1 **Title: EGR4-dependent decrease of UTF1 is associated with failure to reserve**

2 **spermatogonial stem cells in infertile men**

3 Sara Di Persio,[1,#] Tobias Tekath,[2,#] Lara Marie Siebert-Kuss,[1] Jann-Frederik Cremers,[3] Joachim

4 Wistuba,[1] Xiaolin Li,[4] Gerd Meyer zu Hörste,[4] Hannes CA Drexler,[5] Margot Julia Wyrwoll,[3,6]

5 Frank Tüttelmann,[6] Martin Dugas,[2] Sabine Kliesch,[3] Stefan Schlatt,[1] Sandra Laurentino[1,#] and

6 Nina Neuhaus[1,#,*]

7 [1] Centre of Reproductive Medicine and Andrology, University Hospital of Münster, Münster,
8 48149, Germany.

9 [2] Institute of Medical Informatics, University Hospital of Münster, Münster, 48149, Germany.

10 [3] Centre of Reproductive Medicine and Andrology, Department of Clinical and Surgical
11 Andrology, University Hospital of Münster, Münster, 48149, Germany.

12 [4] Department of Neurology with Institute of Translational Neurology, University Hospital of
13 Münster, Münster, 48149, Germany.

14 [5] Bioanalytical Mass Spectrometry Unit, Max Planck Institute for Molecular Biomedicine,
15 Münster, 48149, Germany

16 [6] Institute of Reproductive Genetics, University of Münster, Münster, 48149, Germany.

17 [#] These two first and last authors contributed equally to this work

18 [*] Corresponding author:

19 Nina Neuhaus
20 Centre of Reproductive Medicine and Andrology, University Hospital of Münster, Münster,
21 48149, Germany
22 Tel.: 0251-8356094
23 Nina.Neuhaus@ukmuenster.de

24

25

26

27

28   **One Sentence Summary:** Reserve spermatogonial stem cell depletion in infertile men is

29   regulated by an EGR4-dependent UTF1 decrease, which changes chromatin morphology.

30   **Abstract**

31   Despite the high incidence of male infertility, about 70% of infertile men do not receive a

32   causative diagnosis. To gain insights into the regulatory mechanisms governing human germ

33   cell function in normal and impaired spermatogenesis (cryptozoospermic patients, crypto), we

34   combined single cell RNA sequencing (>30.000 cells), proteome, and histomorphometric

35   analyses of testicular tissues. We found major alterations in the crypto spermatogonial

36   compartment with increased numbers of the most undifferentiated spermatogonia (PIWIL4$^+$

37   State 0 cells). We also observed a transcriptional switch within the spermatogonial

38   compartment driven by the increased and prolonged expression of the transcription factor

39   *EGR4*. Intriguingly, EGR4-regulated genes included the chromatin-associated transcriptional

40   repressor *UTF1*, which was downregulated. Histomorphometrical analyses showed that these

41   transcriptional changes were mirrored at the protein level and accompanied by a change in the

42   chromatin structure of spermatogonia. This resulted in a reduction of A$_{dark}$ spermatogonia -

43   characterized by tightly compacted chromatin and serving as reserve stem cells. These findings

44   suggest that crypto patients are at a disadvantage especially in cases of gonadotoxic damage as

45   they have less cells safeguarding the genetic integrity of the germline. We hypothesize that the

46   more relaxed chromatin status of spermatogonia is dependent on decreased UTF1 expression

47   caused by EGR4 activation. These identified regulators of the spermatogonial compartment will

48   be highly interesting targets to uncover genetic causes of male infertility.

49   **Key words:**

50   Male germline stem cells, spermatogonial stem cells, male infertility, impaired

51   spermatogenesis, single cell RNA sequencing, human spermatogenesis, stem cell

52   differentiation, testis.

Introduction

53

54 Infertility affects 10-15% of couples worldwide, with a male factor in half of the cases *(1)*.

55 Despite this high incidence of male infertility, about 70% of men do not receive a causative but

56 only a descriptive diagnosis of impaired sperm production *(2)*. This knowledge gap needs to be

57 filled so that clinicians can counsel infertile men with regard to causal treatments, chances of

58 sperm retrieval from testicular biopsies, transmission of infertility to the offspring, and potential

59 health risks to the offspring and the males themselves. This latter aspect is of particular interest

60 as male infertility and low sperm counts are associated with an increased risk of cancer,

61 cardiometabolic disease and even premature mortality *(3)*.

62 Of the infertile men, 7% are diagnosed with a severe form of oligoasthenoteratozoospermia

63 termed "cryptozoospermia" (from the greek "kryptós": hidden) as they have less than 0.1

64 million sperm in the ejaculate, rendering natural conception almost impossible *(2)*. The surgical

65 approach of testicular sperm extraction (TESE) often represents the only route to retrieve viable

66 sperm for use in intracytoplasmic sperm injection (ICSI). Lower fertilization as well as

67 implantation rates of transferred embryos *(4)* strongly indicate alterations not only in sperm

68 production but also in the quality of the germ cells themselves. The underlying etiological

69 factors for this severely impaired sperm production often remain unknown at cellular and

70 molecular level but urgently need to be unveiled to improve clinical diagnostics, counselling,

71 and prospectively treatment of infertile men.

72 Spermatogenesis is a highly complex process orchestrated by up to 10% of the male genome

73 and regulated by a delicate interplay between the somatic environment and the germline.

74 Spermatogenesis is supported by the spermatogonial stem cells (SSCs), a subpopulation of

75 diploid spermatogonia, which undergo mitotic divisions, enter meiosis, and ultimately give rise

76 to mature haploid sperm. SSCs are defined based on their functional properties, their ability to

77   self-renew, and to give rise to differentiating germ cells. SSC systems in mice and men show

78   species-specific differences *(5)*. The human SSC system, according to the most accepted model,

79   is classified as a progenitor-buffered system, containing quiescent reserve spermatogonia

80   (A_{dark}) in addition to self-renewing spermatogonia (A_{pale}), the latter generating differentiating

81   spermatogonia (B) *(6–8)*. These three different cell populations are classified based on their

82   morphological appearance and the chromatin structure of their nuclei. In contrast to this, single

83   cell RNA sequencing (scRNA-seq) analyses of the normal human testicular tissues enabled the

84   identification of multiple spermatogonial states *(9–12)* indicative of a SSC compartment that is

85   more heterogeneous than anticipated based on morphological properties. Furthermore, these

86   studies indicated *PIWIL4* and *EGR4* as marker genes of the most undifferentiated

87   spermatogonia *(9, 11)*. As the morphological appearance of these transcriptionally-defined

88   spermatogonia remains largely unknown, it has not been possible so far to link the two

89   classification systems. Importantly, the cellular and molecular alterations of the SSC

90   compartment in stressful conditions such as male infertility, remain to be discovered. This will

91   provide not only valuable information about human SSC biology but might also reveal the

92   mechanisms behind the association of male infertility and cancer or a reduced life span *(3)*.

93   These insights will pave the way to understand the connection between male infertility and

94   general health.

95   Furthermore, infertile men display alterations of the spermatogonial microenvironment, which

96   has been suggested as a regulatory unit for spermatogonial fate decisions. Although the

97   molecular and cellular properties of the human SSC-microenvironment remain largely

98   unknown, an intact interplay between spermatogonia and their microenvironment must be

99   considered a prerequisite for intact spermatogenesis and thereby fertility *(5)*. Key players of

100   the spermatogonial microenvironment include cells in direct physical contact with the

4

101   spermatogonia (e.g. Sertoli and peritubular cells) and cells outside the seminiferous tubules

102   (e.g. Leydig, endothelial, and perivascular cells and macrophages) *(5, 13)*.

103   In this study, we set a framework which combines scRNA-seq, proteomic, histomorphometric,

104   and bioinformatical analyses to unveil the cellular and molecular changes in the spermatogonial

105   compartment and its microenvironment of cryptozoospermic men. We revealed an increase in

106   the number of PIWIL4$^+$ undifferentiated spermatogonia and a depletion of the quiescent A$_{dark}$

107   spermatogonia, considered as reserve stem cells. Moreover, we identified EGR4 as a regulator

108   of these alterations. We found a proinflammatory microenvironment and identified FGF2-

109   FGFR1/3 interaction as a communication channel between advanced germ cells and

110   spermatogonia in cryptozoospermic testes. Intriguingly, our findings highlight alterations of the

111   stem cell compartment and its regulators as an origin of impaired spermatogenesis, representing

112   a milestone in the understanding of pathways underlying male infertility.

113   **Results**

114   **Alterations of the spermatogonial compartment in cryptozoospermic men**

115   To study cellular and molecular changes associated with male infertility, particularly in the

116   spermatogonial compartment, we selected testicular biopsies from cryptozoospermic men

117   (crypto) and normal controls (normal) based on clinical parameters (n=34, table S1). We used

118   a multi-layered approach including quantitative histomorphometrical, scRNA-seq, and

119   proteome analysis (Fig. 1A). Histological quantifications revealed a reduced proportion of

120   tubules containing elongated spermatids as the most advanced germ cell type in the crypto

121   group, while the percentage of tubules with round spermatids, spermatocytes, and

122   spermatogonia remained unchanged (Fig. 1B). To assess the cellular composition of the

123   spermatogonial compartment, we analyzed pan-spermatogonial marker MAGEA4 and

124   undifferentiated spermatogonial marker UTF1 (Fig. 1C). We found a similar number of

125   spermatogonia (MAGEA4$^+$) per tubule in tissues from normal and crypto individuals (n=6

126   each). In contrast, tissues from the crypto group showed a reduction in UTF1$^+$ spermatogonia

127   (Fig. 1D and table S2).

128   In-depth characterization of the molecular changes was performed by scRNA-seq analysis of

129   normal and crypto testicular biopsies (n=3 each). Suitable samples were selected based on

130   ploidy analysis, ensuring presence of haploid spermatids in all samples (Fig. 1E). To exclude

131   genetic causes for male infertility we performed whole exome sequencing analysis. No relevant

132   variants were identified in any of the three cryptozoospermic men rendering the potential

133   genetic cause for their infertility unidentifiable for the time being and making a uniform cause

134   highly unlikely. Following scRNA-seq and quality control, data from 15 546 and 13 144 cells

135   were ultimately included in the analysis for normal and crypto samples, respectively (table S3).

136   Unsupervised clustering resulted in a comparable number of clusters (30 versus 29) between

137   the two groups (fig. S1, A and B). To assign identities to the clusters while still considering

138   eventual biological differences between the groups, we used a two-step approach. Firstly, we

139   used 55 published marker genes to identify 15 cell types in the normal dataset (fig. S1B) *(9–*

140   *12)*. Notably, the latent space replicated the known differentiation process from spermatogonia

141   to late spermatids (spiral-like shape in Fig. 1F). Using this as reference, we projected the same

142   labels *(14)* onto the crypto dataset and found the expected cell types represented (Fig. 1G).

143   Correct cluster assignment was confirmed by evaluating the expression of specific marker genes

144   in the crypto dataset (fig. S1C). Importantly, all samples contributed to all cell type clusters in

145   each dataset (Fig. 1, F and G). Additionally, using marker gene expression, we identified two

146   additional clusters in the crypto group without equivalence in the normal dataset namely,

147   perivascular cells (cluster 7) and immune cells (cluster 24; Fig. 1G).

148   **Cellular and transcriptional changes in germ and somatic cell compartments**

6

149   To explore the similarities and differences between the two groups, we combined all normal

150   and crypto datasets (Fig. 2A). Consistent with the histomorphometrical analysis (Fig. 1D)

151   evaluation of the scRNA-seq dataset showed comparable percentages of spermatogonia in both

152   groups (Fig. 2B). In fact, percentages were similar from spermatogonia up to zygotene

153   spermatocytes (Fig. 2B). However, from the pachytene spermatocyte stage onwards, the crypto

154   group displayed a striking reduction in germ cells (Fig. 2B). This was corroborated by parallel

155   quantitative proteomic analysis, showing comparable levels of spermatogonial marker

156   MAGEA4 and reduced expression of markers for leptotene/pachytene spermatocytes (DPH7,

157   PIWIL1) as well as spermatids (PRM2; Fig. 2C). To assess the changes at the transcriptional

158   level between all the identified cell types, we performed differential gene expression (DGE)

159   analysis (table S4) and identified genes uniquely differentially expressed in each germ cell stage

160   (fig. S2 and table S5). Interestingly, *UTF1* and *EGR4* were down- and up- regulated

161   respectively in the crypto undifferentiated spermatogonia (fig. S2A). Curiously, genes

162   associated with the gene ontology (GO) terms "Cell cycle", "Chromosome organization",

163   "DNA repair", and "Telomere organization" (e.g. *BRCA2*, *CENPA*, *SMC1A*, *SYCP3*, *MEIOB*,

164   *CETN2*) were differentially expressed in differentiating spermatogonia compared to pachytene

165   spermatocytes (fig. S2, B to E).

166   In the somatic cell compartment, the most striking difference was the increase in the proportions

167   of peritubular myoid cells (PMCs), fibrotic PMCs, and macrophages (Fig. 2B); the latter finding

168   in particular was supported by proteomic data, which revealed significantly increased

169   macrophage marker LYZ expression (Fig. 2C). Among somatic cells, PMCs showed the most

170   profound changes in DGE analysis, with 83 down- and 149 up-regulated genes, respectively, in

171   the crypto group. GO analysis revealed an enrichment of genes involved in "extracellular matrix

172   components and organization" (fig. S3A and table S5). Because perivascular and immune cells

173   did not have equivalents in the normal dataset, no direct comparison was conducted between

7

174   normal and crypto samples (fig. S3B). DGE analysis between perivascular cells and PMCs in

175   the crypto dataset revealed 48 genes (fig. S3C and table S6) regulating muscle contraction (fig.

176   S3D). To localize perivascular cells at tissue level we selected the musculoskeletal marker

177   *MUSTN1* from the upregulated genes, which was exclusively expressed in this cluster at RNA

178   level (Fig. 2D). MUSTN1 was specifically expressed in blood vessels, where it co-localized

179   with ACTA2. Indeed, these ACTA2$^+$/MUSTN1$^+$ cells surrounded VWF$^+$ endothelial cells (Fig.

180   2E). Quantification of the total number of blood vessels/mm^2 revealed a significantly higher

181   number of blood vessels in crypto samples (Fig. 2F and table S2). Similar results were obtained

182   for the number of MUSTN1$^+$ blood vessels/mm^2 (Fig. 2F and table S2). We therefore concluded

183   that MUSTN1 is a specific perivascular cell marker and that an increased proportion of

184   MUSTN1$^+$ blood vessels is a specific feature of crypto samples. DGE analysis comparing

185   immune cells to all other crypto cells resulted in 195 differentially expressed genes, including

186   T-cell marker *CD3E* (Fig. 2G, fig. S3E and table S6). To evaluate which types of immune cells

187   were present in crypto testes, we performed pathway analysis and found involvement of

188   cytotoxic T lymphocyte pathways (fig. S3F). Immunofluorescence analysis localized CD3$^+$ T

189   cells in the vicinity of blood vessels in crypto samples but not in the normal (Fig. 2H). To

190   investigate the crosstalk between spermatogonia and their microenvironment, we used

191   CellphoneDB *(15)*. We found 50 significant ligand-receptor interactions considering both

192   datasets (Fig. 3). A significant interaction was detected between the spermatogonial-based

193   receptors *FGFR1* and *3* and the ligand *FGF2* produced by pachytene and diplotene

194   spermatocytes in the crypto group. Moreover, spermatogonial-located *ACKR2* showed

195   significant interaction with its ligands secreted by endothelial cells, perivascular cells,

196   macrophages, and immune cells (*CCL2*, *CCL3*, *CCL4*, *CCL5, CCL3L1* and *CCL14*) in the

197   crypto datasets but not in the normal.

198   **Crypto undifferentiated spermatogonia activate the EGR4 regulatory network**

8

199   SCENIC was used to identify the gene regulatory network changes in the crypto somatic and

200   germ cell populations *(16)*. This pipeline identifies groups of genes, named regulons, co-

201   regulated by a transcription factor. In total, in the two datasets we identified 403 regulons, each

202   containing up to 2 450 genes (table S7). To identify differentially activated regulons between

203   the normal and the crypto dataset, we selected the 15 regulons with the highest specificity score

204   for each cluster in the normal dataset. We then focused on those regulons in the crypto dataset

205   that showed at least 20% change in the proportion of cells per cluster in which that specific

206   regulon was active (Fig. 4A and table S8). Comparative evaluation of the binarized AUC (area

207   under the curve) score of each regulon revealed 13 regulatory networks that were differentially

208   activated between the two datasets (bold in Fig. 4A). As perivascular and immune cells were

209   uniquely represented in the crypto dataset, they were not included in this analysis. In the somatic

210   compartment, although PMC and fibrotic PMC clusters presented the most striking differences

211   in cell proportions, no regulon was differentially activated in these cells. Similarly, we could

212   not find differentially activated regulons for Leydig cells. However, one differentially activated

213   regulon was found in the macrophages in the crypto dataset (activation of RUNX3, 75 genes,

214   fig. S4).

215   Contrary to these minor changes in somatic cells, germ cells showed deeper alterations in their

216   gene regulatory networks. Undifferentiated spermatogonia displayed four regulons with

217   significantly enhanced activation in the crypto samples: ASCL2 (88 genes), EGR4 (594 genes),

218   HOXC9 (139 genes), and DLX5 (104 genes) (Fig. 4B-E). Intriguingly, among the genes

219   regulated by EGR4, we identified *ASCL2*, *HOXC9*, and *DLX5* whereas none of the other

220   transcription factors seemed to regulate the expression of *EGR4*. Comparison between the

221   regulon and the DGE analysis of the undifferentiated spermatogonia showed that 23 of the

222   differentially expressed genes were predicted to be regulated by EGR4 (Fig. 4, F and G).

223   Follow-up histological analysis showed specific expression of EGR4 in spermatogonia (Fig.

224   4H). Moreover, the E2F1 regulon (685 genes) showed a significant activation in differentiating

9

225   spermatogonia (fig. S4A). While no regulons in the crypto dataset were differentially active

226   specifically in spermatocytes, early spermatids showed activation of RORA (fig. S4B), and late

227   spermatids showed deactivation of STAT3, TBX4, ATF2, and RFX2, and activation of ETV2

228   (fig. S4, C to G).

229   **Crypto spermatogonial compartment shows an increased number of PIWIL4$^+$**

230   **spermatogonia**

231   To further scrutinize the changes in the spermatogonial compartment, we identified specific

232   spermatogonial subtypes based on marker genes reported in existing single cell studies *(9, 11)*.

233   For this, we subset and re-clustered undifferentiated and differentiating spermatogonial clusters

234   from both datasets (fig. S5A). Upon analyzing the expression of the different published markers

235   in the normal dataset, we could assign the clusters and identify six spermatogonial states: State

236   0 (*PIWIL4, PHDGH, EGR4*), State 0A (*UTF1, SERPINE2, FGFR3*), State 0B (*NANOS2*), State

237   1 (*GFRA1, GFRA2, NANOS3, ID4*), State 2 (*KIT, MKI67, DMRT1, DNMT1*), and State 3

238   (*STRA, SYCP3*) (fig. S5B and Fig. 5A). Notably, the markers mentioned display the highest,

239   but not unique, expression levels in the respective states. Using the data from normal

240   spermatogenesis as reference, these labels were projected onto the crypto dataset (Fig. 5B). For

241   further analyses, both datasets were integrated (Fig. 5C). To compare our assignment with those

242   published by Guo et al., and Sohni et al., we performed correlation analysis among the three

243   datasets *(9, 11)*. States 0, 0A and 0B in our assignment showed the highest correlation with

244   State 0 in Guo et al. (fig. S5C) and with SSC1-B,-A, and -C respectively in Sohni et al. (fig.

245   S5D).

246   We assessed the cellular composition of the spermatogonial compartment and found an

247   increased number of cells in State 0 and a significantly decreased proportion of cells in State

248   0A in the crypto dataset. There were no differences in cell numbers in any other states (Fig.

249   5D).

250   To reveal whether this shift in spermatogonial states was also represented at the cellular level,

251   and considering that the number of UTF1$^+$ spermatogonia is significantly decreased (Fig. 1, C

252   and D), we analyzed *PIWIL4*, *FGFR3*, *GFRA1,* and *NANOS3* (Fig. 5, E and F) using

253   quantitative histomorphometric analysis. The altered proportions of the spermatogonial states

254   identified based on scRNA-seq data were largely corroborated by cell quantification. We found

255   a significant increase in PIWIL4$^+$ spermatogonia (State 0) per seminiferous tubule, and no

256   significant changes in FGFR3$^+$, GFRA1$^+$, and NANOS3$^+$ spermatogonia (Fig. 5G and table S2).

257   To further unravel the alterations of the spermatogonial compartment, RNA velocity analysis

258   *(17)* was performed setting State 0 as the origin of the spermatogenic differentiation process.

259   *(9)* The velocity streamlines (which depend on the mRNA maturation in each cell) indicate that

260   States 0 and 0A in the normal samples are transcriptionally not directed toward State 1, unlike

261   State 0B (Fig. 4I). In contrast, in the crypto samples the velocity streamlines in States 0, 0A,

262   and 0B appear to be preferentially oriented toward State 1 (Fig. 5H). This was evident in all

263   three crypto samples (fig. S5E).

264   **The crypto spermatogonial compartment shows reduced numbers of A$_{dark}$ spermatogonia**

265   To evaluate the expression profile of the morphologically-defined A$_{dark}$ spermatogonia, we

266   assessed how many A$_{dark}$ spermatogonia were PIWIL4$^+$, UTF1$^+$, FGFR3$^+$, GFRA1$^+$, or

267   NANOS3$^+$. Semi-quantitative evaluation in tissues showed the majority of A$_{dark}$ spermatogonia

268   in the control group were UTF1$^+$ (78.4%), whereas only 5.8% were PIWIL4$^+$, 5.7% were

269   FGFR3$^+$, 18.1% were GFRA1$^+$, and no A$_{dark}$ cells were NANOS3$^+$ (Fig. 6, A to D).

270   We then asked whether the percentage of A$_{dark}$ spermatogonia is altered in the crypto samples.

271   Assessment of the A$_{dark}$ spermatogonia among all spermatogonia (MAGEA4$^+$) indeed unveiled

272   a significant reduction of this cell type in crypto samples (Fig. 6E). Specifically, we found a

273   general reduction in A$_{dark}$ cells among the PIWIL4$^+$, UTF1$^+$, FGFR3$^+$, as well as GFRA1$^+$

274   spermatogonia (Fig. 6, F to I). Finally, we questioned whether the A$_{dark}$ reduction was

275   dependent on the most advanced germ cell stage present in the respective tubules.

276   Quantifications revealed that the number of A_{dark} spermatogonia in the crypto samples was

277   reduced independently of the spermatogenic state of the tubules (Fig. 6J), suggesting that the

278   loss of this cell type is a general process in crypto tissues.

279   **Altered EGR4 and HOXC9 expression in crypto State 0-0A spermatogonia**

280   To assess potential changes in gene expression observed in the undifferentiated spermatogonia

281   of crypto samples, we employed trajectory-based differential expression analysis (tradeSeq)

282   *(18)*. For this, cells were aligned along the latent time (fig. S6A) and then subdivided into six

283   groups containing equal cell numbers (knotgroups) using seven knots (fig. S6B). We focused

284   our attention on the most undifferentiated spermatogonia, State 0 and 0A, which are mostly

285   represented by knotgroups 1 and 2.

286   Comparing knotgroup 1, which includes 78.2% of State 0 spermatogonia, between the normal

287   and crypto datasets revealed 21 up- and 12 downregulated genes (Fig. 7A) involved in

288   regulating cellular development and differentiation (Fig. 7B); among those genes, we found

289   *EGR4* itself and 16 genes belonging to the EGR4 regulon (*ANXA11*, *CEP131*, *CRIP2*, *EGR4*,

290   *IDH2*, *NUDT14*, *PDLIM4*, *PLPP2*, *PODXL2*, *PSIP1*, *RAC3*, *SH2B2*, *SMARCD3*, *SPINT1*,

291   *ST3GAL4,* and importantly *UTF1*) (Fig. 7C and fig. S6D). We had already identified an

292   enhanced activation of the EGR4 regulon in undifferentiated spermatogonia of the crypto

293   samples (Fig. 4C), highlighting this regulon and gene as important regulators of spermatogonia.

294   We then analyzed its expression in the spermatogonial states and found that, while expression

295   of *EGR4* was restricted to State 0 in the normal situation, in the crypto dataset it showed a

296   prolonged expression, being present also in States 0A and 0B (Fig. 7D). Afterwards, the

297   expression of all EGR4-regulated genes was examined according to the spermatogonial states,

298 confirming their modulation (fig. S6D). Interestingly, in the crypto group, we found a reduced

299 expression of *UTF1* in State 0, 0A, and 0B spermatogonia (Fig. 7D).

300 For knotgroup 2, representing mostly State 0A cells, we identified 82 upregulated genes in the

301 crypto cells (Fig. 7E), most of which are involved in cell differentiation (Fig. 7F). Again, *EGR4*

302 was one of the differentially expressed genes, together with *HOXC9*, which we had already

303 identified in the crypto samples as an activated regulon in undifferentiated spermatogonia (Fig.

304 4D). Among the differentially expressed genes in knotgroup 2, we found 25 genes that are

305 regulated only by EGR4 (Fig. 7G, fig. S6C and fig. S7A), 1 gene that is regulated only by

306 HOXC9 (Fig. 7H), and 5 genes that are regulated by both transcription factors, including

307 *HOXC9* itself (Fig. 7I and fig. S7C). Among the 25 EGR4-regulated genes in knotgroup 2, 12

308 were also upregulated in knotgroup 1 (fig. S6C).

309 Assessing the expression of these genes according to the spermatogonial states, we observed

310 that, similar to *EGR4*, *HOXC9* also displayed a prolonged expression in the crypto samples

311 compared to the normal situation, where it is exclusively expressed in State 0 (Fig. 7L). This

312 altered expression pattern was also observed for additional genes regulated by one or both of

313 these transcription factors (Fig. 7, J to L and fig. S7, B to D), indicating that crypto samples

314 show persistent expression of State 0 genes in undifferentiated spermatogonia (State 0 through

315 1).

Discussion

317 In this work we pioneered the study of human male germ cell defects using state-of-the-art

318 approaches. These led us to identify novel target genes and mechanisms for explaining the

319 etiology of infertility. At cellular level we unveiled major alterations of the spermatogonial

320 compartment with an increase in the number of the most undifferentiated (PIWIL4$^+$)

321 spermatogonia but a reduction in the morphologically defined A$_{dark}$ reserve stem cells.

13

322   Interestingly, these alterations are contingent on a prolonged expression of EGR4 in crypto

323   samples. Also, we identified receptor-ligand interactions modulating the interplay between

324   spermatogonia and their microenvironment.

325   ScRNA-seq analyses revealed six spermatogonial states based on transcriptional profile, in line

326   with published data *(11)*. Interestingly, we found a representation of the same states in both

327   normal and crypto groups. Although the total number of spermatogonia remained unchanged,

328   the states were represented in different proportions in the two groups. The most prominent

329   finding was an increase, in the crypto group, of PIWIL4$^+$ State 0 cells, which have been

330   suggested as the origin of the spermatogonial differentiation process *(9, 11, 19)*. The increase

331   of PIWIL4$^+$ cells occurred concomitantly with a reduction of the UTF1$^+$ cells, which we

332   observed both at transcriptional and protein level. Moreover, we identified changes in

333   transcriptional dynamics of crypto spermatogonia, as visualized by cells in State 0, uniformly

334   pointing towards the direction of 0A and 0B in the RNA velocity analysis. As, PIWIL4$^+$ and

335   UTF1$^+$ spermatogonia are mostly quiescent *(9, 11, 20, 21)*, we hypothesize that these altered

336   cell proportions are a consequence of a change in transcriptional profiles of spermatogonial

337   subpopulations and not due to changes in cell proliferation (Fig. 8A). We speculate that this

338   increase in PIWIL4$^+$ spermatogonia is the consequence of the reduced spermatogenic efficiency

339   in cryptozoospermic men, however we cannot exclude that the failure to progress to more

340   differentiated states might be the cause.

341   We identified the transcription factor EGR4 as one of the gatekeepers regulating the change in

342   transcriptional profiles of spermatogonia in the crypto group. Supporting this hypothesis we

343   found EGR4 exclusively expressed in spermatogonia at protein level and in the most

344   undifferentiated spermatogonial state (State 0) at transcriptional level, in accordance with recent

345   studies *(9, 11)*. Importantly, trajectory-based differential expression analysis showed prolonged

346   expression of EGR4 in spermatogonial States 0, 0A, and 0B in crypto samples. Moreover, we

14

347 found enhanced activation of the EGR4, ASCL2, HOXC9 and DLX5 regulons in

348 undifferentiated spermatogonia of crypto tissues. These four genes, including *EGR4* itself,

349 contain a putative binding site for EGR4 in their promoter region suggesting EGR4 as an

350 upstream regulator. The crucial role of EGR4 in the germline is corroborated by data

351 demonstrating that it controls multiple differentiation steps during mouse spermatogenesis *(22)*.

352 Furthermore, *EGR4* gene sequence variations as well as altered gene expression profiles have

353 been identified in men with impaired spermatogenesis *(23, 24)*. Importantly, EGR4, ASCL2,

354 HOXC9 and DLX5 have also been assigned pivotal roles in other cell systems. Interestingly, in

355 small cell lung cancer EGR4 regulates proliferation through induction of DLX5 expression

356 among other genes *(25)*. ASCL2 activation is involved in restoring the intestinal stem cell

357 system following damage *(26)* while HOXC9 and DLX5 are involved in neuronal and

358 osteogenic cell differentiation processes, respectively *(27, 28)*. These genes are highly relevant

359 for cell proliferation and differentiation processes in multiple stem cell systems and therefore

360 dysfunction in these pathways may present a link between the germline and other cell systems

361 crucial for general health.

362 One of the genes presenting a putative binding site for EGR4 is *UTF1*, whose expression

363 negatively correlates with that of *EGR4*. UTF1 is a chromatin-associated protein expressed in

364 human undifferentiated, quiescent spermatogonia *(20, 21)*. It is involved in embryonal stem

365 (ES) cell differentiation *(29)* and acts as transcriptional repressor *(30)*. Knock-down of UTF1

366 in ES cells results in extensive chromatin decondensation *(31)*. Remarkably, we found a general

367 reduction of the A_{dark} spermatogonia - characterized by highly condensed chromatin *(6, 7)* - in

368 the crypto group. Marker profiling of A_{dark} spermatogonia shows that almost 80% are UTF1$^+$.

369 We therefore hypothesize that the reduction in UTF1 expression leads to a change in chromatin

370 condensation resulting in the transition from an A_{dark} to an A_{pale} morphology (Fig. 8B). Thereby

371 at cellular level, these cells with a more open chromatin are maintained in a 'ready to react'

372 status. The price of this change in operating status is the reduction of the reserve stem cells

15

373   (A_{dark}), which under normal conditions ensure the recovery of sperm production following a

374   gonadotoxic insult, by their ability to transform into mitotically active A_{pale} spermatogonia *(32)*.

375   This alteration of the stem cell pool may therefore place crypto patients at a disadvantage in the

376   case of gonadotoxic insult. Moreover, the over-recruitment of the reserve spermatogonia into

377   mitotically active spermatogonia may not be sustainable in the long term and is likely to lead

378   to further quantitative and qualitative deterioration of germ cells.

379   The A_{dark} morphology is shared by spermatogonial subpopulations characterized by the

380   expression of PIWIL4, FGFR3, and GFRA1, which define distinct states at transcriptional level

381   (Fig 8A). Intriguingly, only 5.8% of the PIWIL4$^+$ cells showed A_{dark} morphology. These

382   findings indicate that multiple spermatogonial states are able to display A_{dark} spermatogonial

383   morphology, in line with previous data *(33)*, and also that A_{dark} spermatogonia may not be the

384   most undifferentiated spermatogonia in the human testis. This presents a paradigm shift in the

385   field.

386   The fact that the stem cell potential of the A_{dark} spermatogonia will remain untested until it is

387   possible to purify these cells for use in transplantation assays in primates (including the human)

388   constitutes a limitation of this study. Unfortunately, due to profound species differences,

389   including the lack of A_{dark} reserve stem cells in rodents, these cannot be used as a model for

390   mechanistic studies.

391   The microenvironment is pivotal for SSC behavior *(34, 35)*. We identified multiple changes to

392   the SSC microenvironment in the crypto group. Importantly, PMCs, perivascular cells,

393   macrophages, and immune cells showed increased numbers and enhanced expression of

394   chemokines. These changes suggest the presence of a proinflammatory microenvironment in

395   crypto testes. In addition to the changes in the somatic compartment, we identified a reduced

396   number of spermatocytes from the pachytene stage onwards in crypto samples. Intriguingly,

397   this was accompanied by an enhanced interaction between FGF2 - produced by pachytene and

398   diplotene spermatocytes - and FGFR1 and 3 - present on spermatogonia. These interactions

16

399 were solely significant in the crypto group. The effect of FGF2 on human spermatogonia

400 remains largely unknown, however a pro-survival role for FGF2 has been shown *(36)*. Gain of

401 function mutations in FGFR2 and 3 have been found in spermatogonia supporting the survival

402 of these cells and inducing their clonal expansion *(37)*. We suggest that FGF2-FGFR3 pathway

403 may be relevant for maintaining the size of the human spermatogonial compartment in the

404 crypto situation.

405 The revelation of regulators of the human spermatogonial stem cell compartment and its

406 properties provides enticing possibilities. We recommend screening the genes highlighted in

407 this manuscript for pathogenic variants. This may help to decipher genetic defects affecting

408 spermatogonia, and thus help to resolve additional cases of male infertility. Finally, this will

409 result in improved counselling of infertile men about the chances for sperm retrieval prior to

410 surgery. In addition, it will help avoid unnecessary surgical interventions, identify potential

411 health risks for their offspring, and establish preventive measures regarding their own general

412 health risks. Lastly, due to the potential long term exhaustion of the stem cell pool crypto

413 patients may prospectively benefit from cryopreservation of sperm prior to active family

414 planning.

415 **Materials and Methods**

416 **Human testicular biopsies**

417 Adult human testis samples (n=34) were obtained from patients undergoing surgery for

418 microdissection testicular sperm extraction (mTESE) or histological evaluation at the

419 Department of Clinical and Surgical Andrology (University Hospital in Münster, Germany).

420 One additional testicular sample was excised from each patient for this study after written

421 informed consent. Ethical approval was obtained for this study (Ethics Committee of the

422 Medical Faculty of Münster and State Medical Board no. 2008-090-f-S).

Patient cohort selection

We subjected six testicular tissues (normal=3; crypto=3) to scRNA-seq. We selected patient samples with qualitatively and quantitatively normal spermatogenesis (normal), and with cryptozoospermia (crypto). Clinical workup prior to surgery included full physical evaluation, hormonal analysis (including luteinizing hormone (LH), follicle stimulating hormone (FSH) and testosterone (T)) *(38)*, semen analysis *(39)*, and genetic analyses (including karyotype and screening for azoospermia factor (AZF) deletions). Exclusion criteria were known genetic causes of infertility, acute infections, testicular tumors, and a history of cryptorchidism. Normal patients were selected from those diagnosed with obstructive azoospermia. These patients had no sperm in their ejaculate, normal testicular volume, and normal FSH levels. Age-matched cryptozoospermic patients had a sperm concentration <0.1 million/ml (cryptozoospermia) in the ejaculate, a reduced testicular size, and, in most cases, elevated FSH levels. All patients had spermatozoa in their TESE samples, regardless of histological results. Detailed clinical information about all the patients is available in table S1. Moreover, we added independent samples for proteome and histological analyses to further expand the sample cohort. Due to sample limitation, only 2 of the 3 crypto samples could be included in the proteome analysis.

Exome sequencing

The exomes of the three cryptozoospermic men were sequenced. Briefly, target enrichment was performed by SureSelect QXT Target Enrichment Kit according to the manufacturer's protocol using the capture libraries Agilent SureSelect Human All Exon Kit V6. Sequencing was performed on the Illumina NextSeq®500 system. The data was evaluated for rare (minor allele frequency [MAF] in gnomAD <0.01), likely pathogenic variants (stop-, frameshift- and splice site-variants) in 170 candidate genes previously reported to be associated with impaired spermatogenesis according to Oud et al. *(40)*. (for gene list and detected variants see table S9).

447   Additionally, the recently published genes ADAD2, M1AP, MSH4, RAD21L1, RNF212,

448   SHOC1, STAG3, SYCP2, associated with non-obstructive azoospermia, were screened. Only

449   variants with a coverage >10 and detected in genes that quantitatively impair spermatogenesis

450   were considered for further analyses. Variants in recessive genes were only evaluated if at least

451   two were identified in the same individual. For putative dominant and X-linked genes, variants

452   that were also found in fertile controls were not considered for further analyses.

453   **Preparation of single cell suspensions**

454   Single cell suspensions were obtained using a two-step enzymatic digestion as previously

455   described *(41, 42)*. A total of 20 000 cells were used for ploidy analysis, 12 000 cells were

456   prepared for scRNA-seq and 50 000 cells were stored at -80°C for proteomic analysis.

457   **Ploidy analysis**

458   Ploidy analysis was performed on 20 000 cells of each sample prior to scRNA-seq. After

459   centrifugation, the cells were incubated in the dark for 30 minutes with a solution containing

460   50 µg/ml Propidium Iodide (Sigma- Aldrich, Cat# P4170), 1 mg/ml bovine serum albumin

461   (Sigma-Aldrich, Cat# A9647), 0.1% Triton X-100 (Sigma-Aldrich, Cat# 93443), and 10 µg/ml

462   RNase A (Sigma-Aldrich, Cat# R6513) in phosphate-buffered saline (PBS). After incubation,

463   approximately 10 000 cells were analyzed for each sample using a Beckman Coulter CytExpert

464   QC Flow Cytometer. Debris was defined based on forward and side scatter and excluded from

465   the analysis. The cells analyzed for DNA content were measured at 617 nm and assigned to the

466   following categories according to staining intensity: 1) haploid cells (1C: spermatids), 2)

467   diploid cells (2C: spermatogonia, somatic cells (e.g. Leydig peritubular and Sertoli cells)), 3)

468   "double diploid" cells (4C: spermatocytes).

469   **ScRNA-seq library preparation and sequencing**

19

470   12 000 cells per sample (normal=3; crypto=3) were suspended in MEMα at a concentration of

471   500 cells/µl and loaded onto the Chromium Single Cell A Chip. Library preparation was

472   performed following the kit instructions (Chromium Single Cell Kit [v2 chemistry]). Briefly

473   ~6 000 cells per sample were captured by the 10x Genomics Chromium controller, after cDNA

474   synthesis, 12-14 cycles were used for library amplification. The resulting libraries were

475   quantified using the Lab901 TapeStation system (Agilent) before shallow sequencing

476   (NextSeq-550 sequencer, Illumina) for quality control. The final, deep sequencing was

477   performed on a NovaSeq6 000 sequencer (Illumina), using 2×150 bp paired-end sequencing.

478   **ScRNA-seq and proteome analyses**

479   The scRNA-seq and proteome analyses are described in detail Supplementary Materials and

480   Methods.

481   **PAS, immunohistochemical and immunofluorescence staining**

482   Testicular biopsies were fixed in Bouin's solution overnight and then washed in 70% ethanol.

483   The tissues were paraffin-embedded and sectioned at 5 µm. Tissue sections were dewaxed with

484   AppiClear (Applichem, Cat# A4632.2500), rehydrated through decreasing ethanol

485   concentrations, and washed in distilled water. For periodic acid-Schiff/hematoxylin (PAS)

486   staining, the slides were first incubated in 1% PA (Sigma-Aldrich, Cat# 1.005.240.100) and

487   then in Schiff's reagent (Sigma-Aldrich, Cat# 1.090.330.500). Cell nuclei were counterstained

488   with Mayer's hematoxylin solution (Sigma-Aldrich, Cat# 1.092.490.500). After rinsing in tap

489   water, the slides were dehydrated through increasing ethanol concentrations, incubated in

490   AppiClear solution, and closed with Merckoglas (Sigma-Aldrich, Cat# 1.039.730.001).

491   Immunohistochemical staining was performed as previously described *(42)*. Briefly, after

492   rehydration, the sections underwent heat-induced antigen retrieval using sodium citrate buffer

493   pH 6.0. The endogenous peroxidase activity and the unspecific antibody binding were blocked

494 using hydrogen peroxide (Hedinger, Cat# GH06201) and a solution containing goat serum

495 (Sigma-Aldrich, Cat# G6767-100ML) and BSA, respectively. The sections were incubated

496 overnight at 4°C with the primary antibodies diluted in the blocking solution. The following

497 day, sections were incubated with a biotin-labeled secondary antibody and then with

498 streptavidin-horseradish peroxidase. The peroxidase activity was detected using 3,30-

499 diaminobenzidine tetrahydrochloride solution (Applichem, Cat# A0596.0001). The reaction

500 was stopped with distilled water and nuclei were counterstained with Mayer's hematoxylin.

501 Finally, the slides were rehydrated with increasing ethanol concentrations, washed with

502 AppiClear, and closed with Merckoglas. The sections were scanned with Precipoint M8

503 Microscope and Scanner (Precipoint, Freising, Germany) *(43)*.

504 The immunofluorescent staining was performed as previously described *(21)*. After

505 rehydration, the sections underwent heat-induced antigen retrieval using sodium citrate buffer

506 pH 6.0. After cooling to room temperature the tissues were permeabilized with Triton X-100,

507 incubated with 1M glycine (Sigma-Aldrich, Cat# G7126-500G) and then covered with a

508 blocking solution of tween (Sigma-Aldrich, Cat# 655205), BSA, and donkey serum (Sigma-

509 Aldrich, Cat# S30-100 ml). The incubation of the primary antibody was performed overnight

510 at 4°C in blocking solution. The following day the sections were washed and incubated for 1

511 hour with species specific secondary antibodies diluted in blocking solution. Slides were finally

512 mounted with Vectashield Mounting Media with 4,6-diamidino-2-phenylindole as nuclear

513 counterstain (Vector Laboratories, Cat# H-1200). Stainings were digitalized employing the

514 Olympus BX61VS microscope and scanner software VS-ASW-S6 (Olympus, Hamburg,

515 Germany). All antibodies are listed in table S10.

516 **Evaluation of the testicular histology**

21

517 To score the spermatogenic status of each patient two PAS-stained sections from two

518 independent biopsies per testis were evaluated using the Bergmann and Kliesch method *(44)*.

519 A score from 0 to 10 was assigned to each patient according to the percentage of tubules

520 containing elongated spermatids. In the same sections, the percentage of seminiferous tubules

521 containing spermatocytes or spermatogonia as the most advanced germ cell type was assessed,

522 as well as Sertoli cell only and hyalinized tubules (tubular shadows).

523 **A_{dark}, A_{pale}, B spermatogonia identification**

524 The A_{dark}, A_{pale} and B spermatogonia were defined according to morphological criteria as

525 previously published *(6, 7)*. Briefly, the A_{dark} spermatogonia presented a spherical or slightly

526 ovoid nucleus containing uniformly dark stained chromatin with an unstained central

527 rarefaction zone. A_{pale} spermatogonia had an ovoid nucleus containing very weakly stained

528 chromatin and 1-3 deeply stained nucleoli attached to the nuclear membrane. Finally, B

529 spermatogonia showed a spherical nucleus with granulated and weakly stained chromatin and

530 intensely stained nucleoli detached from the nuclear membrane.

531 **Histological quantifications and statistics**

532 Quantification of stained cells following immunofluorescence analyses was performed with

533 Fiji/ImageJ *(45)*. The number of ACTA2$^+$ and MUSTN1$^+$ blood vessels was evaluated in two

534 independent sections of normal (n=12) and crypto (n=13) patients and normalized per area

535 (mm^2). To discern between the ACTA2$^+$ seminiferous tubules and the ACTA2$^+$ blood vessels

536 for each stained section a consecutive section was stained with the endothelial marker VWF. A

537 blood vessel was defined as such provided that it was positive for VWF in the consecutive

538 section.

539 Each immunohistochemical quantification was performed on normal (n=6) and crypto (n=6)

540 samples using the Precipoint Viewpoint software (Precipoint, Freising, Germany). The number

541   of MAGEA4$^+$, PIWIL4$^+$, FGFR3$^+$, GFRA1$^+$, NANOS3$^+$ and UTF1$^+$ spermatogonia per tubule

542   was assessed per round tubule. Tubules were considered round when the ratio between the two

543   diameters was in the range of 1 - 1.5. The percentage of PIWIL4$^+$, FGFR3$^+$, GFRA1$^+$,

544   NANOS3$^+$ and UTF1$^+$ A$_{dark}$ spermatogonia was evaluated counting 100 A$_{dark}$ spermatogonia

545   per sample and determining the proportion of each marker. In case of the crypto samples it was

546   not always possible to reach 100 A$_{dark}$ spermatogonia. To evaluate how many MAGEA4$^+$,

547   PIWIL4$^+$, FGFR3$^+$, GFRA1$^+$ and UTF1$^+$ spermatogonia were A$_{dark}$, we counted 100

548   spermatogonia positive for each marker in each sample and then assessed the proportion of

549   A$_{dark}$ spermatogonia. Tubules of the crypto samples were subdivided into three categories

550   according to the most advanced germ cell type present: spermatogonia, spermatocytes and

551   spermatids. All the quantification results were plotted as box plots (center line: median; box

552   limits: upper and lower quartiles; whiskers: 1.5× interquartile range; points: outliers). Normality

553   and homoscedasticity tests were performed for all variables and differences between groups

554   were assessed by parametric (t-test) or non-parametric tests (Mann-Whitney U test) as

555   appropriate. Kruskal-Wallis rank sum test was used to compare three or more independent

556   groups, followed by multiple pairwise comparisons. Statistical analysis and graphs were

557   executed using R 4.0.0 and packages stats (v4.0.0) and ggplot2 (v3.3.1). Details regarding the

558   number of samples or cells evaluated and the statistical analysis are provided in table S2.

559   **Supplementary material**

560   Supplementary Materials and Methods

561   Fig. S1. Clustering analysis of normal and crypto datasets.

562   Fig. S2. DGE and GO analysis between the different germ cell clusters of the normal and crypto

563   datasets.

564   Fig. S3. DGE and GO analysis between the different somatic cell clusters of the normal and

565   crypto datasets.

566   Fig. S4. Differential regulon activation in the normal and crypto datasets.

567   Fig. S5. Cluster analysis and assignment of normal and crypto spermatogonial datasets.

568   Fig. S6. EGR4-regulated genes in the knotgroup 1 spermatogonia.

569   Fig. S7. EGR4- and HOXC9-regulated genes in knotgroup 2 spermatogonia.

570   Table S1. Clinical parameters.

571   Table S2. Histomorphometric and statistical analysis.

572   Table S3. Single cell RNA sequencing stats.

573   Table S4. Differential expression analysis results.

574   Table S5. Genes uniquely differentially expressed in germ or somatic cell clusters and relative

575   gene ontology analysis.

576   Table S6. Differential gene expression analysis of perivascular cells compared toPMCs and of

577   immune cells compared to all other cell clusters.

578   Table S7. List of the 403 regulons obtained from the Scenic analysis.

579   Table S8. List of regulons obtained from the Scenic analysis plotted in the heatmap in Fig. 4.

580   Table S9. Genes with at least one potentially pathogenic variant described according to Oud et

581   al. 2019.

582   Table S10. Reagents & Software.

583 **References**

584 1. World Health Organization, Infertility is a global public health issue., (2017).

585 2. F. Tüttelmann, C. Ruckert, A. Röpke, Disorders of spermatogenesis: Perspectives for
586 novel genetic diagnostics after 20 years of unchanged routine, *medgen* **30**, 12–20 (2018).

587 3. A. M. Kasman, F. Del Giudice, M. L. Eisenberg, New insights to guide patient care: the
588 bidirectional relationship between male infertility and male health, *Fertility and Sterility* **113**,
589 469–477 (2020).

590 4. V. Nordhoff, R. K. Fricke, A. N. Schüring, M. Zitzmann, S. Kliesch, Treatment strategies for
591 severe oligoasthenoteratozoospermia (OAT) (<0.1 million/mL) patients, *Andrology* **3**, 856–
592 863 (2015).

593 5. S. Sharma, J. Wistuba, T. Pock, S. Schlatt, N. Neuhaus, Spermatogonial stem cells:
594 updates from specification to clinical relevance, *Human Reproduction Update* **25**, 275–297
595 (2019).

596 6. Y. Clermont, The cycle of the seminiferous epithelium in man, *Am. J. Anat.* **112**, 35–51
597 (1963).

598 7. Y. Clermont, Renewal of spermatogonia in man, *Am. J. Anat.* **118**, 509–524 (1966).

599 8. J. Ehmcke, J. Wistuba, S. Schlatt, Spermatogonial stem cells: questions, models and
600 perspectives, *Human Reproduction Update* **12**, 275–282 (2006).

601 9. J. Guo, E. J. Grow, H. Mlcochova, G. J. Maher, C. Lindskog, X. Nie, Y. Guo, Y. Takei, J.
602 Yun, L. Cai, R. Kim, D. T. Carrell, A. Goriely, J. M. Hotaling, B. R. Cairns, The adult human
603 testis transcriptional cell atlas, *Cell Res* **28**, 1141–1157 (2018).

604 10. B. P. Hermann, K. Cheng, A. Singh, L. Roa-De La Cruz, K. N. Mutoji, I.-C. Chen, H.
605 Gildersleeve, J. D. Lehle, M. Mayo, B. Westernströer, N. C. Law, M. J. Oatley, E. K. Velte, B.
606 A. Niedenberger, D. Fritze, S. Silber, C. B. Geyer, J. M. Oatley, J. R. McCarrey, The
607 Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to
608 Spermatids, *Cell Reports* **25**, 1650-1667.e8 (2018).

609 11. A. Sohni, K. Tan, H.-W. Song, D. Burow, D. G. de Rooij, L. Laurent, T.-C. Hsieh, R.
610 Rabah, S. S. Hammoud, E. Vicini, M. F. Wilkinson, The Neonatal and Adult Human Testis
611 Defined at the Single-Cell Level, *Cell Reports* **26**, 1501-1517.e4 (2019).

612 12. M. Wang, X. Liu, G. Chang, Y. Chen, G. An, L. Yan, S. Gao, Y. Xu, Y. Cui, J. Dong, Y.
613 Chen, X. Fan, Y. Hu, K. Song, X. Zhu, Y. Gao, Z. Yao, S. Bian, Y. Hou, J. Lu, R. Wang, Y.
614 Fan, Y. Lian, W. Tang, Y. Wang, J. Liu, L. Zhao, L. Wang, Z. Liu, R. Yuan, Y. Shi, B. Hu, X.
615 Ren, F. Tang, X.-Y. Zhao, J. Qiao, Single-Cell RNA Sequencing Analysis Reveals Sequential
616 Cell Fate Transition during Human Spermatogenesis, *Cell Stem Cell* **23**, 599-614.e4 (2018).

617 13. Nieschlag, Eberhard, Behre, Hermann M., Nieschlag, Susan, *Andrology*.

618 14. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M.
619 Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data, *Cell* **177**,
620 1888-1902.e21 (2019).

621   15. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, CellPhoneDB:
622   inferring cell–cell communication from combined expression of multi-subunit ligand–receptor
623   complexes, *Nat Protoc* **15**, 1484–1506 (2020).

624   16. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G.
625   Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J.
626   Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering, *Nat*
627   *Methods* **14**, 1083–1086 (2017).

628   17. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K.
629   Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van
630   Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I.
631   Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells, *Nature* **560**, 494–
632   498 (2018).

633   18. K. Van den Berge, H. Roux de Bézieux, K. Street, W. Saelens, R. Cannoodt, Y. Saeys,
634   S. Dudoit, L. Clement, Trajectory-based differential expression analysis for single-cell
635   sequencing data, *Nat Commun* **11**, 1201 (2020).

636   19. A. N. Shami, X. Zheng, S. K. Munyoki, Q. Ma, G. L. Manske, C. D. Green, M. Sukhwani,
637   K. E. Orwig, J. Z. Li, S. S. Hammoud, Single-Cell RNA Sequencing of Human, Macaque, and
638   Mouse Testes Uncovers Conserved and Divergent Features of Mammalian
639   Spermatogenesis, *Dev Cell* **54**, 529-547.e12 (2020).

640   20. K. von Kopylow, H. Staege, W. Schulze, H. Will, C. Kirchhoff, Fibroblast growth factor
641   receptor 3 is highly expressed in rarely dividing human type A spermatogonia, *Histochem*
642   *Cell Biol* **138**, 759–772 (2012).

643   21. S. Di Persio, R. Saracino, S. Fera, B. Muciaccia, V. Esposito, C. Boitani, B. P. Berloco, F.
644   Nudo, G. Spadetta, M. Stefanini, D. G. de Rooij, E. Vicini, Spermatogonial kinetics in
645   humans, *Development* **144**, 3430–3439 (2017).

646   22. C. A. Hogarth, D. Mitchell, C. Small, M. Griswold, EGR4 displays both a cell- and
647   intracellular-specific localization pattern in the developing murine testis, *Dev. Dyn.* **239**,
648   3106–3114 (2010).

649   23. S. R. Sung, S. H. Song, K. M. Kang, J. E. Park, Y. J. Nam, Y. Shin, D. H. Cha, J. T. Seo,
650   T. K. Yoon, S. H. Shim, Sequence variations of the EGR4 gene in Korean men with
651   spermatogenesis impairment, *BMC Med Genet* **18**, 47 (2017).

652   24. F. Hadziselimovic, N. O. Hadziselimovic, P. Demougin, G. Krey, B. Hoecht, E. J.
653   Oakeley, EGR4 Is a Master Gene Responsible for Fertility in Cryptorchidism, *Sex Dev* **3**,
654   253–263 (2009).

655   25. T. Matsuo, L. T. Dat, M. Komatsu, T. Yoshimaru, K. Daizumoto, S. Sone, Y. Nishioka, T.
656   Katagiri, Early growth response 4 is involved in cell proliferation of small cell lung cancer
657   through transcriptional activation of its downstream genes, *PLoS One* **9**, e113606 (2014).

658   26. K. Murata, U. Jadhav, S. Madha, J. van Es, J. Dean, A. Cavazza, K. Wucherpfennig, F.
659   Michor, H. Clevers, R. A. Shivdasani, Ascl2-Dependent Cell Dedifferentiation Drives
660   Regeneration of Ablated Intestinal Stem Cells, *Cell Stem Cell* **26**, 377-390.e6 (2020).

661   27. X. Wang, J.-H. Choi, J. Ding, L. Yang, L. C. Ngoka, E. J. Lee, Y. Zha, L. Mao, B. Jin, M.
662   Ren, J. Cowell, S. Huang, H. Shi, H. Cui, H.-F. Ding, HOXC9 directly regulates distinct sets
663   of genes to coordinate diverse cellular processes during neuronal differentiation, *BMC*
664   *Genomics* **14**, 830 (2013).

665   28. N. Holleville, S. Matéos, M. Bontoux, K. Bollerot, A. Monsoro-Burq, Dlx5 drives Runx2
666   expression and osteogenic differentiation in developing cranial suture mesenchyme,
667   *Developmental Biology* **304**, 860–874 (2007).

668   29. V. van den Boom, S. M. Kooistra, M. Boesjes, B. Geverts, A. B. Houtsmuller, K. Monzen,
669   I. Komuro, J. Essers, L. J. Drenth-Diephuis, B. J. L. Eggen, UTF1 is a chromatin-associated
670   protein involved in ES cell differentiation, *Journal of Cell Biology* **178**, 913–924 (2007).

671   30. S. M. Kooistra, R. P. Thummer, B. J. L. Eggen, Characterization of human UTF1, a
672   chromatin-associated protein with repressor activity expressed in pluripotent cells, *Stem Cell*
673   *Research* **2**, 211–218 (2009).

674   31. S. M. Kooistra, V. van den Boom, R. P. Thummer, F. Johannes, R. Wardenaar, B. M.
675   Tesson, L. M. Veenhoff, F. Fusetti, L. P. O'Neill, B. M. Turner, G. de Haan, B. J. L. Eggen,
676   Undifferentiated Embryonic Cell Transcription Factor 1 Regulates ESC Chromatin
677   Organization and Gene Expression, *STEM CELLS* **28**, 1703–1714 (2010).

678   32. M. M. van Alphen, H. J. van de Kant, D. G. de Rooij, Depletion of the spermatogonia
679   from the seminiferous epithelium of the rhesus monkey after X irradiation, *Radiat Res* **113**,
680   473–486 (1988).

681   33. S. Z. Jan, T. L. Vormer, A. Jongejan, M. D. Röling, S. J. Silber, D. G. de Rooij, G. Hamer,
682   S. Repping, A. M. M. van Pelt, Unraveling transcriptome dynamics in human
683   spermatogenesis, *Development* **144**, 3659–3673 (2017).

684   34. T. DeFalco, S. J. Potter, A. V. Williams, B. Waller, M. J. Kan, B. Capel, Macrophages
685   Contribute to the Spermatogonial Niche in the Adult Testis, *Cell Reports* **12**, 1107–1119
686   (2015).

687   35. S. Yoshida, M. Sukeno, Y. -i. Nabeshima, A Vasculature-Associated Niche for
688   Undifferentiated Spermatogonia in the Mouse Testis, *Science* **317**, 1722–1726 (2007).

689   36. K. Tan, H.-W. Song, M. Thompson, S. Munyoki, M. Sukhwani, T.-C. Hsieh, K. E. Orwig,
690   M. F. Wilkinson, Transcriptome profiling reveals signaling conditions dictating human
691   spermatogonia fate in vitro, *Proc Natl Acad Sci U S A* **117**, 17832–17841 (2020).

692   37. A. Goriely, A. O. M. Wilkie, Paternal Age Effect Mutations and Selfish Spermatogonial
693   Selection: Causes and Consequences for Human Disease, *The American Journal of Human*
694   *Genetics* **90**, 175–200 (2012).

695   38. E. Leitão, S. Di Persio, S. Laurentino, M. Wöste, M. Dugas, S. Kliesch, N. Neuhaus, B.
696   Horsthemke, The sperm epigenome does not display recurrent epimutations in patients with
697   severely impaired spermatogenesis, *Clin Epigenet* **12**, 61 (2020).

698   39. World Health Organization, *WHO laboratory manual for the Examination and processing*
699   *of human semen* (Geneva, 5th edn., 2010; https://www.who.int/docs/default-source/srhr-
700   documents/infertility/examination-and-processing-of-human-semen-5ed-
701   eng.pdf?sfvrsn=5227886e_2).

702   40. M. S. Oud, L. Volozonoka, R. M. Smits, L. E. L. M. Vissers, L. Ramos, J. A. Veltman, A
703   systematic review and standardized clinical validity assessment of male infertility genes,
704   *Human Reproduction* **34**, 932–941 (2019).

705   41. N. Kossack, N. Terwort, J. Wistuba, J. Ehmcke, S. Schlatt, H. Schöler, S. Kliesch, J.
706   Gromoll, A combined approach facilitates the reliable detection of human spermatogonia in
707   vitro, *Human Reproduction* **28**, 3012–3025 (2013).

708 42. S. Laurentino, L. Heckmann, S. Di Persio, X. Li, G. Meyer Zu Hörste, J. Wistuba, J.-F.
709 Cremers, J. Gromoll, S. Kliesch, S. Schlatt, N. Neuhaus, High-resolution analysis of germ
710 cells from men with sex chromosomal aneuploidies reveals normal transcriptome but
711 impaired imprinting, *Clin Epigenetics* **11**, 127 (2019).

712 43. S. Albert, J. Wistuba, K. Eildermann, J. Ehmcke, S. Schlatt, J. Gromoll, N. Kossack,
713 Comparative Marker Analysis after Isolation and Culture of Testicular Cells from the
714 Immature Marmoset, *Cells Tissues Organs* **196**, 543–554 (2012).

715 44. Bergmann and Kliesch, in *Nieschlag E., Behre H.M., Nieschlag S. (eds) Andrology*,
716 (Springer, Berlin, Heidelberg, 2010), pp. 155–167.

717 45. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S.
718 Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K.
719 Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image
720 analysis, *Nat Methods* **9**, 676–682 (2012).

721 46. J. A. Vizcaíno, E. W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J. A. Dianes,
722 Z. Sun, T. Farrah, N. Bandeira, P.-A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A.
723 Campos, R. J. Chalkley, H.-J. Kraus, J. P. Albar, S. Martinez-Bartolomé, R. Apweiler, G. S.
724 Omenn, L. Martens, A. R. Jones, H. Hermjakob, ProteomeXchange provides globally
725 coordinated proteomics data submission and dissemination, *Nat Biotechnol* **32**, 223–226
726 (2014).

727 47. A. Srivastava, L. Malik, T. Smith, I. Sudbery, R. Patro, Alevin efficiently estimates
728 accurate gene abundances from dscRNA-seq data, *Genome Biol* **20**, 65 (2019).

729 48. A. Frankish, M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M.
730 Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J.
731 Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M.
732 Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin, L.
733 Martínez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B. M.
734 Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A. Yates, D.
735 Zerbino, Y. Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M.
736 Kellis, B. Paten, A. Reymond, M. L. Tress, P. Flicek, GENCODE reference annotation for the
737 human and mouse genomes, *Nucleic Acids Research* **47**, D766–D773 (2019).

738 49. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell
739 transcriptomic data across different conditions, technologies, and species, *Nat Biotechnol* **36**,
740 411–420 (2018).

741 50. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq
742 data using regularized negative binomial regression, *Genome Biol* **20**, 296 (2019).

743 51. L. Waltman, N. J. van Eck, A smart local moving algorithm for large-scale modularity-
744 based community detection, *Eur. Phys. J. B* **86**, 471 (2013).

745 52. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection
746 for Dimension Reduction, *arXiv:1802.03426 [cs, stat]* (2018) (available at
747 http://arxiv.org/abs/1802.03426).

748 53. G. Finak, A. McDavid, M. Yajima, J. Deng, V. Gersuk, A. K. Shalek, C. K. Slichter, H. W.
749 Miller, M. J. McElrath, M. Prlic, P. S. Linsley, R. Gottardo, MAST: a flexible statistical
750 framework for assessing transcriptional changes and characterizing heterogeneity in single-
751 cell RNA sequencing data, *Genome Biol* **16**, 278 (2015).

752 54. F. A. Wolf, F. K. Hamey, M. Plass, J. Solana, J. S. Dahlin, B. Göttgens, N. Rajewsky, L.
753 Simon, F. J. Theis, PAGA: graph abstraction reconciles clustering with trajectory inference
754 through a topology preserving map of single cells, *Genome Biol* **20**, 59 (2019).

755 55. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression
756 data analysis, *Genome Biol* **19**, 15 (2018).

757 56. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to
758 transient cell states through dynamical modeling, *Nat Biotechnol* **38**, 1408–1414 (2020).

759 57. T. Moerman, S. Aibar Santos, C. Bravo González-Blas, J. Simm, Y. Moreau, J. Aerts, S.
760 Aerts, J. Kelso, Ed. GRNBoost2 and Arboreto: efficient and scalable inference of gene
761 regulatory networks, *Bioinformatics* **35**, 2159–2161 (2019).

762 58. S. Suo, Q. Zhu, A. Saadatpour, L. Fei, G. Guo, G.-C. Yuan, Revealing the Critical
763 Regulators of Cell Identity in the Mouse Cell Atlas, *Cell Reports* **25**, 1436-1445.e3 (2018).

764 59. K. Van den Berge, C. Soneson, M. D. Robinson, L. Clement, stageR: a general stage-
765 wise method for controlling the gene-level false discovery rate in differential expression and
766 differential transcript usage, *Genome Biol* **18**, 151 (2017).

767 60. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized
768 p.p.b.-range mass accuracies and proteome-wide protein quantification, *Nat Biotechnol* **26**,
769 1367–1372 (2008).

770 61. V. G. Tusher, R. Tibshirani, G. Chu, Significance analysis of microarrays applied to the
771 ionizing radiation response, *Proceedings of the National Academy of Sciences* **98**, 5116–
772 5121 (2001).

773 **Acknowledgements**

779 **Funding**

783 **Author Contributions**

784   S.D.P., T.T., S.L. and N.N. designed the study, performed data analysis, data interpretation and

785   wrote the manuscript; H.C.A.D. performed low-input proteome analyses; M.W., F.T.

786   performed whole exome sequencing and data analyses; X.L., G.M.Z.H. provided the

787   infrastructure and expertise for single cell RNA sequencing and prepared sequencing libraries;

788   T.T., M.D. performed bioinformatic analyses of single cell RNA sequencing datasets,

789   L.M.S.K., S.D.P. performed experiments; J.F.C., S.K. supervised clinical patient care and

790   documentation in the database, performed surgeries and were responsible for clinical and

791   laboratory testing; J.W., S.S. provided expertise for histomorphometrical analyses and data

792   interpretation; N.N. coordinated the project. All authors approved the final version of the

793   manuscript.

794   **Competing interests**

795   The authors declare no competing interests.

796   **Data and materials availability**

797   All mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium

798   (http://proteomecentral.proteomexchange.org)*(46)* via the PRIDE partner repository. The data

799   will be provided upon request.

800   The sequencing data generated during this study as well as Alevin counts, integrated counts and

801   a cell metadata table are available at the NIH Gene Expression Omnibus (GEO) under accession

802   number GSE153947 (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153947). The data

803   will be provided upon request.

804

805   **Figures**

806

807

808
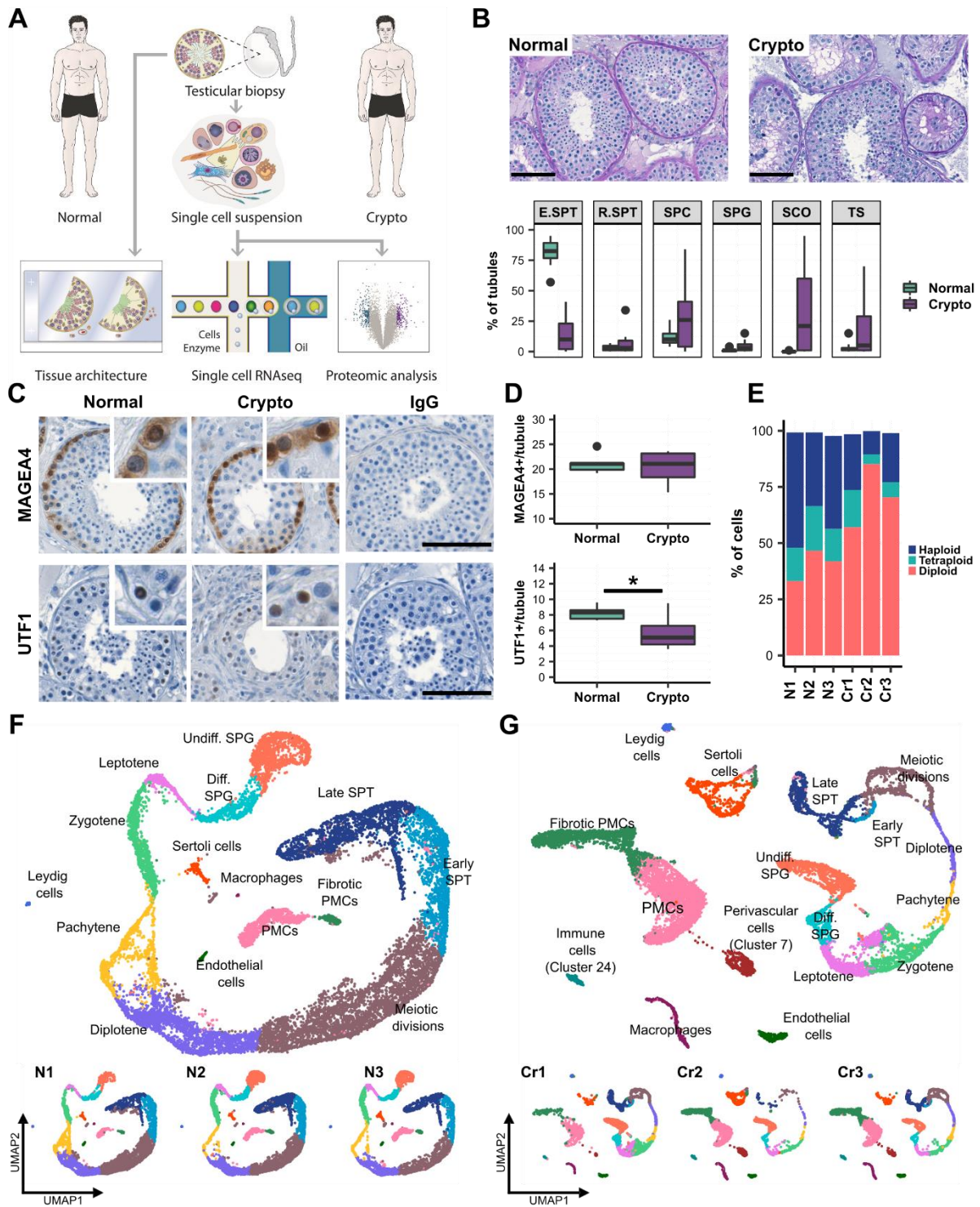
809 **Fig. 1. Crypto tissues show stable numbers of MAGEA4$^+$ spermatogonia but reduced**

810 **numbers of UTF1$^+$ spermatogonia.**

811 **(A)** Experimental outline. **(B)** Upper panel: Representative Periodic acid-Schiff stained

812 micrographs of the testicular tissue of one normal (left) and one crypto (right) sample. Scale

813 bars = 100 µm. Lower panel: Box plots showing the percentages of tubules containing germ

814 cells (most advanced germ cell type), only Sertoli cells, or tubular shadows in the patient cohort.

815 Values per patient can be found in table S1. * p<0.05, *** p<0.001. **(C)** Micrographs showing

816 the MAGEA4 and UTF1 staining in the normal and crypto testicular tissue. Inlays show

817 examples of positive and negative cells for each staining. IgG negative controls show no

818 staining. Scale bar = 100 µm. **(D)** Quantification of MAGEA4$^+$ (upper) and UTF1$^+$ (lower)

819 spermatogonia per tubule in normal (n=6) and crypto samples (n=6). No statistical difference

820 was found in the number of MAGEA4$^+$ spermatogonia per tubule between normal and crypto

821 samples. A significant reduction was found in the number of UTF1$^+$ spermatogonia per tubule

822 in the crypto samples (* p<0.05; Statistical details are available in table S2). **(E)** Stacked bar

823 plots showing the ploidy of the single cell suspensions used for scRNA-seq analysis of each

824 patient sample. **(F)** Upper panel: Uniform manifold approximation and projection (UMAP) plot

825 of the integrated normal dataset. Clusters were assigned based on the expression of 55 markers

826 genes (fig. S1B). Lower panel: scRNA-seq data from the 3 normal samples. Each color

827 represents a different cell type. **(G)** Upper panel: UMAP plot of the integrated crypto dataset.

828 The clusters were assigned by anchoring integration using the normal dataset as reference.

829 Lower panel: scRNA-seq data from the 3 crypto samples.

830

831

832

32

833
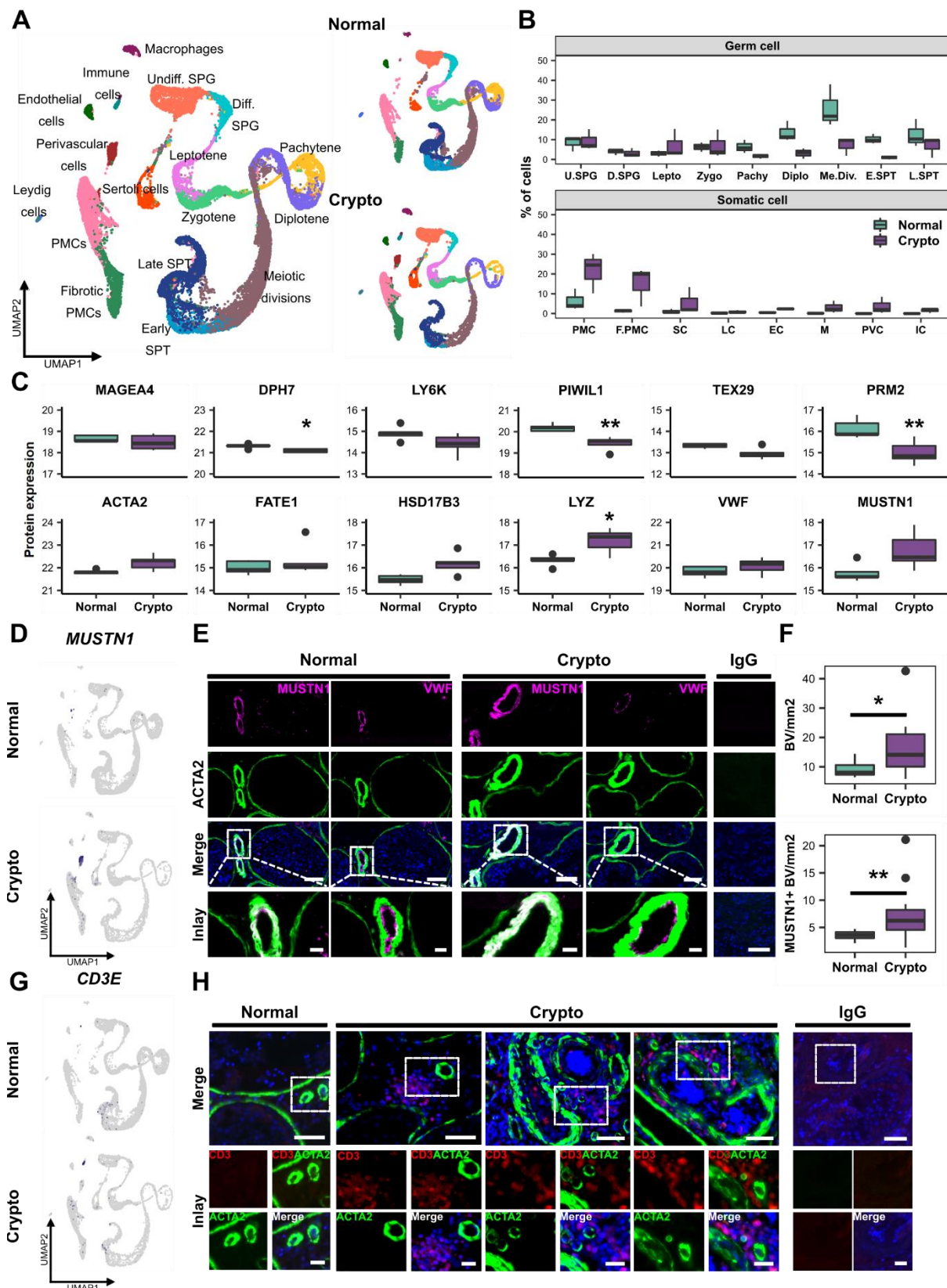
Fig. 2. Exploration of cellular and transcriptional changes in crypto testicular tissues.

(A) Left panel: UMAP plot of the all integrated dataset. Right panel: Contribution of the normal

(15 546 cells) and crypto (13 144 cells) datasets to the all integrated dataset. The cells are color

837   coded according to the respective cell types. **(B)** Box plot comparing the germ and somatic cell

838   proportions in normal and crypto scRNA-seq datasets. **(C)** Panel of box plots representing the

839   protein expression of 12 marker genes used in the cluster assignment (MAGEA4:

840   Spermatogonia; DPH7: Leptotene spermatocytes; LY6K: Zygotene spermatocytes; PIWIL1:

841   Pachytene spermatocytes; TEX29: Early spermatids; PRM2: Elongated spermatids; ACTA2:

842   Peritubular myoid cells (PMCs); FATE1: Sertoli cells; HSD17B3: Leydig cells; VWF:

843   Endothelial cells; LYZ: Macrophages; MUSTN1: Perivascular cells). Protein expression was

844   measured performing bottom-up mass spectrometry analysis on single cell suspensions (normal

845   n=5, crypto n=4 patients) and shows a reduction of germ cells, starting from the leptotene

846   spermatocytes, and an increase of macrophages in crypto patients. * $p<0.05$, ** $p<0.01$. **(D)**

847   Feature plots highlighting the expression of *MUSTN1* (Perivascular cell marker gene) in normal

848   and crypto datasets. **(E)** Representative micrographs showing blood vessels in consecutive

849   sections of normal and crypto testicular tissues stained for MUSTN1 (magenta)/ACTA2 (green)

850   and for VWF (magenta)/ACTA2 (green). The tissue sections were counterstained with DAPI

851   (blue). Double positive MUSTN1/ACTA2 cells were found surrounding the endothelial layer

852   of the blood vessels. The IgG control showed no immunological staining. Scale bar = 100 μm

853   (main) and 20 μm (inlays). **(F)** Box plots representing the number of ACTA2$^+$ blood vessels

854   (Upper) and MUSTN1$^+$ blood vessels (lower) per mm^2 of tissue in the normal (n=12) and crypto

855   samples (n=13). A significant increase of blood vessels and MUSTN1$^+$ blood vessels was found

856   between the two cohorts (Statistical details are available in table S2). **(G)** Feature plots

857   highlighting the expression of *CD3E* (Immune cell marker gene) in normal and crypto datasets.

858   **(H)** Representative micrographs showing CD3$^+$ immune cells (Red) and ACTA2$^+$ blood vessels

859   (Green) in normal and crypto testicular tissues. The tissue sections were counterstained with

860   DAPI (Blue). CD3$^+$ immune cells were found solely in the crypto group in close proximity to

861   blood vessels. The IgG control showed no immunological staining. Scale bar = 50 μm (main)

862   and 20 μm (inlays).

34
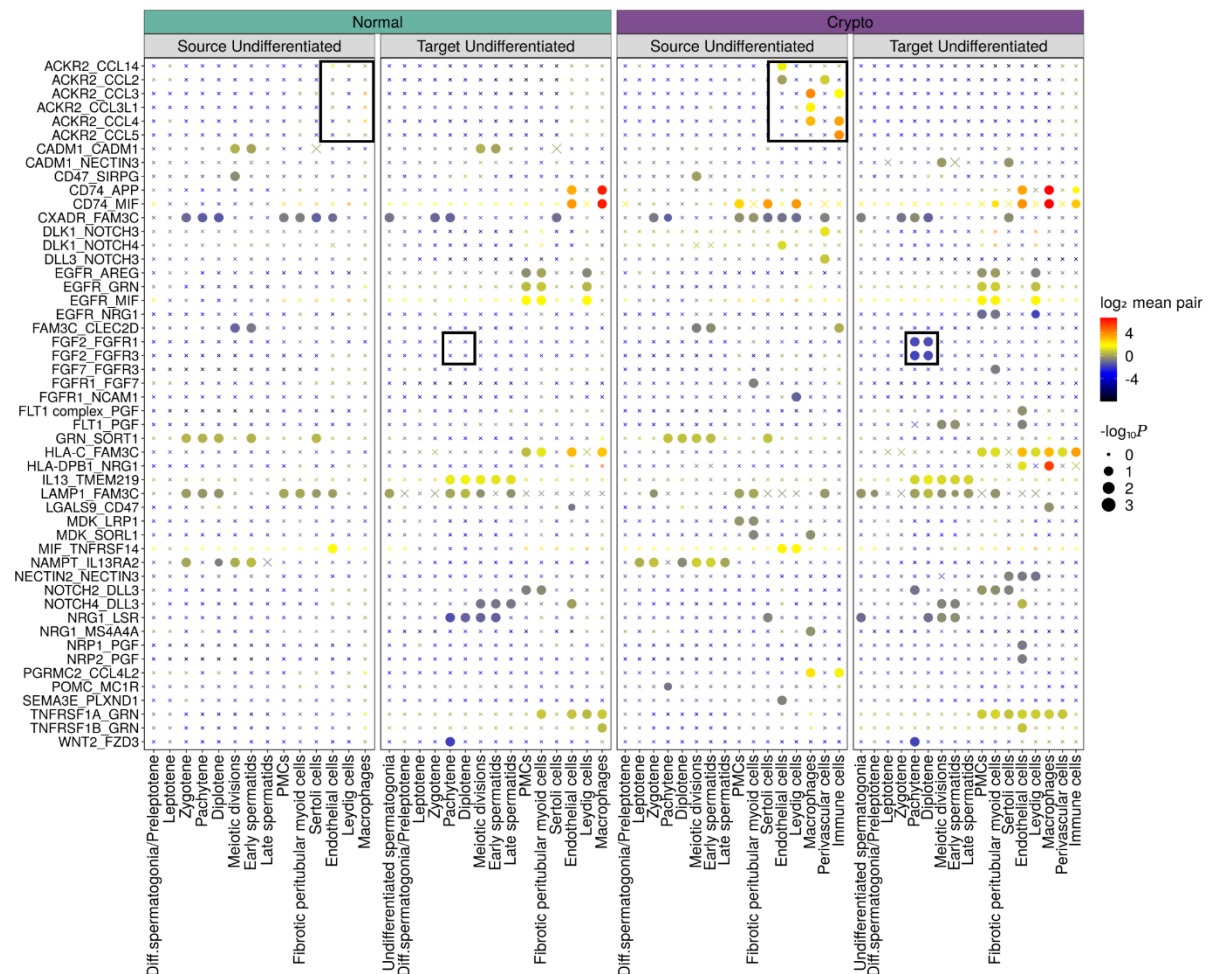
Fig. 3. Ligand-receptor interaction analysis.

Summary of ligand–receptor interactions between the normal and crypto undifferentiated spermatogonia and the other cell types in the tissues. P–values are represented by the size of each circle. The color gradient indicates the level of interaction. Black squares indicate selected interactions that show a different significance in the normal and crypto datasets.
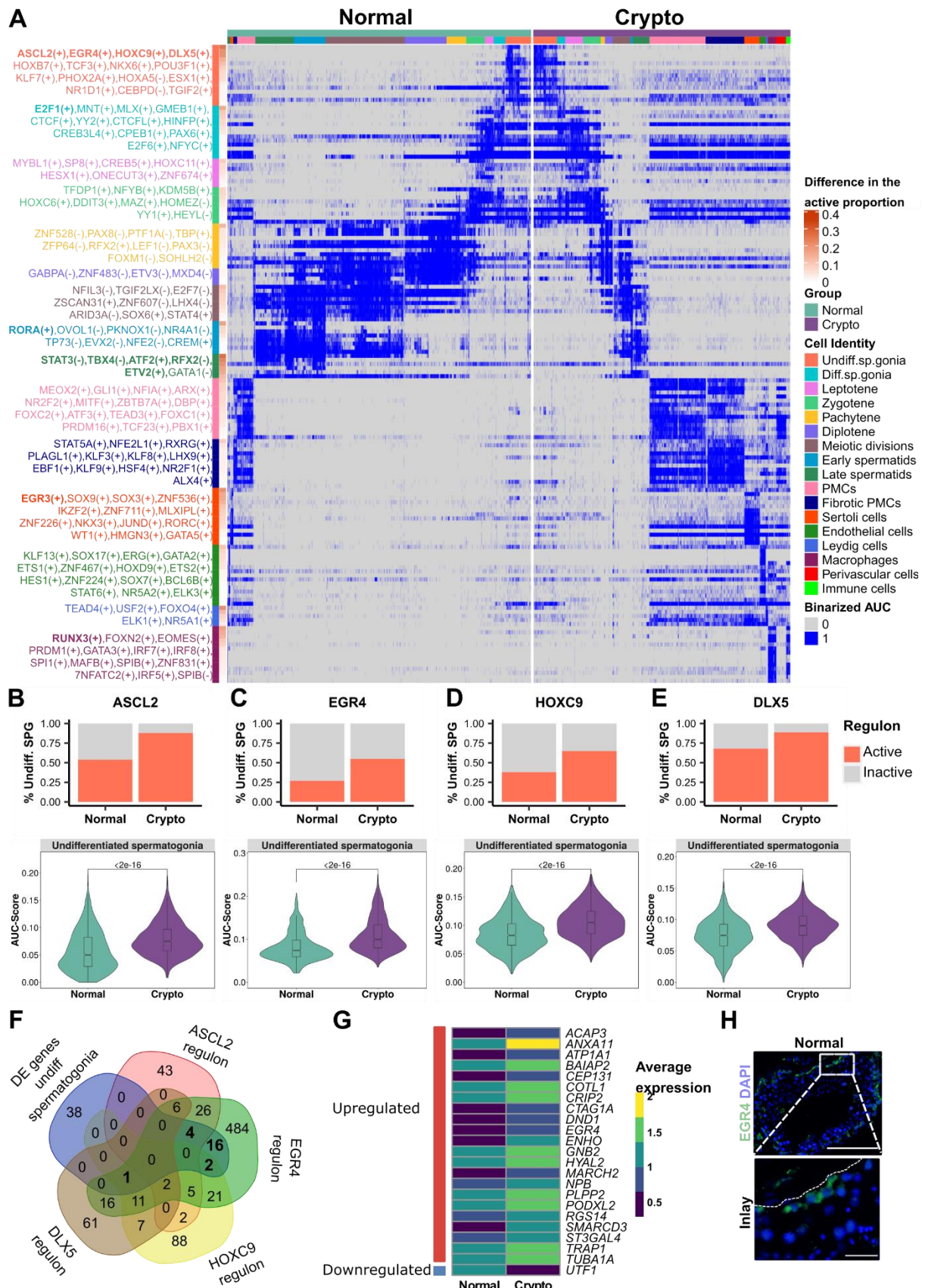
877 **Fig. 4. Undifferentiated spermatogonia activate EGR4 regulatory network in the crypto**

878 **group.**

879 **(A)** Binarized double heatmap showing the AUC score (area under the recovery curve; it scores

880 the activity of regulons) of the identified regulons plotted separately for the normal (teal, left)

881 and crypto (purple, right) datasets. For each cellular cluster, the first 15 regulons showing the

882 highest regulon specificity score (RSS) in the normal dataset were plotted. The regulons written

883 in bold are those showing a difference of at least 20% in the proportion of cells with the active

884 regulon in the crypto dataset and a significant difference in the regulon specific AUC score

885 between normal and crypto datasets. **(B-E)** Upper panel: Stacked bar plots comparing the

886 proportion of undifferentiated spermatogonia with active **(B)** ASCL2, **(C)** EGR4, **(D)** HOXC9

887 and **(E)** DLX5 regulons. Lower panel: Violin plots comparing the AUC score of the **(B)**

888 ASCL2, **(C)** EGR4, **(D)** HOXC9 and **(E)** DLX5 regulons in the normal and crypto

889 undifferentiated spermatogonia. A significant increase was found for all four regulons in the

890 crypto dataset (BH corrected p-value of Mann-Whitney U test). A complete list of the regulons

891 and their RSS and AUC scores is available in table S8. **(F)** Overlap among the 61 genes

892 uniquely differentially expressed in undifferentiated spermatogonia and the genes putatively

893 regulated by ASCL2, EGR4, HOXC9 and DLX5. **(G)** Heatmap showing the expression in the

894 normal and crypto undifferentiated spermatogonia of the 23 differentially expressed genes

895 putatively regulated by EGR4. **(H)** Micrographs showing the expression of EGR4 in

896 spermatogonia in normal testicular tissue. Scale bars: main=100 µm; inlay=20 µm.

897

898

899

900

901

37

902

903
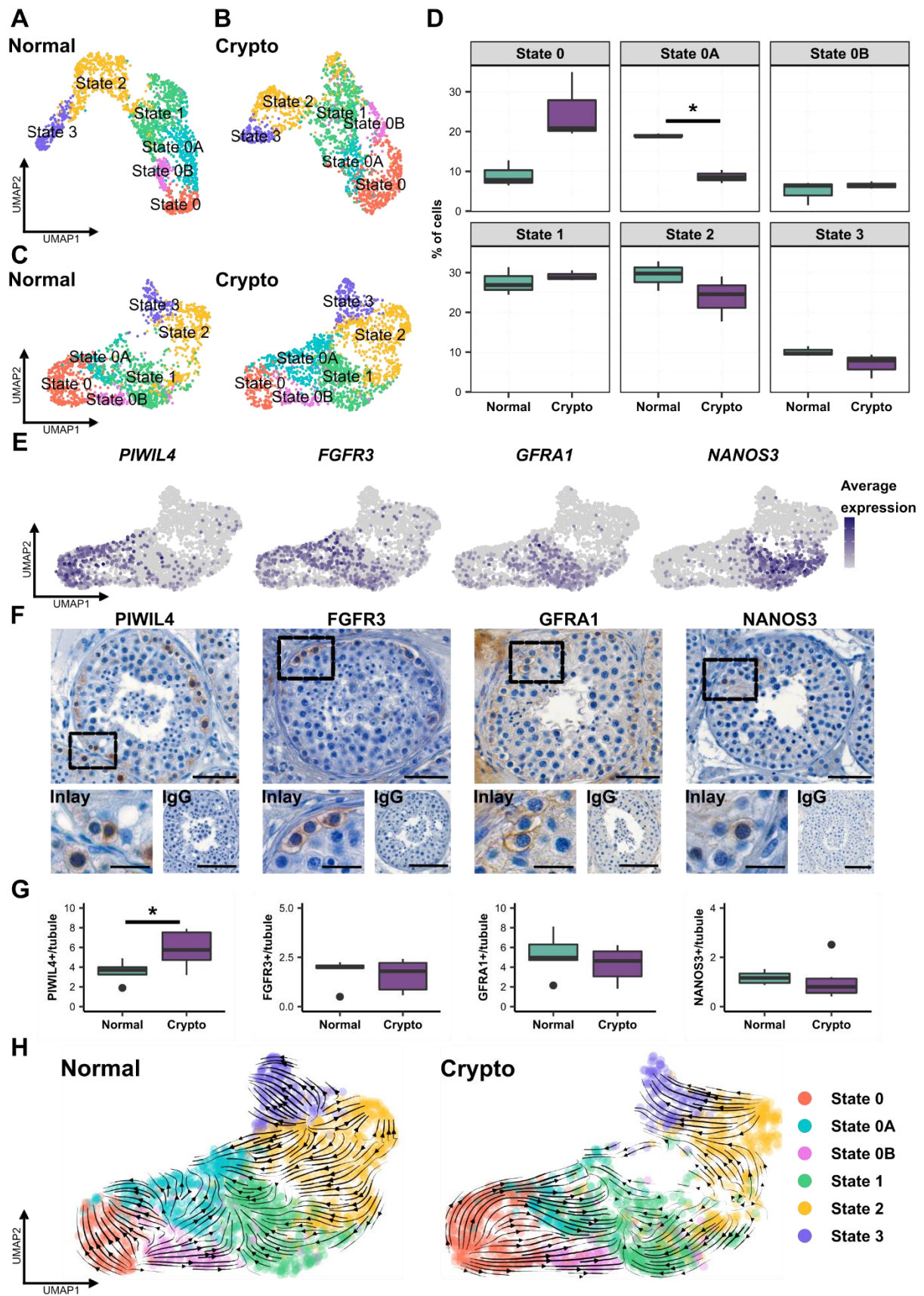
904 **Fig. 5. The crypto SSC compartment shows altered composition with an increased**

905 **number of PIWIL4$^+$ spermatogonia.**

906 **(A)** UMAP plot showing the cluster assignment of the normal dataset. Six clusters were

907 assigned based on the expression of markers shown in fig. S5B: State 0, 0A, 0B, 1, 2, and 3.

908 Each color represents a different cluster. **(B)** UMAP plot showing the cluster assignment of the

909 crypto dataset. The clusters were assigned by the anchoring integration method using the normal

910 dataset as reference. Each color represents a different cluster, as in the normal dataset. **(C)**

911 UMAP plot showing the integrated normal and crypto spermatogonial datasets. Each color

912 represents a different cluster, as in the normal dataset. **(D)** Box plots representing the different

913 cell proportions in the normal and crypto spermatogonial datasets. The proportion of State 0A

914 spermatogonia was significantly reduced in the crypto dataset (* $p<0.05$). **(E)** Feature plots

915 showing the expression level of *PIWIL4, FGFR3, GFRA1,* and *NANOS3* in the all integrated

916 datasets. **(F)** Micrographs showing the PIWIL4, FGFR3, GFRA1 and NANOS3 staining in the

917 testicular tissue. Inlays show examples of positive and negative cells for each staining. IgG

918 negative controls show no staining. Scale bars: main $= 50\,\mu m$, Inlay$=20\,\mu m$; IgG$=100\,\mu m$.

919 **(G)** Quantification of PIWIL4$^+$, FGFR3$^+$, GFRA1$^+$, and NANOS3$^+$ spermatogonia per tubule

920 in normal (n=6) and crypto samples (n=6). A significant increase was found in the number of

921 PIWIL4$^+$ spermatogonia per tubule in the crypto samples (* $p<0.05$). No statistical difference

922 was found in the total number of GFRA1$^+$, FGFR3$^+$ and NANOS3$^+$ spermatogonia between

923 normal and crypto samples. Statistical details are available in table S2. **(H)** RNA velocities

924 derived from the scVelo dynamical model for the normal and crypto spermatogonial dataset are

925 visualized as streamlines in UMAP plots. A change in the arrow directionality can be observed

926 in the crypto dataset.

927

928

929
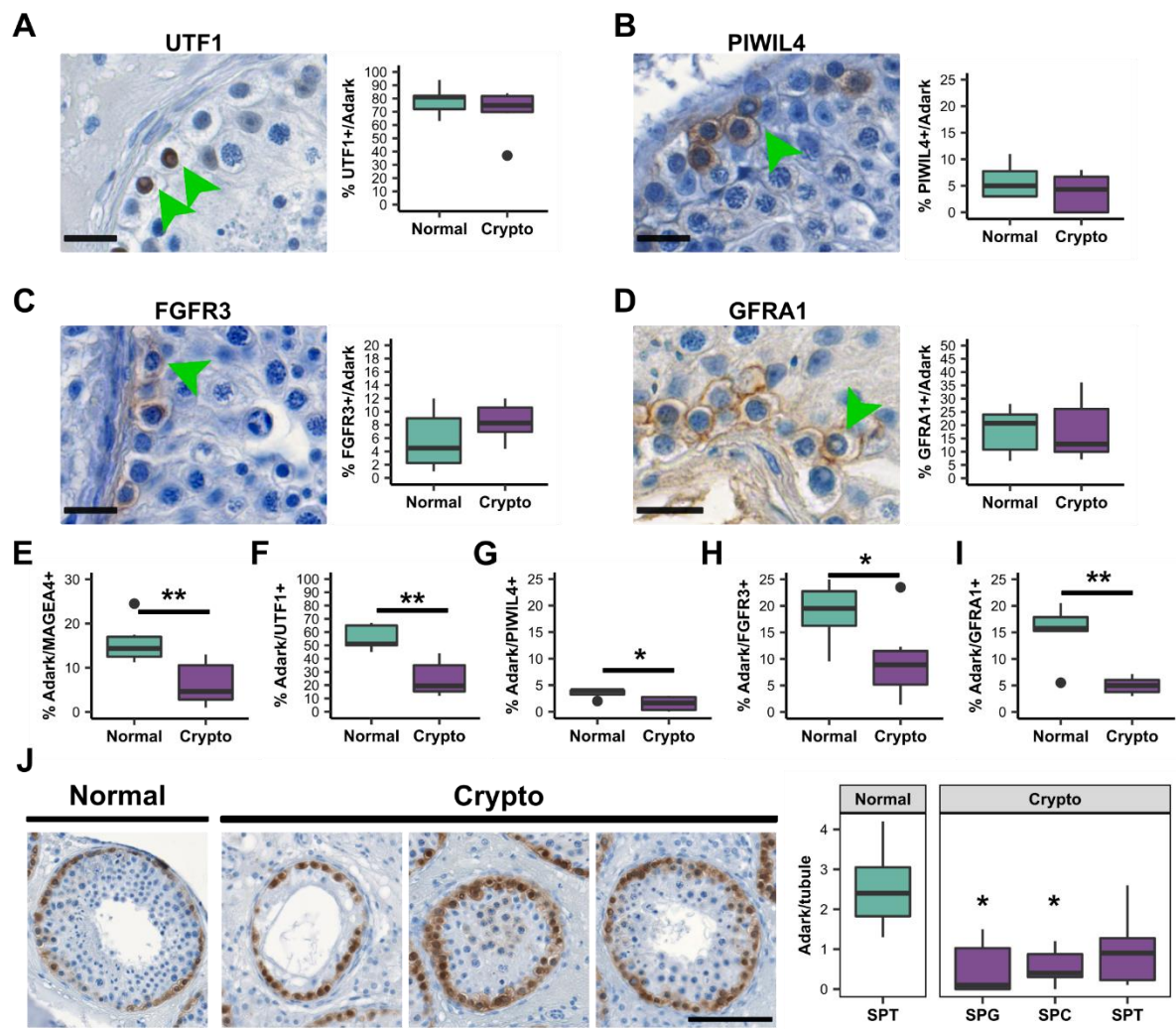
Fig. 6. The crypto group shows altered numbers of A_{dark} spermatogonia.

(A-D) Right panel: Micrographs showing **(A)** UTF1$^+$, **(B)** PIWIL4$^+$, **(C)** FGFR3$^+$, **(D)** GFRA1$^+$ A_{dark} spermatogonia. Green arrow heads indicate A_{dark} spermatogonia. Scale bars=20 µm. Left panel: Boxplots showing the percentage of UTF1$^+$ **(A)**, PIWIL4$^+$ **(B)**, FGFR3$^+$ **(C)** and GFRA1$^+$ **(D)** spermatogonia among the A_{dark} population in normal (n=6) and crypto samples (n=6). A mean of 78.4% of the A_{dark} spermatogonia were UTF1$^+$, 5.8% were PIWIL4$^+$, 5.7% were FGFR3$^+$ and 18.1% were GFRA1$^+$ in the normal dataset. No significant difference was found in the crypto samples. **(E-I)** Quantification of the percentage of A_{dark} spermatogonia among MAGEA4$^+$ **(E)**, UTF1$^+$ **(F)**, PIWIL4$^+$ **(G)**, FGFR3$^+$ **(H)** and GFRA1$^+$ **(I)** spermatogonial populations in normal (n=6) and crypto samples (n=6). A significant reduction of A_{dark} spermatogonia was found among all the cell types (**p <0.01, *p <0.05). **(J)** Quantification of

40

942   the proportion of A_{dark} spermatogonia per tubule according to the most advanced germ cell type

943   present. Left: Representative micrographs showing in the order one tubule from a normal

944   sample containing spermatids as the most differentiated germ cell and three tubules from a

945   crypto sample containing either spermatogonia (SPG), spermatocytes (SPC), or spermatids

946   (SPT) as the most differentiated germ cell. Scale bar = 100µm. Right panel: Boxplot showing

947   the quantification of A_{dark} spermatogonia per tubule in normal (n=6) and crypto samples (n=6).

948   *p <0.05. The asterisks refer to the comparison with the normal samples. Statistical details are

949   available in table S2.

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

41
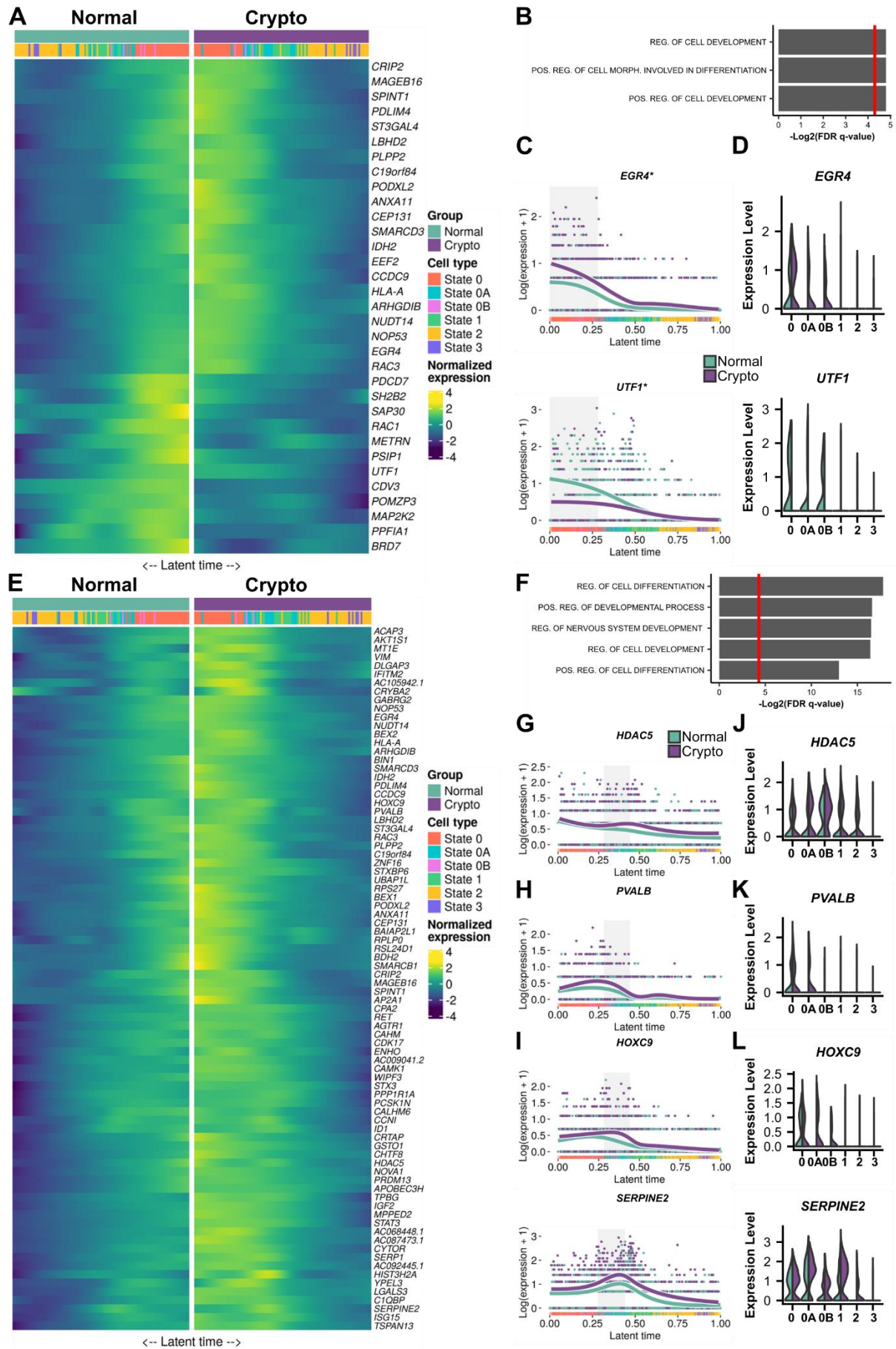
Fig. 7. Crypto State0 and 0A spermatogonia show enhanced expression of EGR4 and HOXC9.

(A) Double heatmap showing the normalized expression of the 33 genes identified by tradeSeq analysis as differentially expressed in the knotgroup 1 of the crypto dataset. The cells are plotted along the latent time with the normal cells on the left side (teal) and crypto cells on the right side (purple). **(B)** The bar plot shows GO analysis of the 33 differentially expressed genes in the knotgroup 1. The results showed an enrichment in GO terms related to cellular differentiation and development. The red line on the bar plot represents p=0.05. **(C)** Line plots showing the expression of *EGR4* and *UTF1* in the normal (teal) and crypto (purple) cells along the latent time. The gray area highlights the cells belonging to the knotgroup 1. The plots show enhanced expression of *EGR4* and reduced expression of *UTF1* in the crypto knotgroup 1. **(D)** Double violin plots comparing the expression level of *EGR4* and *UTF1* in the spermatogonial states in the normal (teal, left) and crypto (purple, right) datasets. **(E)** Double heatmap showing the normalized expression of the 82 genes identified by the tradeSeq analysis as differentially expressed in the knotgroup 2 of the crypto dataset. The cells are plotted along the latent time with the normal cells on the left side (teal) and crypto cells on the right side (purple). **(F)** The bar plot shows the GO analysis of the 82 differentially expressed genes in the knotgroup 2. The results showed an enrichment in GO terms related to cellular differentiation and development. The red line on the bar plot represents the p=0.05. **(G-I)** Line plots showing the expression along the latent time of *HDAC5* (**G**; regulated by EGR4 only), *PVALB* (**H**; regulated by HOXC9 only) and *SERPINE2* and *HOXC9* (**I**; regulated by EGR4 and HOXC9) in the normal (teal) and crypto (purple) cells. The gray area highlights the cells belonging to the knotgroup 2. The plots show enhanced expression of all the genes in the crypto knotgroup 2. **(J-L)** Double violin plots comparing the expression level of *HDAC5* (**J**; present in the EGR4 regulon only), *PVALB* (**K**; present in the HOXC9 regulon only), *SERPINE2* and *HOXC9* (**L**; present in both the EGR4 and

43

993   HOXC9 regulons) in the normal (teal) and crypto (purple) cells along the latent time. The plots

994   show enhanced expression of the all the genes in the crypto State 0, 0A, 0B spermatogonia.

995
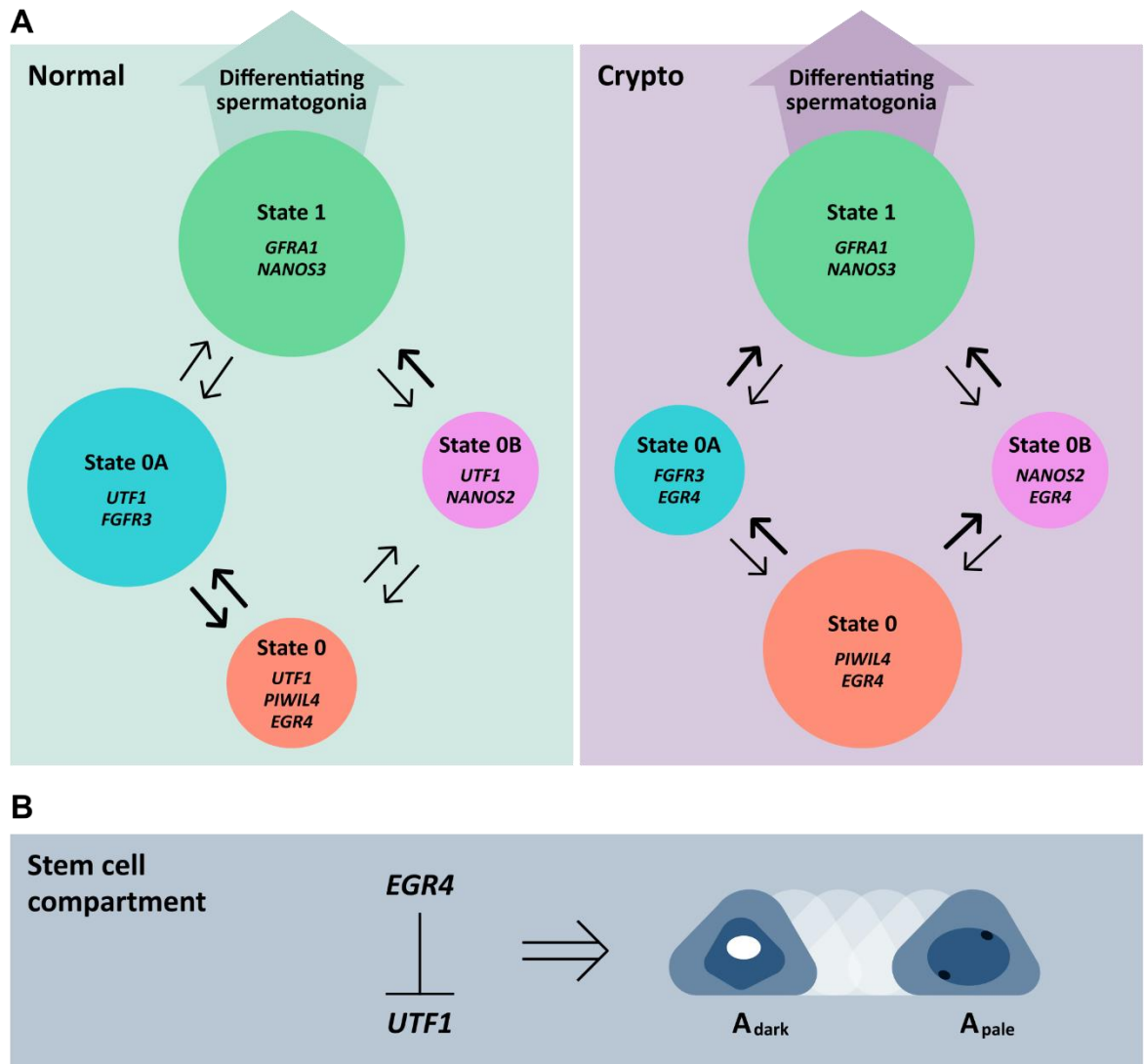
996   **Fig. 8. Model of the human spermatogonial compartment in normal and crypto samples.**

997   **(A)** The circles represent the State 0 (Red), State 0A (Blue), State 0B (Magenta) and State 1

998   (Green) spermatogonia. The thickness of the black arrows indicates the proportion of cells

999   changing their expression profile. (**B**) Suggested mechanism: the EGR4 mediated UTF1

1000   downregulation induces the chromatin remodeling resulting in the transition from A_{dark} to A_{pale}

1001   morphology.

1002

1003

Supplementary Material

Material and Methods

ScRNA quantification and UMI/CB correction

Alevin *(47)* was used to deduplicate and quantify the single-cell sequencing data and to discern the valid cellular barcodes (CB) from the background noise. The Gencode *(48)* human reference v30 was used as transcriptomic reference, together with the flags --useCorrelation and --chromium in the quantification. Additionally, the --expectCells parameter was set to 6 000 (50% of the initial loading of 12 000 cells) to help the CB knee finding method. This approach resulted in a mean of 4 783 cells (SD: 587) per sample and a mean/median number of 2 460 (SD: 799) expressed genes per cell. All software is listed in table S10.

Normalization, sample integration, dimensional reduction and label transfer

Alevin counts were processed with Seurat *(14, 49)*. To reduce initial noise, genes expressed in less than 0.1% of the cells were excluded, together with cells containing fewer than 200 counts. In total, only 9 cells were removed by this filter. The analysis of mitochondrial content in regard to the total measured expression revealed 471 cells (1.6%) with a strong mitochondrial effect (>25% of mitochondrial counts), 306 of them (65%) later identified as Sertoli cells. It is unknown at the moment if this high mitochondrial content is a sign of cell death or a normal feature of Sertoli cells, therefore these cells were not excluded from the analysis. Mitochondrial content and total RNA counts were used as a regression variable in the normalization procedure sctransform *(50)*, with the number of variable features set to 5 000. Anchor-based sample integration was performed on the normalized counts, setting the number of features in the anchor finding process to 5 000. We integrated normal and crypto samples (n=3, each) in three ways, resulting in an integrated dataset of all normal samples, an integrated dataset of all crypto samples and an integrated dataset consisting of all six samples. As proposed by the Seurat

45

1028   developers, scaling was disabled in the FindIntegrationAnchors() method, due to the prior

1029   normalization by sctransform. All available genes were kept in the integrated dataset,

1030   irrespective of their usage as anchors or not. Integrated counts were scaled and centered for

1031   principal component (PC) analysis and dimensional reduction. For each scaled dataset, 100 PCs

1032   were calculated using the most variable genes. Elbow-Plots of the PCs were generated to

1033   determine the inflection point, confirming 25 PCs for further analysis. A shared nearest

1034   neighbor graph was constructed based on the PCs. Cell clusters were identified via the smart

1035   local moving algorithm *(51)* with a resolution parameter of 1 100 random starts and a maximum

1036   of 100 iterations. For data visualization, the non-linear dimensional reduction method UMAP

1037   *(52)* was used based on the specified 25 PCs.

1038   To enable an unbiased verification of the cluster identities, the top marker genes per cluster

1039   were computationally determined with the FindAllMarkers() function, testing with MAST *(53)*

1040   only for positive marker genes with a minimal expressed-in fraction of 30% ($p < 0.001$). These

1041   computational markers were used with known cell type markers from the literature *(9–12)* to

1042   carefully assign cluster identities. Following the marker-based cluster identity assignment in

1043   the normal dataset, identity labels were transferred to the crypto dataset utilizing the Seurat

1044   label transfer functionality. Imputed cell identity labels were carefully inspected and again

1045   confirmed by marker expression. The same clustering and dimensional reduction approach was

1046   performed on the subgroup of spermatogonial cells of the normal and crypto dataset.

1047   **Differential gene expression analysis**

1048   Intra- and inter-dataset differential gene expression analysis was performed using MAST *(53)*.

1049   The original Alevin expression counts (after default Seurat normalization) were used to avoid

1050   bias by the scaling effect of sample-wise sctransform normalization or sample integration. Filter

1051   criteria for differentially expressed genes (DEG) were: default log fold change threshold of

1052   0.25, FDR-corrected p-value below 0.01 and expression in at least 10% of the cells of one

1053   comparison group.

PAGA trajectory inference

1055   Partition based graph abstraction (PAGA) *(54)* from the Scanpy package *(55)* was used for

1056   trajectory inference. The neighborhood graph was based on the top 25 PCs from Seurat with

1057   the number of nearest neighbors set to 20. Cell identity labels were used as partition categorical

1058   for the PAGA construction. Based on the PAGA, a new UMAP embedding was computed to

1059   better reflect the global topology, using 100 optimization iterations (epochs).

RNA velocity estimation and embedding

1061   RNA velocity is the time derivative of the measured mRNA abundance (mature spliced/ nascent

1062   unspliced transcripts) and allows to estimate the future developmental directionality of each

1063   cell. As the abundance estimation relies on genomically aligned nascent reads, it is not possible

1064   to use the transcriptomic pseudo-alignment of Alevin. Therefore, we used the 10x Genomics

1065   Cell Ranger pipeline to create a genomic BAM file for each sample. Subsequently, aligned

1066   reads were used separately as input for velocyto *(17)*, as well as the 10x Genomics genomic

1067   annotation file and UCSC expressed repeats annotation file. Velocyto analysis of each sample

1068   was limited to cell barcodes validated by Alevin. Due to differences in the barcode correction

1069   methods of Alevin and Cell Ranger, RNA velocity could not be estimated for a total of 24 cells

1070   (0.08% of total cells). Per sample abundance estimates were combined with the loompy

1071   package, analogous to the scheme in the Seurat integration step. Combined abundance data was

1072   merged with the PAGA anndata object using scVelo *(56)*. Abundance estimates were

1073   normalized and filtered, requiring a minimal abundance of 5 per gene, simultaneously in the

1074   spliced and unspliced counts. First and second order moments were calculated utilizing the first

1075   30 PCs and 30 neighboring cells from the neighbor graph. The likelihood-based dynamical

1076   model from scVelo was applied, recovering full splicing kinetics before estimating the

1077   velocities. In addition to the velocity graph and the projection of the velocities into the UMAP

1078   embedding, a gene-shared latent time was calculated. The latent time is solely based on the

1079   transcriptional dynamics and represents the internal clock of a cell. For the latent time

1080   calculation, a cell from the most undifferentiated cell type was used as a root cell, if applicable.

1081   **Gene regulatory network inference**

1082   The python version of the SCENIC *(16)* pipeline was used for gene regulatory network

1083   inference. Briefly, SCENIC links cis-regulatory sequences to single-cell gene expression data

1084   to predict interactions between transcription factors (TFs) and target genes. Pairs of genes and

1085   co-expressed TFs were identified with the GRNBoost2 algorithm from Arboreto *(57)*, using the

1086   standard normalized Alevin counts. Additionally, a list of 1390 curated human TFs was used,

1087   provided in the SCENIC repository. Co-expression modules derived from this analysis were

1088   pruned to remove indirect targets and false positives (using human motifs v9 and hg38

1089   databases from cisTargetDB). In accordance with the SCENIC publication we refer to modules

1090   with significant motif enrichment of the correct upstream regulator as 'regulons'. Cells with

1091   enriched expression for genes in a regulon were marked as active for this specific regulon.

1092   Additionally, a Jensen-Shannon divergence based regulon specificity score *(58)* (RSS) was

1093   used for ranking purposes.

1094   As we experienced highly variable results for different SCENIC runs on the same data, we

1095   choose a strategy to increase the stability of the adjacency calculation, i.e. the target gene and

1096   transcription factor co-expression measurement and weighting. Because of its nondeterministic

1097   nature, we decided to perform this calculation 10 times and only carry over gene–TF links that

1098   appeared in each of the runs. This "meta-adjacency" table, consisting of the median weight per

1099   link, was used in subsequent SCENIC steps.

Differential expression analysis along latent time

We utilized the R package tradeSeq *(18)* to perform differential expression analysis along the common gene-shared latent time between the spermatogonial cells of the normal and crypto datasets. Standard normalized Alevin counts were filtered, excluding genes with a total expression below 10 (excluding 28% of 33 679 genes). For each gene, a negative binomial generalized additive model was fitted to seven knots along the latent time, in accordance to the results of evaluateK(). The knots are equally distributed among the cell density along the trajectory, with the first and last knot representing the minimal and maximal trajectory value, respectively. Knots were comprised to knot groups, where all cells between adjacent knots are included. We focused on identifying genes differentially expressed exclusively in a single knot group. For this, we adopted the stageR *(59)* two-staged testing scheme, performing a whole-trajectory patternTest() for the screening stage and a earlyDETest() for each knot group in the confirmation stage. All tests were performed against a foldchange of $\log_2 (1.2)$ with an overall false discovery rate of 0.05. The fitted distributions of the significant genes of each earlyDETest() were clustered, using 100 cluster points and a minimal cluster size of 20, except for knotgroup 1, where a minimal cluster size of 10 was used.

Correlation analysis of spermatogonial cell types

Cell types of the subset of spermatogonial cells were assigned based on the conjunction of markers of spermatogonial states, as described by Guo et al. and Sohni et al. *(9, 11)*. To asses that the cell type assignment worked as expected, correlation analysis of spermatogonial states between our dataset and the datasets of Guo and Sohni was performed. Spearman correlation of common genes was based on an "average" cell, consisting of the genewise mean of all cells of the specific state.

Cell-cell communication analysis

1124   Analysis of cell-cell communication was performed with CellphoneDB *(15)* (Version 2.1.2),

1125   using the "statistical_analysis" method. Ligands and receptors were required to be expressed in

1126   at least 50% of the cells (--threshold) and the number of iterations was set to 10 000. Significant

1127   interactions ($p < 0.05$) were subsequently visualized.

1128   **Gene ontology and pathway analysis**

1129   Gene ontology analysis was performed using the Gene Set Enrichment Analysis (GSEA)

1130   (https://www.gsea-msigdb.org/gsea/index.jsp). Pathway analysis was performed using

1131   Ingenuity Pathway Analysis (Qiagen).

1132   **Sample preparation and LC-MS/MS analysis**

1133   Single cell suspensions obtained by digestion of a biopsy (see "Single cell suspension

1134   preparation" section) from either normal (n=5) or crypto patients (n=4) were prepared for

1135   bottom-up mass spectrometry analysis, combining a one-pot sample preparation method (iST-

1136   NHS; Preomics) with isobaric tandem mass tags (TMT) labeling and subsequent stepwise

1137   reversed-phase fractionation at high pH (Thermo Fisher Scientific, Pierce, Cat# 84868). Cell

1138   lysis, digestion and labeling were performed according to the manufacturer's protocol, except

1139   for the two samples containing $1.0x10^6$ cells (assuming a protein content of ~200 µg), where

1140   volumes of buffers and reagents were adjusted accordingly (N1, N2). In addition, to support

1141   complete degradation of released chromatin, 1 µl of Benzonase was added to each sample after

1142   the initial lysis/denaturation step and incubated for 10 minutes at room temperature (RT). TMT

1143   labeling was performed following digestion by directly adding the TMT labeling reagent

1144   resuspended in 41 µl dry acetonitrile to the corresponding digest (RT, 500 rpm, 1hr). Prior to

1145   quenching the labeling reaction with 5% hydroxylamine, a "label check" was performed by

1146   analyzing 1 µl of each sample by mass spectrometry. A second mass spectrometry analysis was

1147   performed by mixing 1 µl of each sample to obtain normalization factors for the correct 1:1

50

1148   ratio across all channels. Following desalting and lyophilization of the mixture, labeled peptides

1149   were resuspended in 300 µl of 0.1% TFA and submitted to a reversed-phase fractionation

1150   procedure at high pH on spin columns according to the manufacturer's instructions (5% - 50%

1151   acetonitrile, 0.1% triethylamine). This resulted in 8 fractions (plus desalted flowthrough and 2

1152   wash fractions) that were dried down in an Eppendorf concentrator before being resuspended

1153   in 0.1% formic acid for mass spectrometry analysis. All samples were measured twice as

1154   technical replicates. Samples were measured on an Easy nLC 1200 system coupled to a Q

1155   Exactive HF MS via a Nanospray Flex ion source (ThermoFisher Scientific). Peptides were

1156   dissolved in buffer A (0.1% formic acid) and separated on a 25 cm column, in-house packed

1157   with 1.9 µm C18 beads (Reprosil -Pur C18 AQ, Dr. Maisch) using a multi-linear gradient from

1158   5%-18% buffer B (80% acetonitrile; 0.1% formic acid) and from 18%-40% B in 55 min each,

1159   followed by an increase to 60% B in 10 min, a final washout for 7 min at 90% B and re-

1160   equilibration at starting conditions (100% buffer A; flow rate 250 nl/min). The Q-Exactive HF

1161   mass spectrometer was operated in data-dependent acquisition mode (spray voltage 2.1 kV;

1162   column temperature maintained at 45°C using a PRSO-V1 column oven (Sonation, Biberach)).

1163   MS1 scan resolution was set to 120 000 at m/z 200 and the mass range to m/z 350-1 600. AGC

1164   target value was 3E6 with a maximum fill time of 50 ms. Fragmentation of peptides was

1165   achieved by higher-energy collisional dissociation (HCD) using a top15 method (MS2 scan

1166   resolution 60 000 at 200 m/z; AGC Target value $1E^5$; maximum fill time 108 ms; isolation

1167   width 1.3 m/z; normalized collision energy 32). Dynamic exclusion of previously identified

1168   peptides was allowed and set to 30 s, singly charged peptides and peptides assigned with a

1169   charge > 8 were excluded from the analysis. Data were recorded using Xcalibur software

1170   (ThermoFisher Scientific).

1171   **MS data analysis and quantification**

51

1172   Raw MS files were processed using MaxQuant (version 1.6.6.0) *(60)*. Identification of peptides

1173   and proteins was enabled by the built-in Andromeda search engine by querying the

1174   concatenated forward and reverse mouse Uniprot database (UP000005640_9606.fasta; version

1175   from 04/2019) including common lab contaminants. Allowed initial mass deviations were set

1176   to 7 ppm and 20 ppm, respectively, in the search for precursor and fragment ions. Trypsin with

1177   full enzyme specificity and only peptides with a minimum length of 7 amino acids was selected.

1178   A maximum of two missed cleavages was allowed; the 'match between runs' option was turned

1179   on. Carbamidomethylation (Cys) was set as fixed modification, while Oxidation (Met) and N -

1180   acetylation at the protein N-terminus were defined as variable modifications. For peptide and

1181   protein identifications, a minimum false discovery rate (FDR) of 1% was required. First, the

1182   list of identified proteins was filtered and potential contaminants, reverse hits derived from the

1183   target-decoy search as well as proteins that were identified only by a single modified peptide

1184   were removed. Only proteins identified by at least one unique peptide were retained for further

1185   analysis. Next, a sample loading normalization step was performed using an R script

1186   (https://pwilmart.github.io/TMT_analysis_examples/CarbonSources_MQ.html), which was

1187   adapted to our own experimental setup. The list of proteins containing the normalized reporter

1188   ion intensities was imported into the Perseus bioinformatics tool. Proteins that did not show

1189   reporter ion intensities in all 10 TMT channels were removed, and reporter ion intensities were

1190 $\log_2$ transformed. To screen for differentially expressed proteins, a two-sided Student's t-test

1191   was performed using permutation-based FDR calculation as truncation (FDR = 0.05), also

1192   taking into account the contribution of the difference of the means (s0 = 0.1).*(61)*

1193

1194

1195

1196

52

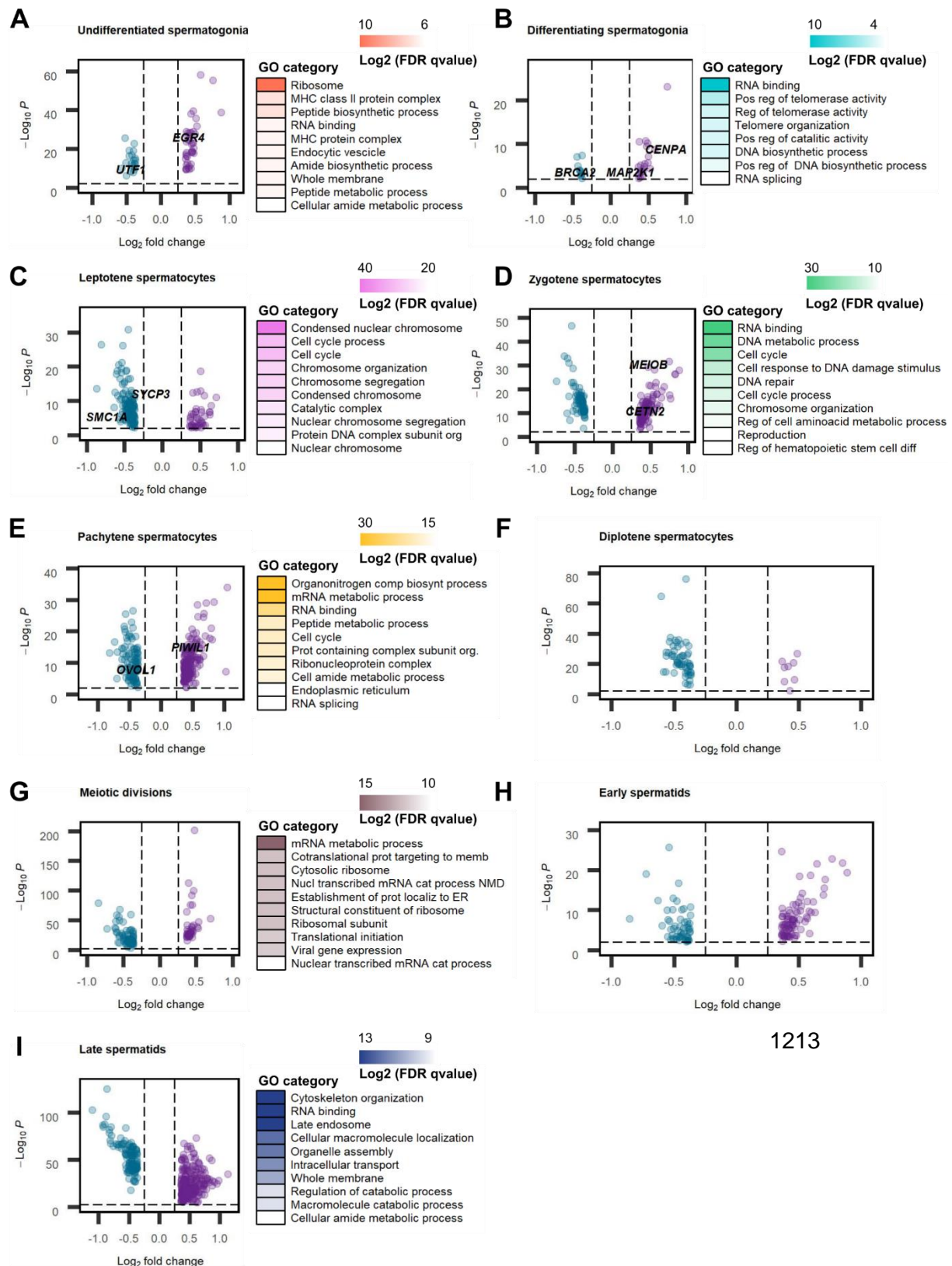1197 **Supplementary figures**
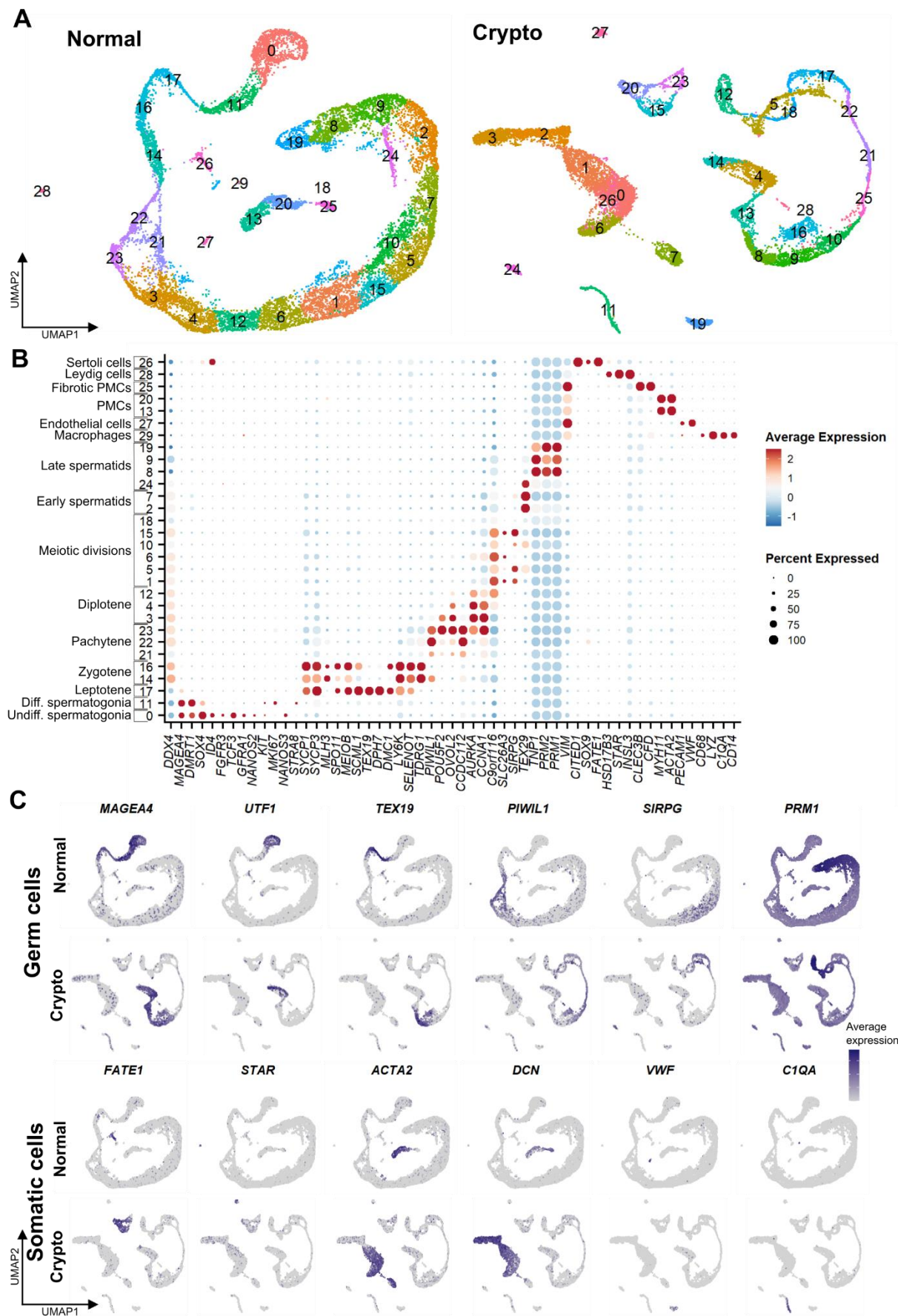
1198

1199 **Fig. S1. Clustering analysis of normal and crypto datasets.**

1200 **(A)** Uniform manifold approximation and projection (UMAP) plot of the clustering analysis of

1201 the integrated (left) normal (15 546 cells, 30 clusters) and (right) crypto (13 144 cells, 29

1202 clusters) datasets. **(B)** Dot plot showing the relative expression of 55 marker genes in the 30

1203 normal clusters. Cell identity was assigned according to marker expression. The color and size

1204 of the dots represent the average expression and the percentage of cells expressing each marker

1205 in a cluster, as indicated in the key. **(C)** Feature plots showing the expression of germ cell and

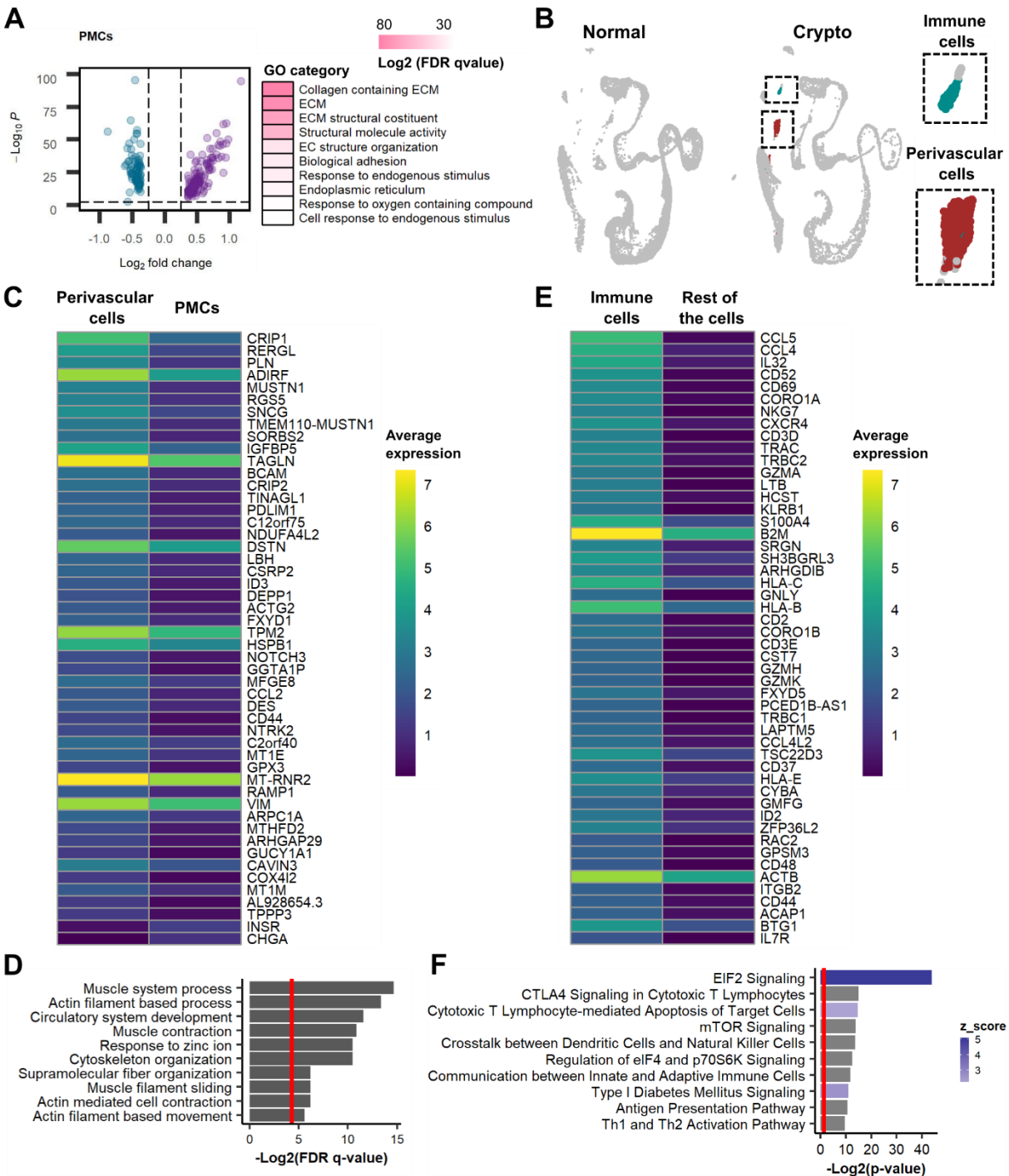1206 somatic cell marker genes in the normal and crypto datasets.

1213

1216

1217

1218 **Fig. S2. DGE and GO analysis between the different germ cell clusters of the normal**
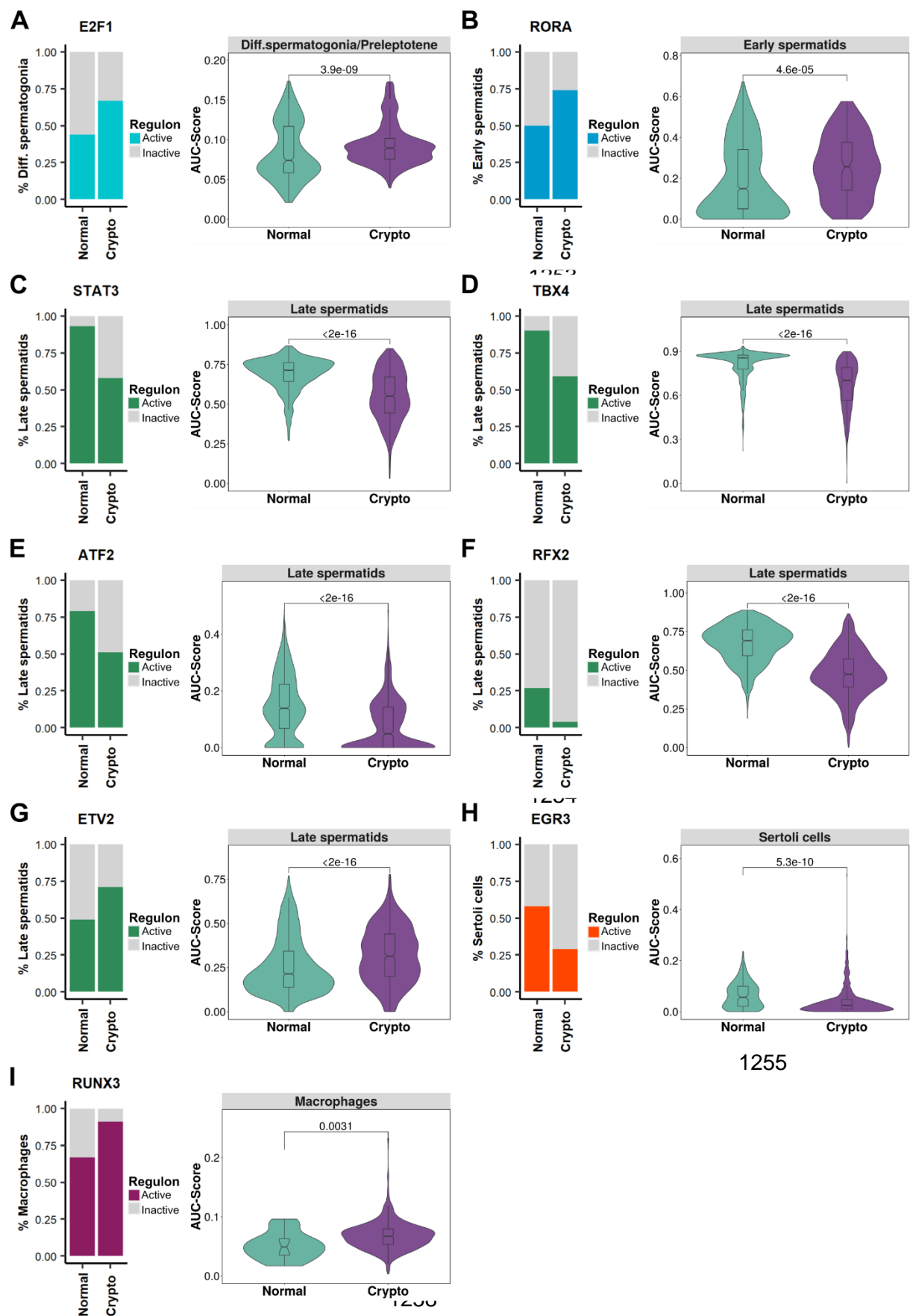
1219 **and crypto datasets.**

1220 **(A)** Volcano plot (Left) and gene ontology (GO) analysis (Right) of the 61 genes uniquely

1221 differentially expressed in the undifferentiated spermatogonial cluster. **(B)** Volcano plot (Left)

1222 and GO analysis (Right) of the 32 genes uniquely differentially expressed in the differentiating

1223 spermatogonial cluster. **(C)** Volcano plot (Left) and GO analysis (Right) of the 164 genes

1224 uniquely differentially expressed in the leptotene spermatocyte cluster. **(D)** Volcano plot (Left)

1225 and GO analysis (Right) of the 180 genes uniquely differentially expressed in the zygotene

1226 spermatocyte cluster. **(E)** Volcano plot (Left) and GO analysis (Right) of the 271 genes

1227 uniquely differentially expressed in the pachytene spermatocyte cluster. **(F)** Volcano plot of

1228 the 66 genes uniquely differentially expressed in the diplotene spermatocyte cluster. **(G)**

1229 Volcano plot (Left) and GO analysis (Right) of the 121 genes uniquely differentially expressed

1230 in the meiotic division cluster. **(H)** Volcano plot of the 142 genes uniquely differentially

1231 expressed in the early spermatid cluster. **(I)** Volcano plot (Left) and GO analysis (Right) of the

1232 405 genes uniquely differentially expressed in the late spermatid cluster. See table S5 for the

1233 statistical details.

1234

Fig. S3. DGE and GO analysis between the different somatic cell clusters of the normal and crypto datasets.

(**A**) Volcano plot (Left) and gene ontology (GO) analysis (Right) of the 232 genes uniquely differentially expressed in the PMC cluster. See Supplementary information, Table S5 for the statistical details. (**B**) Uniform manifold approximation and projection (UMAP) plot of the

1240   normal (Left) and crypto (Right) datasets showing the unique presence of the peritubular (Red)

1241   and immune (Green) cell clusters in the crypto dataset. **(C)** Heatmap showing the differential

1242   gene expression analysis between the perivascular cell clusters and the PMCs in the crypto

1243   dataset. The analysis identified 48 differentially expressed genes (DEG), including MUSTN1,

1244   with a higher expression in perivascular cells (table S6). **(D)** The bar plot shows the GO analysis

1245   of the 48 DEG found in perivascular cells. The results were enriched in GO terms related to

1246   muscle and cytoskeletal contraction. The red line on the bar plot represent p=0.05. **(E)** Heatmap

1247   showing the first 50 genes of the differential gene expression analysis between immune cells

1248   and the rest of the cells in the crypto dataset. A total of 195 genes showed higher expression in

1249   cluster 24 (A complete list of the DEG is available table S6). **(F)** The bar plot shows the pathway

1250   analysis of the 195 DEG. The results were enriched in immune cell associated pathways. The

1251   red line on the bar plot represents the p=0.05. The z_score represents the activation level of

1252   each pathway.

58

1259

1260   **Fig. S4. Differential regulon activation in the normal and crypto datasets.**

1261   **(A-I)** Left panel: Stacked bar plot comparing the proportion of cells with active regulons. Right

1262   panel: Violin plots comparing the AUC score of the regulons in the normal and crypto dataset.

1263   The regulons are organized as follow: **(A)** E2F1 in the differentiating

1264   spermatogonia/preleptotene, **(B)** RORA in the early spermatids, **(C)** STAT3, **(D)** TBX4, **(E)**

1265   ATF2, **(F)** RFX2 and **(G)** ETV2 in the late spermatids, **(H)** EGR3 in the Sertoli cells and **(I)**

1266   RUNX3 in the macrophages. A significant change was found for all the nine regulons in the
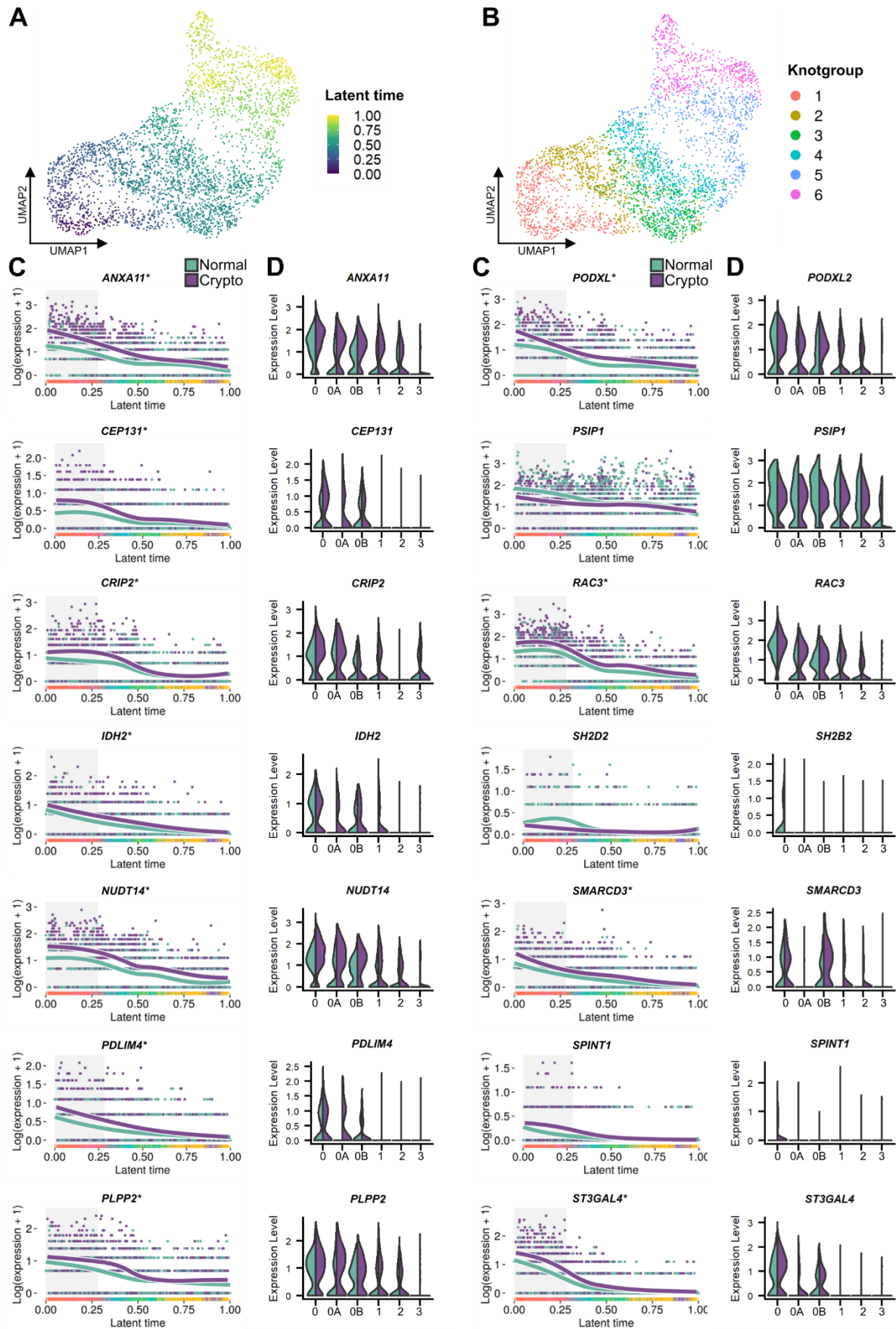
1267   crypto dataset.
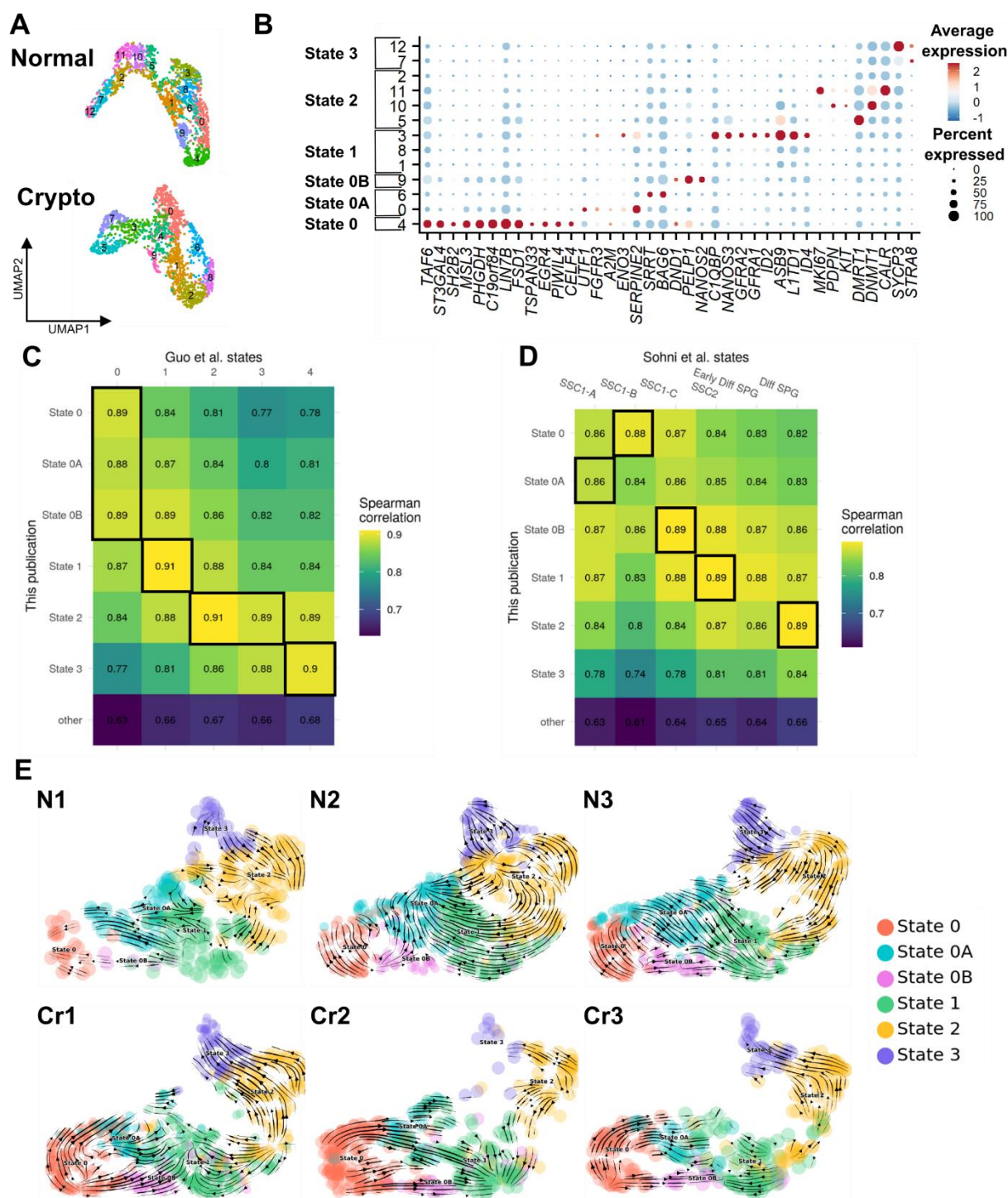
Fig. S5. Cluster analysis and assignment of normal and crypto spermatogonial datasets.

(A) Uniform manifold approximation and projection (UMAP) plot of the integrated normal (upper panel, 13 clusters) and crypto (lower panel, 10 clusters) spermatogonial datasets. The cells are color coded according to the cluster. **(B)** Dot plot showing the relative expression of selected marker genes in the 13 normal clusters. The color and size of the dots represent the

61

1274   average expression and the percentage of cells expressing each marker in a cluster, as indicated

1275   in the key. According to the marker expression, six clusters were defined: State 0, 0A, 0B, 1, 2,

1276   and 3. **(C-D)** Correlation analysis comparing the transcriptome of the different spermatogonial

1277   states defined in this publication with those defined in Guo et al., 2018 **(C)** and Sohni et al.,

1278   2019 **(D)**. The 'other' state is comprised of all non-spermatogonial cell types available in this

1279   publication. The black boxes show the states where we expected the highest correlation in each

1280   state. **(E)** Contribution of each normal and crypto sample to the RNA velocity derived from the

1281   scVelo dynamical model visualized as streamlines in UMAP plots.

1282

1283

1284 **Fig. S6. EGR4-regulated genes in the knotgroup 1 spermatogonia.**

1285 **(A)**Uniform manifold approximation and projection (UMAP) plot showing the integrated

1286 normal-crypto spermatogonial dataset aligned along the latent time. The cells are color-coded

1287 according to their progression along the latent time. State 0 was set as starting point of the

1288 differentiation process. **(B)** UMAP plot showing the subdivision of the integrated normal-

1289 crypto spermatogonial dataset into six knotgroups. The cells are color coded according to their

1290 respective knotgroup. **(C)** Line plots show the expression along the latent time of the EGR4-

1291 regulated differentially expressed genes in normal (teal) and crypto (purple) spermatogonia

1292 included in the knotgroup 1: ANXA11, CEP131, CRIP2, IDH2, NUDT14, PDLIM4, PLPP2,

1293 PODXL2, PSIP1, RAC3, SH2B2, SMARCD3, SPINT1, and ST3GAL4. The gray area

1294 highlights the cells belonging to the knotgroup 1. The asterisk indicates EGR4-regulated genes

1295 that were also differentially expressed in knotgroup 2. **(D)** Double violin plots comparing the

1296 expression levels of EGR4-regulated ANXA11, CEP131, CRIP2, IDH2, NUDT14, PDLIM4,

1297 PLPP2, PODXL2, PSIP1, RAC3, SH2B2, SMARCD3, SPINT1, and ST3GAL4 in the

1298 spermatogonial states of the normal (teal) and crypto (purple) datasets.

64

1299

1300

1301
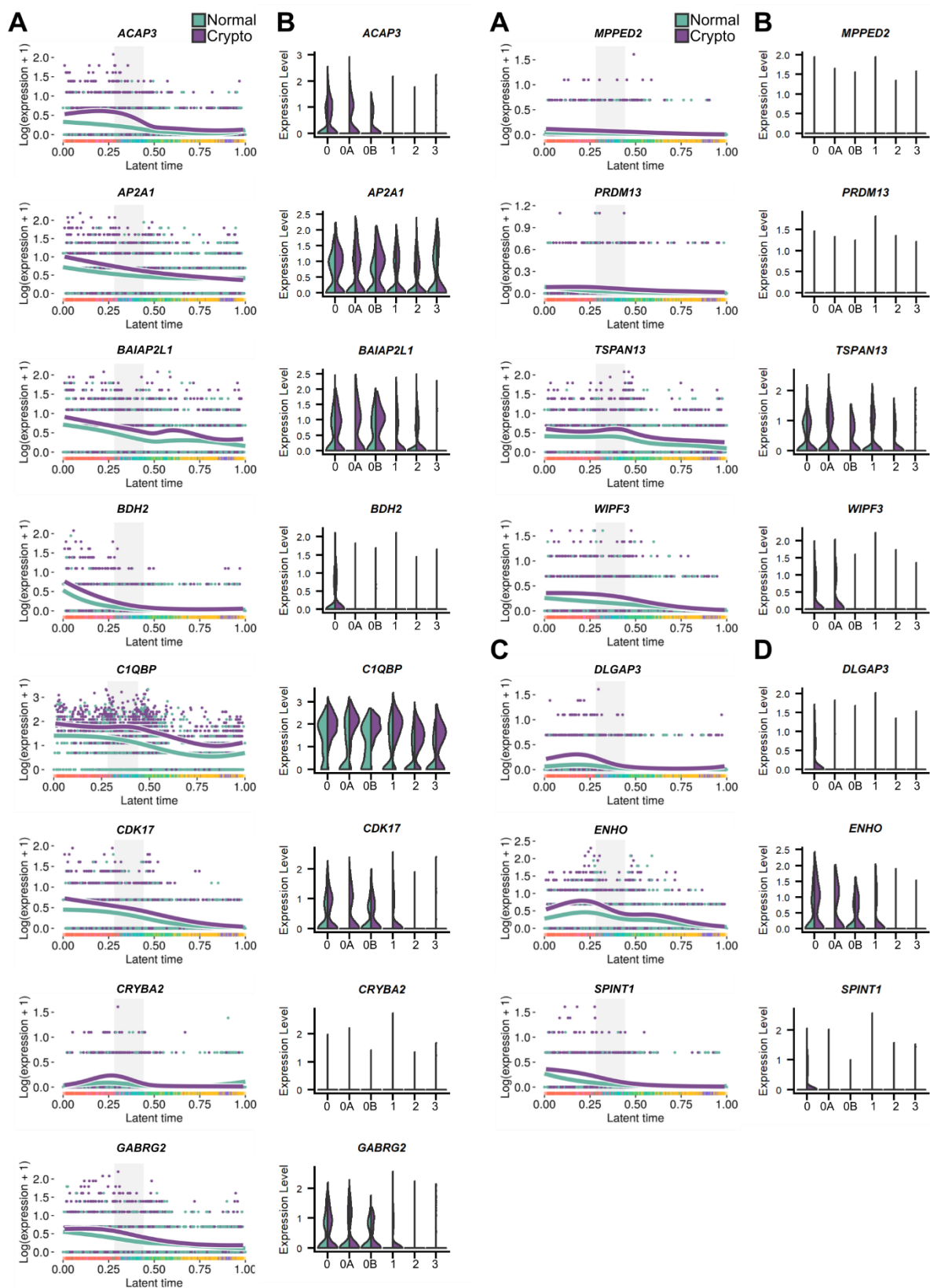
1302 **Fig. S7. EGR4- and HOXC9-regulated genes in knotgroup 2 spermatogonia.**

1303 **(A)** Line plots showing the expression along the latent time of EGR4-regulated, differentially

1304 expressed genes between the normal (teal) and crypto (purple) spermatogonia in the knotgroup

1305 2: ACAP3, AP2A1, BAIAP2L1, BDH2, C1QBP, CDK17, CRYBA2, GABRG2, MPPED2,

1306 PRDM13, TSPAN13, and WIPF3. The gray area highlights the cells belonging to the knotgroup

1307 2. **(B)** Double violin plots comparing the expression levels of EGR4-regulated genes ACAP3,

1308 AP2A1, BAIAP2L1, BDH2, C1QBP, CDK17, CRYBA2, GABRG2, MPPED2, PRDM13,

1309 TSPAN13, and WIPF3 between the normal (teal) and crypto (purple) spermatogonial states.

1310 **(C)** Line plots showing the expression along the latent time of the EGR4- and HOXC9-

1311 regulated, differentially expressed genes between the normal (teal) and crypto (purple)

1312 spermatogonia in knotgroup 2: DLGAP3, ENHO, and SPINT1. The gray area highlights the

1313 knotgroup 2. **(D)** double violin plots comparing the expression levels of EGR4- and HOXC9-

1314 regulated genes DLGAP3, ENHO, and SPINT1 between the normal (teal) and crypto (purple)

1315 spermatogonial states.

1316

1317 **Supplementary tables**

1318 Data files table S1-S10