features comparison.

1258 Violin plot representation of selected expression and genomic features for all expressed mRNAs and
1259 long noncoding (lncRNAs) and novel unannotated transcripts (NUTs): **(A)** maximum abundance
1260 (log), **(B)** Shannon entropy, **(C)** sequence conservation (phastCons score), **(D)** cumulative exons size
1261 (log), **(E)** number of exons (log), and **(F)** GC content (%GC).

1262

1263 **Supplementary Figure S3** Early sexually dimorphic expression patterns during human gonad
1264 development.

1265 **(A)** Heatmap representation of 1,479 early sexually dimorphic transcripts (early-SDTs), distributed
1266 into 14 expression patterns (P1 to P14), across seven developmental stages for both testes and
1267 ovaries. Each row is a transcript, and each column is an experimental condition, i.e. the average of
1268 testes or ovaries from a given a gestational week (GW). The standardized abundance of transcripts is
1269 color-coded according to the scale bar, red corresponding to the highest expression level, blue to the
1270 lowest. **(B)** Repartition of known markers involved in gonad differentiation and development within
1271 early-SDT expression patterns. Note that several isoforms of a given gene can be assembled and
1272 display distinct expression. **(C)** Gene ontology (GO) terms found to be enriched (BH corrected p-
1273 value <0.05) in each expression pattern. **(D)** The comparison of the 1,479 early-SDT with the human
1274 reference transcriptome by Cuffcompare (Pollier *et al.*, 2013) classified them as known isoform
1275 (class code “=”), novel isoforms (class code “j”), novel unannotated transcripts (NUTs) in intronic
1276 regions (class code “i”), intergenic regions (“u”), antisense of known transcripts (class code “x”) or

1277 other ambiguous biotypes. Proportion of mRNAs, lncRNAs and NUTS in early-SDTs (pie chart) and
1278 within each cluster is given (barplot). Total number of transcripts in each cluster is indicated on the
1279 right side of the barplot.

1280

1281 **Supplementary Figure S4** Testicular and ovarian cell-sorting by flow cytometry.

1282 **(A)** Representative immunofluorescence of hEpA-FITC (green) staining of cord cells and KIT-PE
1283 (red) staining of germ cells in testis sections of a 7 PCW old embryo. **(B)** Representative Side (SSC)
1284 versus Forward (FSC) scatter plot showing hEpA/Mast/stem cell growth factor receptor Kit (KIT)
1285 dot plots according to size (FSC-H) and cellular granularity (SSC-H). Example of gating strategy for
1286 flow cytometry sorting of Sertoli cells (hEpA+/KIT-, green dots), germ cells (hEpA-/KIT+, red
1287 plots) and other cells types (hEpA-/KIT-, grey plots). **(C)** Quantitative RT-PCR of KIT proto-
1288 oncogene, receptor tyrosine kinase (KIT), nuclear receptor subfamily 2 group F member 2
1289 (NR2F2) and SRY-box transcription factor 9 (SOX9) was performed on the different sorted cell
1290 populations of germ cells (KIT+, red bars) Sertoli cells (hEpA+, green bars) and other cells (KIT-/
1291 hEpA-, grey bars). Each column shows a pool of sorted cells from five fetal (6.9-7.3 PCW) testes.
1292 Each bar represents the mean \pm SEM of the fold change in target gene expression relative to the
1293 reference genes ribosomal protein lateral stalk subunit P0 (RPLP0) and ribosomal protein S20
1294 (RPS20). **(D)** Representative immunofluorescence of KIT-PE (red) staining of germ cells in sections
1295 of an ovary at 11 PCW. **(E)** SSC versus FSC scatter plot showing KIT dot plots according to size
1296 (FSC-H) and cellular granularity (SSC-H). Example of gating strategy for flow cytometry sorting of
1297 germ cells (KIT+, red plots) and other cells types (KIT-, grey plots). **(F)** Quantitative RT-PCR of
1298 KIT, NR2F2 and forkhead box L2 (FOXL2) was performed on the different sorted cell populations
1299 of germ cells (KIT+, red bars) and other cells (KIT-, grey bars). Each column shows a pool of sorted
1300 cells from seven early differentiating (6.7-8.7 PCW, 7-9 PCW) and three fetal (10.6-11.7 PCW, 10-

1301 12 PCW) ovaries. Each bar represents the mean \pm SEM of the fold change in target gene expression
1302 relative to the reference genes RPLP0 and RPS20. Scale bars: 100 μ M.

1303

1304 **Table legends**

1305 **Table I. List of primers that were used for q-PCR experiments.**

1306 **Table II. Antibodies used for immunofluorescence and immunohistochemistry.**

1307 **Supplementary Table SI. Statistics for read mapping and transcript assembly.**

1308 **Supplementary Table SII. Early sexually dimorphic transcripts encoding transcription factors.**

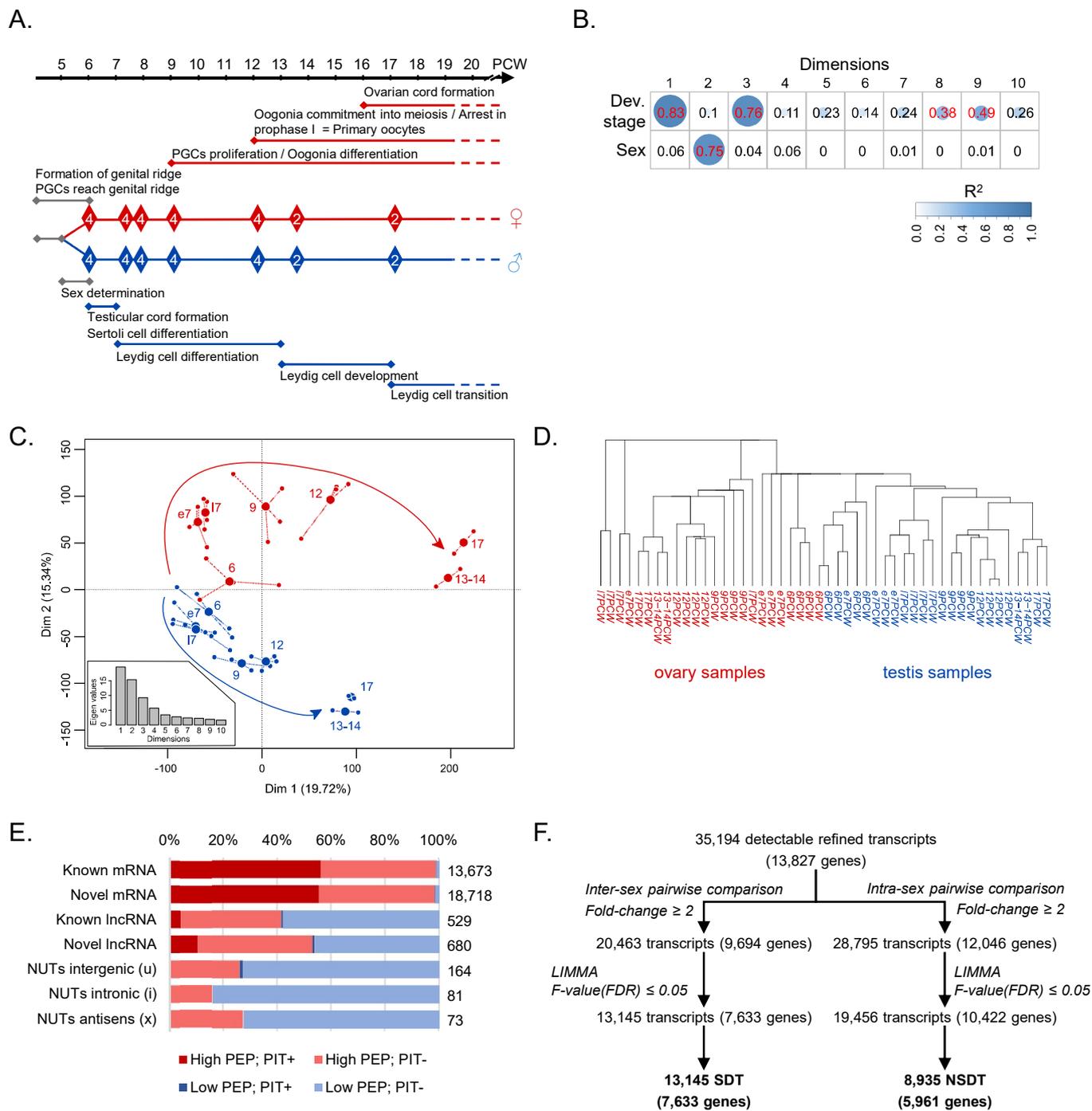
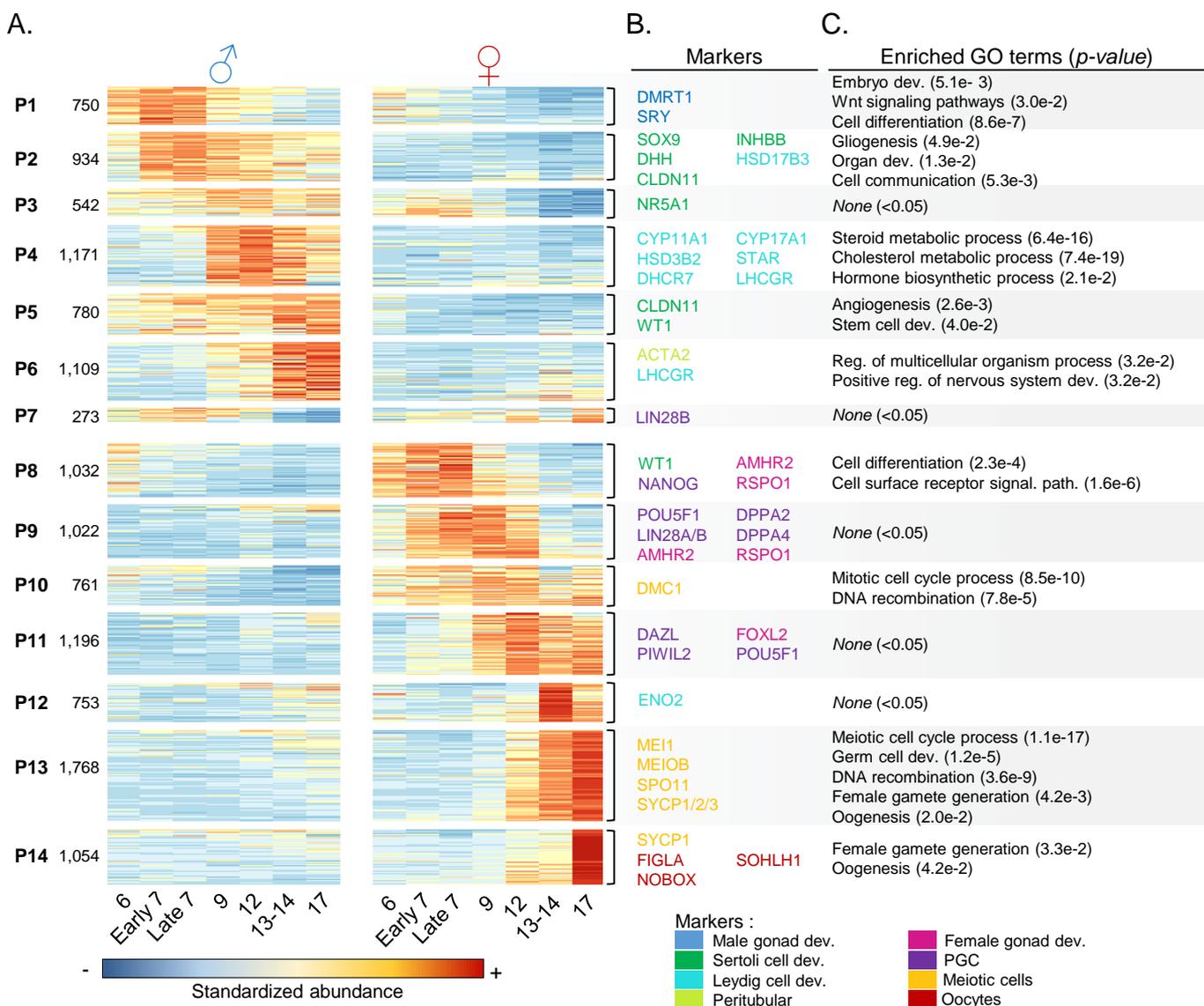


Fig. 1



D.

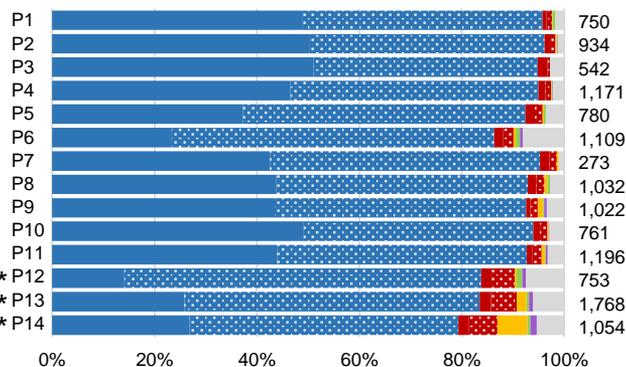
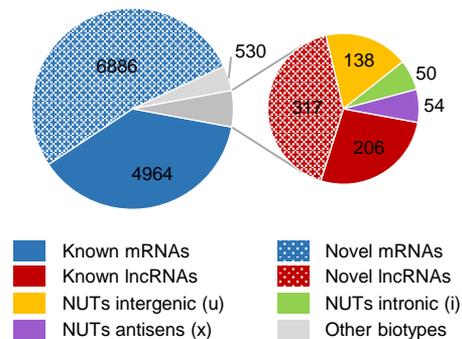


Fig. 2

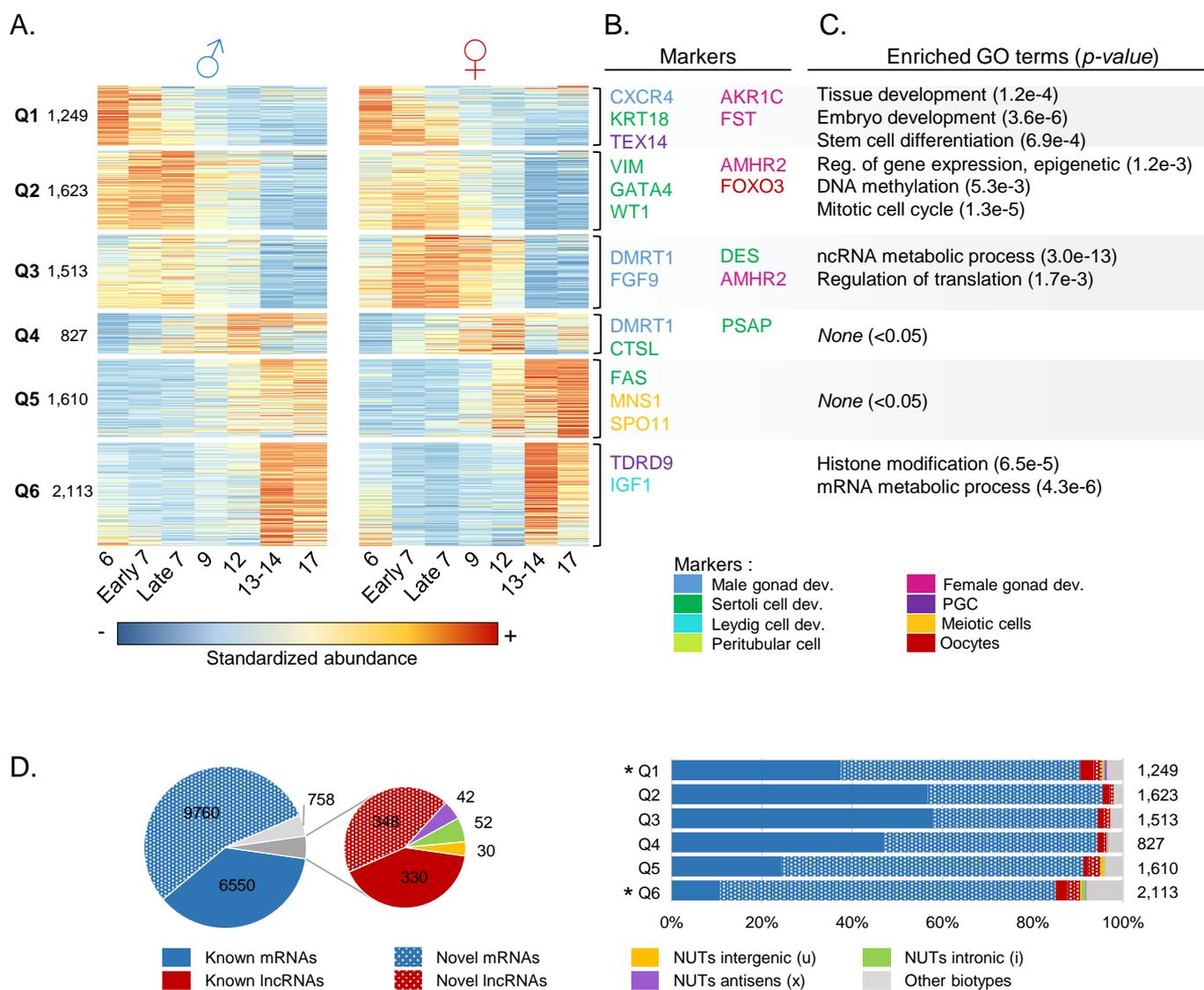


Fig. 3

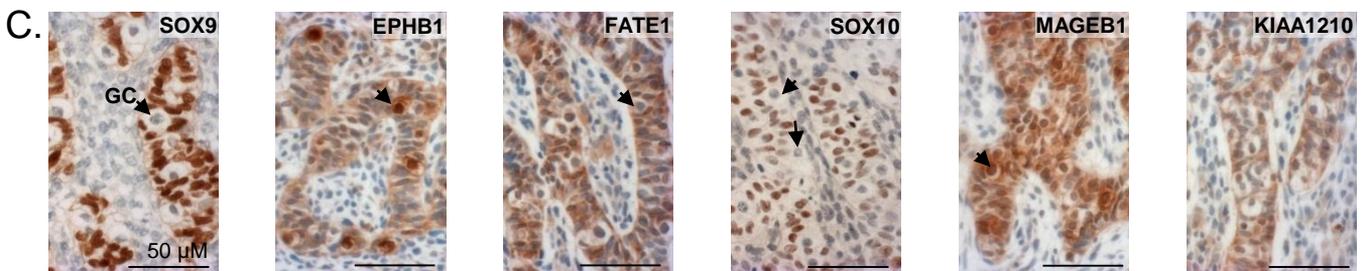
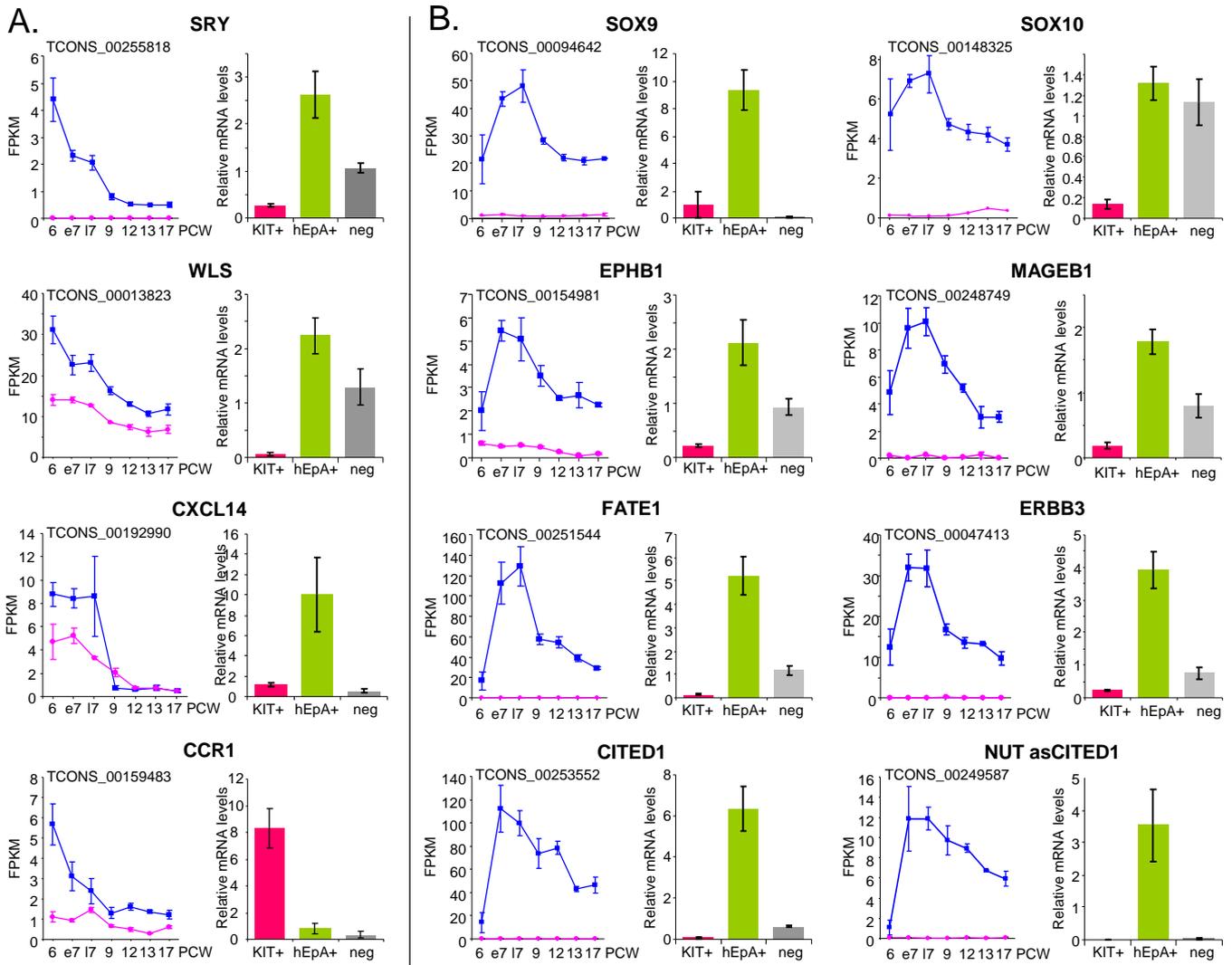


Fig. 4

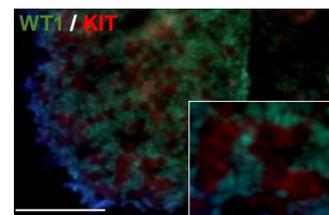
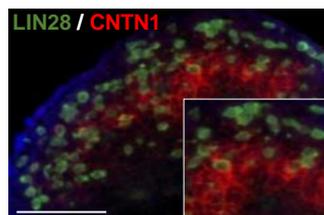
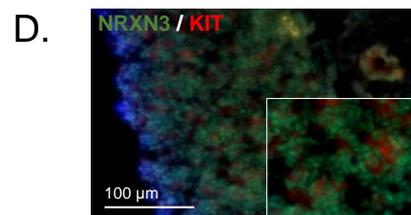
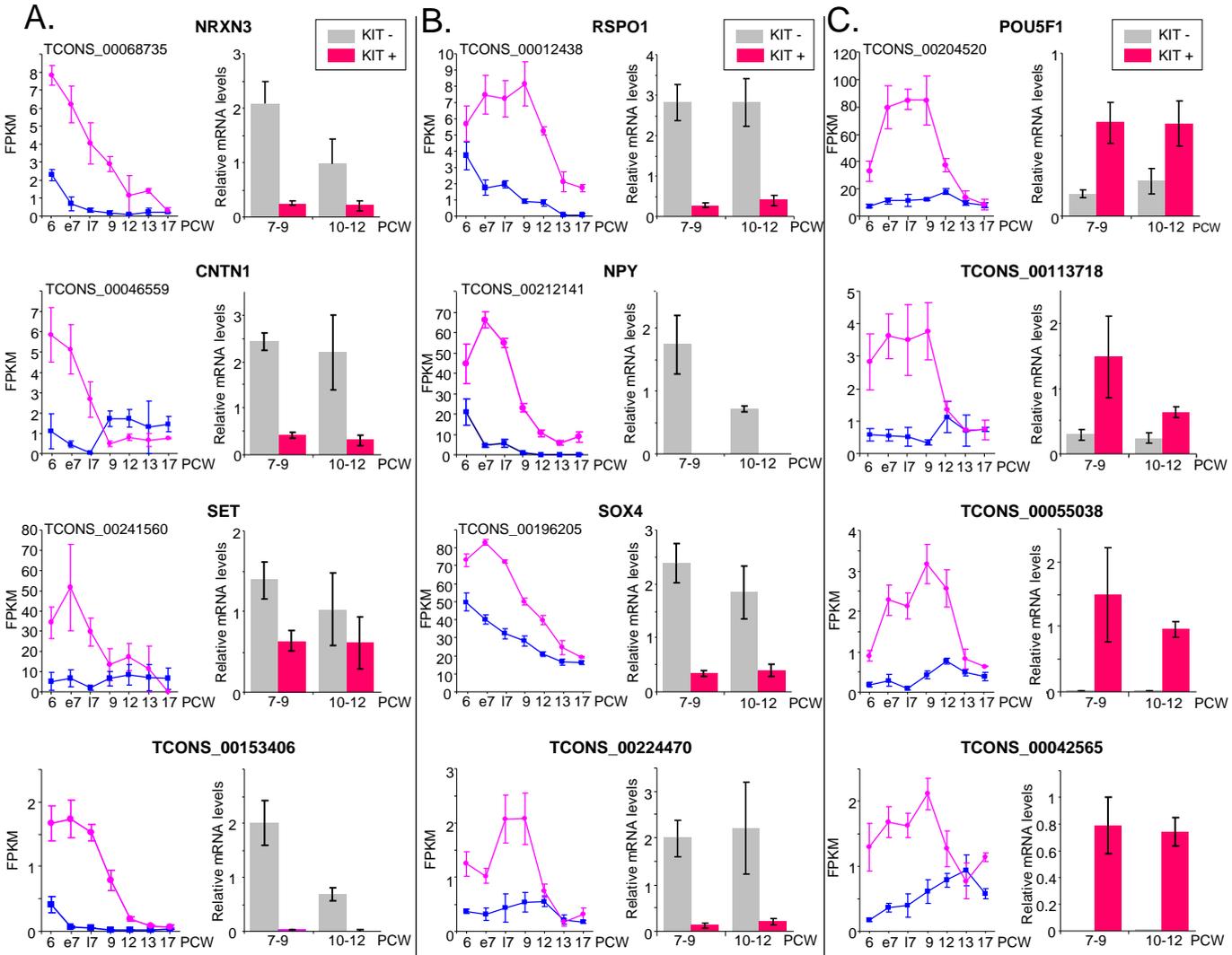


Fig. 5

180,242 assembled transcripts

(174,717 genes)

↓ ≥ 1 *FPKM*

60,437 detectable transcripts

(35,457 genes)

↓ ≥ 200 *nt*

60,136 long & detectable transcripts

(35,202 genes)



14,739 known isoforms ('=')

(10,899 genes)

41,015 novel isoforms ('j')

or novel genes ('i', 'u', 'x')

(30,392 genes)

↓ ≥ 2 *exons*

20,455 novel multi-exonic transcripts

(10,037 genes)



35,194 detectable refined transcripts

(13,827 genes)

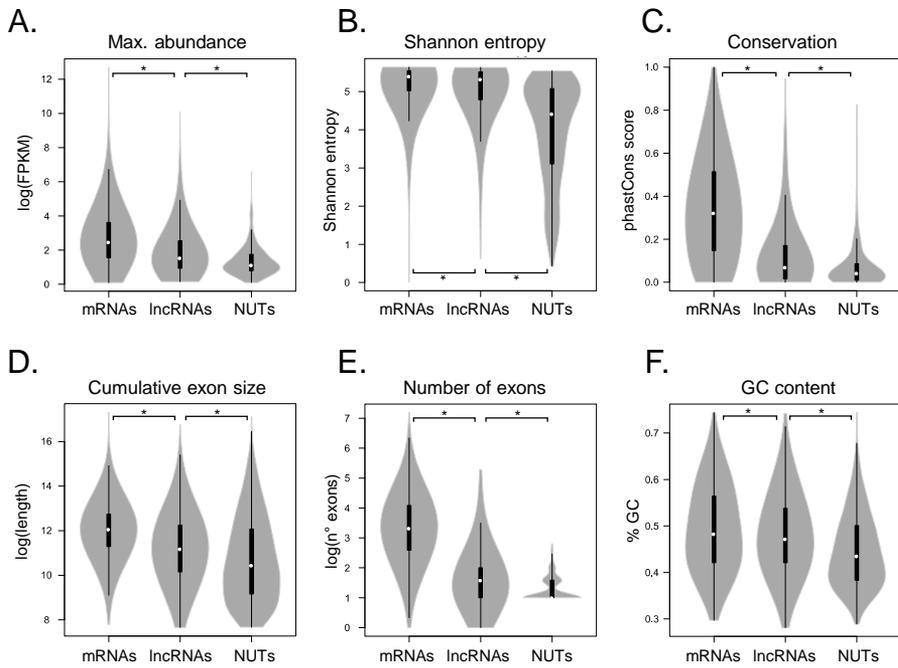


Fig. S2

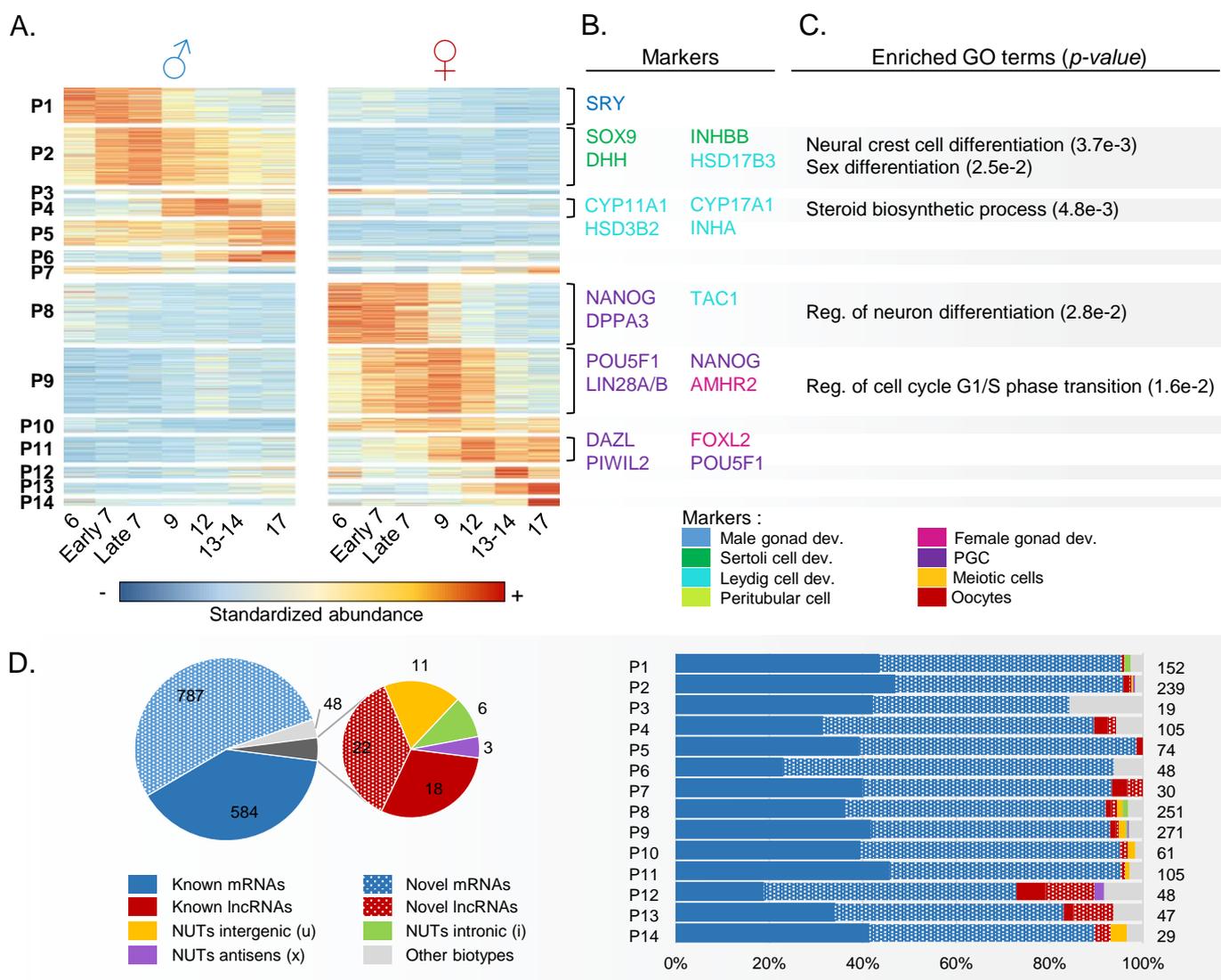


Fig. S3

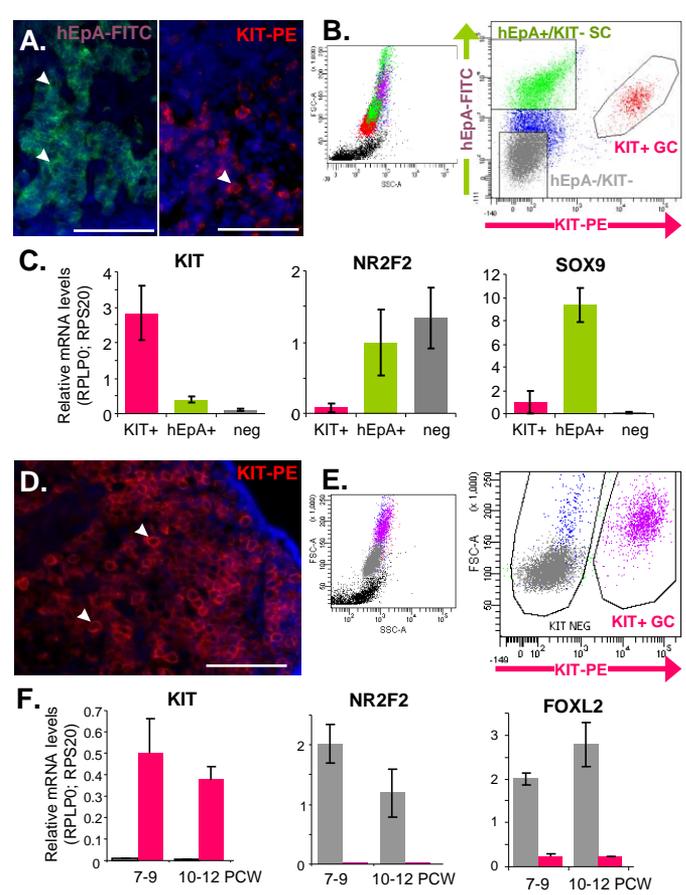


Fig. S4

Table 1. List of primers that were used for q-PCR experiments

| Gene | Forward (5'-3') |
|----------------|--------------------------|
| CCR1 | CAGAAAGCCCCAGAAACAAA |
| CITED1 | TGCACTTGATGTCAAGGGTG |
| CNTN1 | TTGGGAAGATGGTAGCTTGG |
| CXCL14 | ATGAAGCCAAAGTACCCGCA |
| EPHB1 | AGAGGAGGGAAAAGGACCAGG |
| ERBB3 | ACAGCCCCAGATCTGCAC |
| FATE1 | GGCAATTTCCAAGGCATACG |
| FOXL2 | GCGAAGTTCCCGTTCTACGA |
| KIT | TTCTTACCAGGTGGCAAAGG |
| MAGEB1 | CTATGGGGAACCCCGTAAGT |
| NPY | CTACATCAACCTCATCACCAGG |
| NR2F2 | GCCATAGTCCTGTTACCTCA |
| NRXN3 | GCTGAGAACAACCCCAATA |
| POU5F1 | TACTCCTCGGTCCCTTTC |
| RPLP0 | TCTACAACCCTGAAGTGCTTGAT |
| RPS20 | AACAAGCCGCAACGTAAAATC |
| RSPO1 | ACACTTCCCAGCATCTGAGACCAA |
| SET | AAATCAAATGGAAATCTGGAAAGG |
| SOX10 | AAGCCTCACATCGACTTCGG |
| SOX4 | GACCTGCTCGACCTGAACC |
| SOX9 | AACGCCTTCATGGTGTGG |
| SRY | ACAGTAAAGGCAACGTCCAG |
| TCONS_00042565 | GCGGCCCTAAGACAAAGAAC |
| TCONS_00055038 | CCCACCTTCCTCCTTCCTG |
| TCONS_00113718 | GCCCAACCACAGAAGGTTT |
| TCONS_00153406 | TCTACTTGTTTCTGGAGCTGAAG |
| TCONS_00224470 | CCTGGGCTACACTGGTCTTT |
| TCONS_00249587 | GGGAGAAAAGTAGCCCCAAG |
| WLS | CCTTGGTTCCAATTCATGCT |

| Reverse (5'-3') | Size (bp) | Reference |
|--------------------------|------------------|------------------|
| GGTGTTTGGAGTTTCCATCC | 80 | Primer 3 |
| GTTGTAGGAGAGCCTATTGG | 201 | PMID:22703800 |
| TGATAACAAGGGTCCAGTGC | 116 | PMID: 26855587 |
| TCTCGTTCCAGGCGTTGTAC | 148 | PMID: 24700803 |
| GGTTTCCCACGGCATCTC | 183 | PMID: 24121831 |
| GTTGGGCGAATGTTCTCATC | 78 | PMID: 23991224 |
| CTAGTCTGCGCCACTGCATC | 68 | PMID: 17761949 |
| CTCGTTGAGGCTGAGGTTGT | 75 | Primer BLAST |
| AAATGCTTTCAGGTGCCATC | 209 | PMID: 21668453 |
| GGTTTCAGCATAGGCTCTCG | 130 | Primer 3 |
| TCACCACATTGCAGGGTCT | 133 | PMID: 27722841 |
| AATCTCGTCGGCTGGTTG | 131 | PMID: 24318875 |
| ATGCTGGCTGTAGAGCGATT | 179 | PMID: 28013231 |
| CAAAAACCCTGGCACAAACT | 131 | PMID: 24743772 |
| CAATCTGCAGACAGACTGG | 96 | PMID: 24743772 |
| ACGATCCCACGTCTTAGAACC | 166 | PMID: 24743772 |
| TGCTGAACAGGATGGGAAGAAGGT | 146 | PMID: 23617070 |
| AAAGAAGCTCTCTGGTTCCTCATG | 101 | PMID: 22677993 |
| TCCATGTTGGACATTACCTCGT | 67 | PMID: 23338937 |
| CCGGGCTCGAAGTAAAATCC | 107 | PMID: 28535514 |
| TCTCGCTCTCGTTCAGAAGTC | 124 | Primer3 |
| ATCTGCGGGAAGCAAACCTGC | 293 | PMID: 11869379 |
| TCTGACCAGAAAATCGCTTC | 104 | Primer3 |
| TCAGTGCAGAAGAGCCCAA | 86 | Primer3 |
| CTGGGACAGGATGGAGAGG | 97 | Primer3 |
| TGCAGTAACATCCTCCTCCTC | 120 | Primer3 |
| GAGCACCTCATTCTTGGCT | 70 | Primer3 |
| GAAGCAAATGGAGAGACGGA | 111 | Primer 3 |
| TTCAGTCCACTCAGCAAACG | 138 | Primer3 |

Table 2. Antibodies used for immunofluorescence and immunohistochemistry.

| Antigen | | HIER |
|----------------|---|-------------|
| hEpA-FITC | Mouse FITC-coupled anti-human epithelial antigen | - |
| CNTN1 | Contactin 1 | Ci |
| EPHB1 | ephrin type-B receptor 1 | Ci |
| ERBB3 | HER3/ErbB3 (D22C5) XP® Rabbit mAb | Ci |
| FATE1 | fetal and adult testis expressed 1 | Ci |
| KIAA1210 | KIAA1210 | Ci |
| KIT/CD117-PE | mouse R-Phycoerythrin-coupled anti-human KIT/CD117; clone 104D2 | - |
| LIN28 | lin-28 homolog A | Ci |
| MAGEB1 | melanoma-associated antigen B1 | Ci |
| NRXN3 | neurexin 3 | Ci |
| SOX9 | SRY (sex determining region Y)-box9 | Ci |
| SOX10 | SRY (sex determining region Y)-box 10 | Ci |
| WT1 | Wilms tumor 1 | Ci |

| Dilution | Antibody supplier and product number |
|----------------------|---|
| 1 :100 | Dako, F0860 |
| 1 :100 (IF) | R&D systems, AF904 |
| 1:2500- 1:5000 | Sigma, HPA067740 |
| 0,215277778 | Cell Signaling Tech., #12708 |
| 1:1000-1:2500 | Sigma, HPA034604 |
| 1:200- 1:500 | Sigma, HPA048322 |
| 1 :100 | BioLegend, 313204 |
| 1 :100 (IF) | Abcam, Ab46020 |
| 1:100 - 1:250 | Sigma, HPA001193 |
| 1 :200 / 1 :100 (IF) | Sigma, HPA0002727 |
| 0,111111111 | Millipore, AB5535 |
| 1:200- 1:500 | Sigma, HPA068898 |
| 1:100 (IF) | Santa-Cruz Biotech, sc-192 |

Table S1. Statistics for read mapping and transcript assembly.

Average number of raw and mapped reads, percentage of mapping and number of assembled transcripts are indicated for each experimental condition. PCW = post coitum week. The number of non-redundant assembled transcripts is also provided during testicular and ovarian developments, as well as the overall number of non-redundant assembled transcripts in this study.

| | PCW | 6 | e7 | 17 | 9 | 12 | 13-14 | 17 | Total |
|--|------------------------------|--------------------------------|-----------------------------|--------------------------------|---------------------------------|---------------------------------|--------------------------------|--------------------------------|----------------|
| Fetal testis | n° samples | 4 | 4 | 4 | 4 | 4 | 2 | 2 | 24 |
| | Average sequenced | 50 163 366 | 43 637 558 | 45 033 799 | 44 578 922 | 61 408 394 | 46 992 435 | 47 295 757 | 339 110 230 |
| | pairs of read | + - 5 540 241 | + - 3 217 370 | + - 4 891 256 | -6124741 | + - 31 392 665 | + - 3 747 882 | + - 2 131 643 | |
| | Average mapped pairs of read | 45 088 608 + - 4 567 496 | 39 066 943 + - 2 877 937 | 38 915 843 + - 6 718 006 | 36 500 690 + - 3 675 842 | 46 586 645 + - 20 642 653 | 40 120 429 + - 2 994 901 | 38 721 188 + - 4 093 627 | 285 000 344 |
| | % of mapped reads | 89,9 | 89,5 | 86,4 | 81,9 | 75,9 | 85,4 | 81,9 | 84,4 |
| | n° reconstructed transcripts | 164 931 | 129 802 | 103 431 | 116 903 | 107 148 | 164 652 | 147 287 | 142 953 |
| | | | | | | | | | |
| Fetal ovary | n° samples | 4 | 4 | 4 | 4 | 4 | 2 | 2 | 24 |
| | Average sequenced | 51 756 993 | 42 038 738 | 43 167 327 | 45 461 320 | 43 085 238 | 47 163 616 | 38 044 232 | 310 717 463 |
| | pairs of read | + - 3 476 980 | + - 3 950 805 | + - 6 221 180 | + - 8 410 709 | + - 6 565 924 | + - 3 195 188 | + - 2 869 127 | |
| | Average mapped pairs of read | 46 047 326 + - 3 200 842 | 38 087 267 + - 3 581 721 | 33 349 134 + - 3 167 555 | 37 688 289 + - 10 017 251 | 34 516 893 + - 5 232 477 | 38 186 847 + - 1 789 162 | 33 236 072 + - 2 927 283 | 261 111 827 |
| | % of mapped reads | 89 | 90,6 | 77,3 | 82,9 | 80,1 | 81 | 87,4 | 84 |
| | n° reconstructed transcripts | 186 499 | 121 800 | 93 201 | 12 817 | 112 165 | 191 983 | 174 783 | 168 195 |
| n° of non-redundant reconstructed transcripts | | | | | | | | | 180 242 |

Table SII . Early sexually dimorphic transcrip

Early SDT genes encoding transcription factors (column E) are reported. Known interactions including TRRUST (Han *et al.*, 2015) and the T associations with sex reversal, DSDs or testis/ov

| Gene symbol | Ensemble Gene IDs |
|--------------------|--------------------------|
| ADAMTS19 | ENSG00000145808 |
| BACH2 | ENSG00000112182 |
| BCL11B | ENSG00000127152 |
| BIN1 | ENSG00000136717 |
| BRDT | ENSG00000137948 |
| BRIP1 | ENSG00000136492 |
| CBFA2T2 | ENSG00000078699 |
| CDX1 | ENSG00000113722 |
| CDYL | ENSG00000153046 |
| CHD9 | ENSG00000177200 |
| CITED1 | ENSG00000125931 |
| CREBBP | ENSG00000005339 |
| CREM | ENSG00000095794 |
| CUL4B | ENSG00000158290 |
| CUX2 | ENSG00000111249 |
| DEPDC7 | ENSG00000121690 |
| ELF4 | ENSG00000102034 |
| ESR1 | ENSG00000091831 |
| ESR2 | ENSG00000140009 |
| ETV4 | ENSG00000175832 |
| ETV5 | ENSG00000244405 |
| FOSL2 | ENSG00000075426 |
| FOXH1 | ENSG00000160973 |
| FOXI3 | ENSG00000214336 |
| FOXL2 | ENSG00000183770 |
| GABPB2 | ENSG00000143458 |
| GATAD2A | ENSG00000167491 |
| GLI1 | ENSG00000111087 |
| GRIP1 | ENSG00000155974 |
| HEY2 | ENSG00000135547 |
| HIC2 | ENSG00000169635 |
| HIST1H1T | ENSG00000187475 |
| HIVEP2 | ENSG00000010818 |
| IRX1 | ENSG00000170549 |
| KDM4C | ENSG00000107077 |
| KLF16 | ENSG00000129911 |

| | |
|--------------|-----------------|
| KLF4 | ENSG00000136826 |
| KLF8 | ENSG00000102349 |
| LARP1B | ENSG00000138709 |
| LBX2 | ENSG00000179528 |
| LHX2 | ENSG00000106689 |
| LHX9 | ENSG00000143355 |
| LIN28A | ENSG00000131914 |
| LIN28B | ENSG00000187772 |
| MACF1 | ENSG00000127603 |
| MAEL | ENSG00000143194 |
| MBNL2 | ENSG00000139793 |
| MED23 | ENSG00000112282 |
| MTA1 | ENSG00000182979 |
| MYBL2 | ENSG00000101057 |
| MYCL | ENSG00000116990 |
| NANOG | ENSG00000111704 |
| NANOGP1 | ENSG00000176654 |
| NCOA1 | ENSG00000084676 |
| NCOR1 | ENSG00000141027 |
| NFAT5 | ENSG00000102908 |
| NFATC2 | ENSG00000101096 |
| NFE2L3 | ENSG00000050344 |
| NFKB2 | ENSG00000077150 |
| NKRF | ENSG00000186416 |
| NKX3-1 | ENSG00000167034 |
| NR6A1 | ENSG00000148200 |
| NRG1 | ENSG00000157168 |
| NRK | ENSG00000123572 |
| ONECUT1 | ENSG00000169856 |
| PHB2 | ENSG00000215021 |
| PHF8 | ENSG00000172943 |
| PLXNC1 | ENSG00000136040 |
| POU5F1 | ENSG00000204531 |
| POU5F1B | ENSG00000212993 |
| PPARG | ENSG00000132170 |
| PRDM1 | ENSG00000057657 |
| RAD51 | ENSG00000051180 |
| RBM26 | ENSG00000139746 |
| RBMX | ENSG00000147274 |
| REST | ENSG00000084093 |
| RFX2 | ENSG00000087903 |
| RNF125 | ENSG00000101695 |
| RP11-313J2.1 | ENSG00000215146 |
| SALL1 | ENSG00000103449 |

| | |
|--------------|-----------------|
| SALL4 | ENSG00000101115 |
| SAMD11 | ENSG00000187634 |
| SAP18 | ENSG00000150459 |
| SETDB2 | ENSG00000136169 |
| SF3A2 | ENSG00000104897 |
| SMAD9 | ENSG00000120693 |
| SMARCA1 | ENSG00000102038 |
| SOX10 | ENSG00000100146 |
| SOX9 | ENSG00000125398 |
| SP6 | ENSG00000189120 |
| SRX | ENSG00000184895 |
| TBX1 | ENSG00000184058 |
| TCF25 | ENSG00000141002 |
| TCF3 | ENSG00000071564 |
| TEAD4 | ENSG00000197905 |
| TFAP2C | ENSG00000087510 |
| TOX | ENSG00000198846 |
| UBTF | ENSG00000108312 |
| VENTX | ENSG00000151650 |
| WHSC1 | ENSG00000109685 |
| ZBTB1 | ENSG00000126804 |
| ZBTB7C | ENSG00000184828 |
| ZC3H11A | ENSG00000058673 |
| ZC3HAV1 | ENSG00000105939 |
| ZFP30 | ENSG00000120784 |
| ZFP42 | ENSG00000179059 |
| ZFY | ENSG00000067646 |
| ZFYVE20 | ENSG00000131381 |
| ZKSCAN8 | ENSG00000198315 |
| ZMYND11 | ENSG00000015171 |
| ZMYND8 | ENSG00000101040 |
| ZNF138 | ENSG00000197008 |
| ZNF208 | ENSG00000160321 |
| ZNF213 | ENSG00000085644 |
| ZNF217 | ENSG00000171940 |
| ZNF227 | ENSG00000131115 |
| ZNF232 | ENSG00000167840 |
| ZNF281 | ENSG00000162702 |
| ZNF385A | ENSG00000161642 |
| ZNF41 | ENSG00000147124 |

| | |
|--------|-----------------|
| ZNF468 | ENSG00000204604 |
| ZNF560 | ENSG00000198028 |
| ZNF607 | ENSG00000198182 |
| ZNF638 | ENSG00000075292 |
| ZNF66 | ENSG00000160229 |
| ZNF676 | ENSG00000196109 |
| ZNF729 | ENSG00000196350 |
| ZNF76 | ENSG00000065029 |
| ZNF90 | ENSG00000213988 |
| ZNF93 | ENSG00000184635 |
| ZNF98 | ENSG00000197360 |

ption factors.

(columns A-C), their corresponding expression pattern(s) (column D) and their early SDT target genes between transcription factors and their known target genes were extracted from public databases Transcription Factor encyclopedia (Yusuf *et al.*, 2012). Gene names in red, green or blue have known primary cancer, respectively.

| Assembled transcripts IDs | Expression patterns | eSDT target genes |
|--|---------------------|---|
| TCONS_00184938 | P8 | - |
| TCONS_00206773;TCONS_00206776;TCONS_00206777 | P8 | - |
| TCONS_00073504 | P8 | - |
| TCONS_00130031;TCONS_00130033 | P9 | - |
| TCONS_00004335 | P14 | - |
| TCONS_00098944 | P7 | - |
| TCONS_00136955 | P9 | - |
| TCONS_00185901 | P11 | PPARG ;POSTN |
| TCONS_00195284 | P8 | - |
| TCONS_00084978 | P3 | - |
| TCONS_00249587;TCONS_00253552;TCONS_00253553;TCONS_00253554;TCONS_00253556 | P2 | - |
| TCONS_00087391 | P11 | RAD51 |
| TCONS_00022167 | P1 | ACE;TAC1;G6PD |
| TCONS_00254445;TCONS_00254446 | P1 | - |
| TCONS_00049582 | P8 | - |
| TCONS_00033456 | P11 | - |
| TCONS_00254729 | P5 | - |
| TCONS_00201882 | P12 | PMAIP1;GREB1;ESR1;PLAC1 |
| TCONS_00072230 | P11 | - |
| TCONS_00097882;TCONS_00097883 | P9 | - |
| TCONS_00165043 | P9 | - |
| TCONS_00115256 | P6 | CLU |
| TCONS_00236474 | P9 | - |
| TCONS_00128881 | P9 | - |
| TCONS_00163150 | P11 | FOXL2; CYP17A1 ; CYP11A1 |
| TCONS_00005817 | P13 | - |
| TCONS_00107359;TCONS_00107362 | P9 | - |
| TCONS_00047499 | P2 | SOX9 ;SFRP1 |
| TCONS_00053783 | P8 | - |
| TCONS_00200804 | P2 | - |
| TCONS_00145356;TCONS_00145358 | P9 | - |
| TCONS_00204095 | P13 | - |
| TCONS_00208943 | P5 | - |
| TCONS_00180488;TCONS_00180489 | P8 | - |
| TCONS_00237196 | P12 | - |
| TCONS_00110133 | P8 | - |

| | | |
|-------------------------------|-------|---------------------------------------|
| TCONS_00245945 | P9 | LAMA3;LAMA1;N ANOG;PFKP;IFITM 3 |
| TCONS_00249352 | P9 | - |
| TCONS_00170683 | P9 | - |
| TCONS_00128298 | P8 | - |
| TCONS_00241252 | P11 | - |
| TCONS_00007929 | P6 | - |
| TCONS_00001570;TCONS_00001571 | P9 | - |
| TCONS_00199732 | P9 | - |
| TCONS_00002167 | P11 | - |
| TCONS_00006676 | P13 | - |
| TCONS_00060407 | P2 | - |
| TCONS_00208410 | P9 | - |
| TCONS_00070138 | P13 | ESR1 |
| TCONS_00137452 | P9 | MYBL2 |
| TCONS_00012583;TCONS_00012585 | P9 | - |
| TCONS_00044962;TCONS_00044963 | P8;P9 | POU5F1 |
| TCONS_00044969 | P9 | - |
| TCONS_00115064 | P11 | - |
| TCONS_00096460;TCONS_00096465 | P5;P1 | PPARG ;IGFBP3;ES R1 |
| TCONS_00085918 | P7 | CCL2 |
| TCONS_00140516;TCONS_00140524 | P8 | ENPP2 |
| TCONS_00212178 | P9 | - |
| TCONS_00024595 | P2 | - |
| TCONS_00254399 | P8 | - |
| TCONS_00232099;TCONS_00232101 | P5;P2 | ESR1; ACTG2 |
| TCONS_00246680;TCONS_00246681 | P11 | - |
| TCONS_00226393 | P8 | - |
| TCONS_00250221 | P8 | - |
| TCONS_00080027 | P8 | - |
| TCONS_00051081 | P2 | - |
| TCONS_00253265 | P13 | - |
| TCONS_00048895 | P2 | - |
| TCONS_00204519;TCONS_00204520 | P9 | NANOG; ZFP42 |
| TCONS_00230146 | P9 | - |
| TCONS_00149760 | P9 | CAV1;KLF4;GSTA 2 |
| TCONS_00199766;TCONS_00199790 | P9 | - |
| TCONS_00075018 | P8 | - |
| TCONS_00063928 | P1 | - |
| TCONS_00254939 | P12 | KIT |
| TCONS_00168163;TCONS_00168164 | P8 | TAC1;LIN28A;GAB RB3;KCNQ2 |
| TCONS_00110433 | P13 | - |
| TCONS_00101424 | P9 | - |
| TCONS_00027617 | P9 | - |
| TCONS_00088597 | P10 | - |

| | | |
|---|--------|---|
| TCONS_00140537;TCONS_00140540;TCONS_00140543;TC ONS_00140544 | P9 | SALL4;POU5F1 |
| TCONS_00000033;TCONS_00000037 | P1 | - |
| TCONS_00056892 | P8 | - |
| TCONS_00058144;TCONS_00058145;TCONS_00058146 | P8 | - |
| TCONS_00106327;TCONS_00106328;TCONS_00106329 | P2;P5 | - |
| TCONS_00061933 | P7 | - |
| TCONS_00254705 | P2 | - |
| TCONS_00148324;TCONS_00148325 | P2 | MPZ;PLP1;GJB1; E DNRB |
| TCONS_00094642 | P2 | KLF4;COL2A1;HA PLN1;COL9A1;PR AME; SOX10 |
| TCONS_00098228 | P1 | - |
| TCONS_00255818 | P1 | SOX9;PROM1 |
| TCONS_00145221 | P2 | - |
| TCONS_00086893;TCONS_00086894 | P2 | - |
| TCONS_00110107;TCONS_00110108;TCONS_00110111 | P13;P9 | POU5F1 |
| TCONS_00044600 | P9 | - |
| TCONS_00137982;TCONS_00137983 | P9 | ESR1 |
| TCONS_00233028 | P5 | - |
| TCONS_00097957 | P8 | - |
| TCONS_00025976 | P8 | - |
| TCONS_00166063 | P9 | - |
| TCONS_00067741 | P1 | - |
| TCONS_00104942 | P2 | - |
| TCONS_00008281 | P6 | - |
| TCONS_00223757 | P3 | - |
| TCONS_00112190 | P12 | - |
| TCONS_00173057;TCONS_00173058;TCONS_00173059 | P8;P9 | - |
| TCONS_00255507;TCONS_00255508;TCONS_00255510 | P2;P5 | - |
| TCONS_00158407 | P3 | - |
| TCONS_00196743 | P10 | - |
| TCONS_00020765 | P14 | - |
| TCONS_00140371 | P4 | - |
| TCONS_00213671 | P8 | - |
| TCONS_00111586;TCONS_00111591 | P9 | - |
| TCONS_00083148 | P4 | - |
| TCONS_00140595;TCONS_00140597;TCONS_00140603 | P9;P8 | - |
| TCONS_00108714 | P12 | - |
| TCONS_00095744 | P9 | - |
| TCONS_00018508 | P1 | - |
| TCONS_00053148 | P8 | - |
| TCONS_00253003 | P5 | - |

| | | |
|-------------------------------|--------|---|
| TCONS_00113462 | P8 | - |
| TCONS_00110665 | P9 | - |
| TCONS_00112218 | P9 | - |
| TCONS_00117137 | P9 | - |
| TCONS_00107451 | P10 | - |
| TCONS_00111593;TCONS_00111597 | P13;P9 | - |
| TCONS_00107534 | P9 | - |
| TCONS_00197324 | P12 | - |
| TCONS_00107416 | P11 | - |
| TCONS_00107411 | P10 | - |
| TCONS_00111603 | P11 | - |
