

1 **Molecular insights into the pathways underlying naked mole-rat eusociality**  
2  
3 Eskeatnaf Mulugeta<sup>1,2\*</sup>, Lucile Marion-Poll<sup>1,2</sup>, David Gentien<sup>1,2</sup>, Stefanie B. Ganswindt<sup>3</sup>, André  
4 Ganswindt<sup>3</sup>, Nigel C. Bennett<sup>4</sup>, Elizabeth H. Blackburn<sup>5</sup>, Chris G. Faulkes<sup>6\*</sup>, Edith Heard<sup>1,2</sup>  
5

6 **Author Affiliations**

7 1 *Institut Curie, PSL Research University, CNRS UMR3215, INSERM U934, Mammalian  
8 Developmental Epigenetics group, F-75005 Paris, France*

9 2 *Sorbonne Universités, UPMC Univ Paris 6, F-75005 Paris, France*

10 3 *Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, South Africa*

11 4 *Department of Zoology and Entomology, University of Pretoria, Pretoria 0002, South Africa*

12 5 *Salk Institute for Biological Studies, California, USA*

13 6 *School of Biological and Chemical Sciences, Queen Mary University of London, England*

14 \* *Corresponding authors*

15

16 Correspondence to: Eskeatnaf Mulugeta ([eskeatnaf.mulugeta@curie.fr](mailto:eskeatnaf.mulugeta@curie.fr) or [eskeww@gmail.com](mailto:eskeww@gmail.com))

17 Chris G. Faulkes ([c.g.faulkes@qmul.ac.uk](mailto:c.g.faulkes@qmul.ac.uk))

18

19

20

21 **Keywords:**

22 Naked mole-rat, *Heterocephalus*, Bathyergidae, Eusociality, Aromatase, Dopaminergic pathways,  
23 Prolactin, Hyperprolactinemia, Reproductive suppression

24

25

26 **Abstract**

27 **Background:**

28 Eusociality is the highest level of social organization and naked mole-rats (NMRs) are amongst the few  
29 mammals showing this unique social behavior; nevertheless, little is known about the molecular  
30 mechanisms underlying the eusociality of NMRs.

31 **Results:**

32 Gene expression profiling of NMR brain and gonads (ovary and testis), from animals belonging to  
33 different reproductive castes, revealed robust gene expression differences between reproductive and non-  
34 reproductive members of NMR colonies. In the brain, dopaminergic pathways appear to be potential  
35 players in NMR eusocial behaviour. Breeding animals (queens and breeding males) showed increased  
36 expression of genes involved in dopamine metabolism. Using immunohistochemistry, we notably found  
37 these differences to be in dopaminergic hypothalamic areas, which provide inhibitory control over the  
38 secretion of prolactin, amongst other regions. Furthermore, plasma prolactin concentrations were elevated  
39 in many non-breeders (of both sexes), often reaching levels exceeding that of pregnant or lactating  
40 queens, suggesting a role for hyperprolactinaemia in socially-induced reproductive suppression. We also  
41 found that the ovaries of non-breeding females are arrested at pre-pubertal stage. They contained fewer  
42 supporting stromal cells compared to queens, and had very low expression of the aromatase gene  
43 *Cyp19A1* (a key enzyme in estrogen synthesis) compared to non-breeding females. In the testes, genes  
44 involved in post meiosis spermatogenesis and sperm maturation (*Prm1*, *Prm2*, *Odf3* and *Akap4*) were  
45 highly expressed in breeding males compared to non-breeders, explaining the low sperm number and  
46 impaired sperm motility characteristic of non-breeding males.

47 **Conclusions:**

48 Our study suggests that extreme reproductive skew, one of the defining features of eusociality, is  
49 associated with changes in expression of key components of dopamine pathways, which could lead to  
50 hypogonadism and a lifetime of socially-induced sterility for most NMRs.

51

52 **Background**

53 Eusociality is the highest level of social organization and is perceived as one of the major transitions in  
54 the evolution of life [1, 2]. The original definition, derived from studies of social insects, requires three  
55 criteria: a reproductive division of labour, overlapping generations, and cooperative care of young [3, 4],  
56 although it has been argued that physically distinct morphological castes should also be present [5]. This  
57 extraordinary cooperative way of life is famously observed in invertebrates, such as the social  
58 hymenoptera, but also other diverse groups such as crustaceans, and more recently, some mammals [6-9].  
59 The importance of eusociality as a strategy is exemplified by the fact that eusocial species may constitute  
60 75% of the insect biomass in some ecosystems [10, 11]. The taxonomic diversity of eusociality has led to  
61 much debate about its definition, especially among mammals (for review see [12]), but it is generally  
62 accepted that among the African mole-rats (Family: Bathyergidae), the naked mole-rat (*Heterocephalus*  
63 *glaber*) and Damaraland mole-rat (*Fukomys damarensis*) have independently evolved eusociality [8, 13,  
64 14].

65

66 NMRs live entirely underground in the arid regions of East Africa in colonies that may contain up to 300  
67 individuals [15]. Most colonies of NMRs have only one breeding female (the queen), who mates with one  
68 to three selected males, and the rest of the colony (both sexes) are reproductively suppressed [8, 16]. The  
69 reproductive status of the queen and the breeding males may be stable for many years - NMRs can live up  
70 to 32 years in captivity [17]. The queen is dominant in the colony social hierarchy, and evidence suggests  
71 that she exerts a dominant control mechanism of reproductive suppression over the non-breeders of both  
72 sexes, and possibly also the breeding male [18-21]. The remaining members of the colony of both sexes  
73 are morphologically very similar (externally) and do not exhibit sexual behavior, but perform tasks that  
74 are essential for the survival and wellbeing of the colony: foraging, colony defense, maintenance of the  
75 tunnel system, and care of the young [8, 16]. It has been estimated that more than 99% of non-breeders  
76 never reproduce [16]. However, the extreme socially-induced suppression of reproduction is reversible

77 upon removal of the suppressing cues: if the queen or breeding males die or are removed from the colony,  
78 or if non-breeding animals are housed singly or in pairs [18, 20-23].

79  
80 Non-breeding females remain at pre-pubertal anovulatory state despite attaining adult body size, with  
81 small uteri and ovaries, and reduced plasma luteinizing hormone (LH) concentrations that is reflected in  
82 lower plasma and urinary progesterone levels [18, 19, 24]. Upon removal of the queen, and often after  
83 fighting and competition among non-breeding females, one non-breeding female attains breeding status,  
84 and undergoes anatomical, behavioral, and endocrine changes, including elongation of the body,  
85 perforation of the vagina, increased dominance and aggression, and activation of the ovaries and  
86 ovulation, with resulting increased urinary and plasma progesterone [22, 24-26]. The degree of  
87 suppression in non-breeding males is less pronounced than the suppression in non-breeding females in  
88 that gametes are produced. Non-breeding males have lower urinary testosterone levels, plasma luteinizing  
89 hormone (LH), and lower reproductive tract to body mass ratio compared to breeders [20, 27].  
90 Nevertheless, non-breeding males are able to undergo spermatogenesis and mature spermatozoa are  
91 detected in their epididymis and vas deferens, although their number is lower than in breeding males and  
92 the majority are non-motile [28]. This indicates that the low concentration of testosterone and LH in non-  
93 breeding males is sufficient to support the spermatogenetic cycle, but not the final stages of maturation  
94 steps (pituitary FSH may also be inhibited, but to date has not been measured in NMRs). As with females,  
95 suppressive effects are reversible upon withdrawal of non-breeding males from their parent colonies and  
96 housing singly or pairing with a female - increases in concentrations of urinary testosterone levels and  
97 plasma luteinizing hormone (LH) occur rapidly [20].

98  
99 The detailed mechanism underlying this reproductive suppression remains elusive. The central role of  
100 hypothalamic gonadotrophin releasing hormone (GnRH) in integrating environmental cues is well  
101 established [29-31], while reproductive development and puberty may be dependent on another  
102 hypothalamic peptide, kisspeptin (acting via GnRH) [32]. In NMRs, a lack of priming of the pituitary

103 gland by impaired release of hypothalamic GnRH may be a key component in reproductive suppression  
104 [18, 27]. Nevertheless, GnRH is still produced: the number of GnRH-1 immunoreactive cell bodies does  
105 not differ between breeding and non-breeding NMRs within or between the sexes [33]. However,  
106 breeding females have greater numbers of kisspeptin cell bodies in key areas of the hypothalamus,  
107 suggesting that emergence from a socially-induced hypogonadotropic/prepubertal state in female NMRs  
108 may involve kisspeptin-related pathways. Recently, elevated levels of the RF amide related protein 3  
109 (RFRP-3 or GnIH) in the brain of non-breeders have been implicated as a component in the suppression  
110 of reproduction in NMRs, through the inhibition of GnRH secretion [34]. It is currently unknown how  
111 behavioural and other sensory cues (such as signature odours and vocalisations) mediate the extreme  
112 social suppression of reproduction in NMRs.

113

114 In recent years, gene expression profiling of several eusocial insects have provided insight into the  
115 molecular mechanisms and evolutionary paths to eusociality in insects [35-42]. However, although  
116 genomic and transcriptomic analyses of NMRs and other mole-rats have also been published, these have  
117 focused on individual animals or cross species comparisons [43-45]. So far no comprehensive  
118 comparative analysis within and among sexes and reproductive castes of NMRs has been reported. Here  
119 we use RNA-sequencing (RNA-seq) to undertake the extensive transcriptome profiling of breeding and  
120 non-breeding NMR brains and gonads (ovary and testis). This has revealed striking gene expression  
121 differences that point to the possible mechanisms underlying eusociality in a mammal, and extreme  
122 socially-induced reproductive suppression.

123

## 124 **Results**

125

### 126 **Changes in the dopaminergic system are strongly associated with NMR reproductive status**

127 In order to determine the origin of the behavioral and reproductive differences between NMR social  
128 classes and elucidate the molecular pathways that contribute to these differences, we performed gene

129 expression profiling using very high coverage RNA-Seq on whole brains of several animals that  
130 constitute both sexes and different reproductive classes (*Additional file 1A*: animals used and *Additional*  
131 *file 1B*: RNA-seq information). We first explored the global expression levels, and based on principal  
132 component analysis (PCA), the first principal components clearly captured the reproductive caste  
133 differences: with the queens and other castes distinctly clustering in their respective groups. Similarly,  
134 using a hierarchical clustering technique, non-pregnant NMR queens (Qs) cluster separately to the rest of  
135 the colony members (*Supplementary Figure 1A,B*). The rest of the colony members - non-breeding  
136 females (NBFs), non-breeding males (NBM), breeding males (BM), and a pregnant Q clustered as  
137 another group, with the majority of NBFs and NBM clustering as one group, independent of sex  
138 (*Supplementary Figure 1A,B*). In order to define the gene expression differences between brains, we  
139 performed differential expression analysis (see Methods) and identified several genes that show  
140 significant gene expression differences (FDR<0.05 Benjamini-Hochberg multiple testing correction; a  
141 log<sub>2</sub> fold change >1; and >1 CPM, counts per million) between breeding and non-breeding animals  
142 (*Figure 1A,B*). The number of differentially expressed genes (DEGs) follows social and reproductive  
143 status (*Figure 1A,B*), with the highest number of DEGs between breeding animals (Qs vs BMs), or  
144 breeding animals compared to non-breeding animals, and with minor differences between subordinate  
145 animals (*Figure 1A,B*).

146  
147 To elucidate the nature of these gene expression differences in more detail we examined each group in  
148 pairwise comparisons. We first focused on differences between Qs (n=2) and NBFs (n=3). Based on a  
149 global gene expression comparison, the Qs and NBFs clearly cluster as two distinct groups, suggesting  
150 the clear gene expression difference in the brain of Qs and NBFs (*Supplementary Figure 2A,B*). In total,  
151 854 genes were differentially expressed (389 higher expression in the Qs and, 465 lower expression in  
152 Qs) (*Figure 1A,B see Qs vs NBFs, Supplementary Figure 2C, Additional file 2*). These DEGs are enriched  
153 for pathway and biological process terms that include synaptic signaling, dopaminergic synapse, positive  
154 regulation of transport, and neurotransmitter receptor activity (*Supplementary Figure 2D,E,F,G, and*

155 *Additional files 3, 4, 5*). Interestingly, genes related to dopamine were particularly enriched in the DEGs.  
156 Dopamine is a catecholamine (a monoamine compound derived from the amino acid tyrosine) and has a  
157 neurotransmitter function. Several key-genes, such as the tyrosine hydroxylase (*Th*) gene (the rate-  
158 limiting enzyme in the synthesis of catecholamines), the vesicular monoamine transporter (*Vmat2* or  
159 *Slc18a2*) which transports monoamine neurotransmitters into the vesicles to be released at the synapse,  
160 and the dopamine transporter (*Dat* or *Slc6a3*) responsible for the reuptake of dopamine were found to be  
161 highly expressed in the Qs compared to the NBFs (*Additional file 2*). *Th* and *Dat* were both in the top 10  
162 of our DEG list, suggesting higher dopamine production in the Qs brain. In addition, the most abundant  
163 dopamine receptors *Drd1a*, *Drd2* and *Drd3* were all down-regulated in the Qs, while the *Drd5* receptor  
164 gene was up-regulated (*Additional file 2*). Genes related to the other catecholamines adrenalin and  
165 noradrenalin (dopamine beta-hydroxylase *Dbh*, phenylethanolamine N-methyltransferase *Pnmt*, and the  
166 different adrenergic receptors alpha and beta) were not differentially expressed between Qs and NBFs,  
167 indicating that the differences are specific to the dopaminergic system.

168  
169 Next, we looked at the gene expression differences between BMs and NBMs. As breeding status of male  
170 NMRs can be difficult to determine [28], we used two criteria to define their reproductive status in the  
171 colony: 1) long-term observational information, and 2) testis to body mass ratio (*Supplementary Figure*  
172 *3A*). On PCA and hierarchical clustering (using global gene expression), unlike the distinction between  
173 Qs and NBFs, BMs and NBMs do not fully resolve into separate clusters (*Supplementary Figure 1B*,  
174 *Supplementary Figure 3B,C*). For animals that fulfill the two criteria above (3 BMs and 3 NBMs,  
175 excluding one NBM whose status was uncertain), we performed differential gene expression analysis and  
176 identified 193 genes (with 152 showing higher expression in BMs and 42 lower expression in BMs;  
177 *Figure 1A B see Qs vs BMs, Supplementary Figure 3 D, Additional file 6*). These DEGs are enriched for  
178 pathways and biological processes that are involved in nitrogen compound transport, regulation of  
179 nervous system development, synaptic signaling, dopaminergic neuron differentiation, behavior  
180 (*Supplementary Figure 3E,F,G,H and Additional files 7, 8, 9*). Similarly to the comparison between Qs vs

181 NBFs, we also found *Th* and *Dat* to be amongst the most highly expressed genes in BMs compared to  
182 NBMs (*Additional file 6*), suggesting that similar mechanisms, involving the dopaminergic pathway, may  
183 account for the behavioral and reproductive differences observed between BMs and NBMs.

184

185 Breeding animals (Qs and BMs) are the most dominant and aggressive in the NMR colony. These  
186 animals, Qs and BMs, clearly cluster into independent groups (*Supplementary Figure 1A,B*,  
187 *Supplementary Figure 4A*), and also show the highest number of DEGs (*Figure 1A, B, see Qs vs BMs*). In  
188 total 1246 genes show significant expression difference between Qs and BMs (444 higher expression in  
189 Qs and 802 lower expression in Qs; *Figure 1A,B, Supplementary Figure 4C, and Additional file 10*).  
190 These DEGs are enriched for pathways and biological processes that are involved in protein targeting to  
191 membrane, mRNA metabolic process, cellular macromolecule catabolic process, regulation of cellular  
192 response to stress and related pathways (*Supplementary Figure 4D,E,F,G, and Additional file 11, 12, 13*).  
193 The enrichment in these pathways, biological processes, and molecular functions, when breeding females  
194 are compared to breeding males, is very different from the findings in non-breeding animals. In contrast  
195 to the major gene expression differences in the brains of breeding animals (Qs and BMs), non-breeding  
196 animals (NBMs and NBFs) show very similar global expression profiles (*Supplementary Figure 1A,B*,  
197 *Supplementary Figure 5A*), with only very few genes showing significant gene expression differences  
198 (only 28 DEGs; *Figure 1A,B see NBFs vs NBMs, Supplementary Figure 5C, Additional file 14*). This is in  
199 agreement with previous observations showing the absence of behavioral differences (including sexual  
200 characteristics), and sexual differentiation (including similarity in external genitals) between NBMs and  
201 NBFs [46].

202

203 We went on to identify gene expression profiles that are specific to dominant breeding Qs (Q genes). Q  
204 genes were identified by taking genes that show differential expression in the comparison between Qs vs  
205 NBFs, Qs vs NBMs (*Additional file 15*), and Qs vs BMs (*Figure 1A,B, Supplementary Figure 6 A,B,C*).  
206 However, as BMs also have reproductive status and are high in the dominance hierarchy of the colony,

207 we focused on genes that are found in the comparison between Qs vs NBFs and Qs vs NBMs (*Figure 1C*,  
208 *Supplementary Figure 6D,E, Additional file 16*). Around half of the genes that were differentially  
209 expressed between Qs vs NBFs are shared with Qs vs NBMs (*Figure 1C*). These Q genes are enriched in  
210 pathways and biological process terms that are involved in behavior, synaptic signaling, multi-organism  
211 behavior, dopaminergic synapse, neurotransmitter transport, dopamine binding, reproductive behavior  
212 and other related terms (*Figure 1D,E, Additional file 17, 18, 19*). We performed a similar analysis to  
213 define genes that are unique to the BMs by taking genes that are common in the comparison between  
214 BMs vs NBFs (*Additional file 20*) and BMs vs NBMs (*Figure 1A,B, Additional file 21*). In general, while  
215 we found fewer such BM genes compared to Q genes (*Supplementary Figure 6F,G,H*), these genes are  
216 involved in similar pathways as the Q genes. Similar to the Q genes, BM genes are also enriched for  
217 visual perception, nitrogen compound transport, behavior, neuron projection morphogenesis,  
218 dopaminergic synapse (*Additional file 22, 23, 24*). In addition, similarly to the Q gene list, *Th* and *Dat*  
219 were also present in the BM gene list (unique in BMs vs NBMs and BMs vs NBFs). These results (Q  
220 genes and BM genes) highlight the possible role of dopamine pathway in the NMR colony social  
221 behavior and reproductive suppression.

222  
223 Our detailed analysis of NMR brain RNA-seq has identified the dopamine system as a probable key  
224 player in the breeding status differences. There are several cell groups in the brain which are potentially  
225 dopaminergic (named A8 to A16), linked to different functions [47]. In order to independently confirm  
226 the differences in the dopamine system between reproductive castes and to explore their localization, we  
227 collected brains from a different batch of animals (*Additional file 1A*). Even though the Qs were bigger in  
228 size compared to the other animals ( $56 \pm 16$  g versus  $34 \pm 3$  g), all the brains had similar sizes with no  
229 obvious neuromorphological differences across the castes. We performed immunostaining against  
230 Tyrosine hydroxylase (TH) on several sections along the rostrocaudal axis (*Figure 2A*). The most caudal  
231 sections contained the A9 (substantia nigra compacta) and A10 (ventral tegmental area), nuclei that  
232 comprise the vast majority of dopaminergic neurons. Their respective main targets, the dorsal striatum

233 and nucleus accumbens, can be seen in the most rostral sections. In addition, we looked at the staining in  
234 the hypothalamic nuclei A12, A13 and A14. We observed that across the whole brain, the intensity of the  
235 TH staining tended to be higher in breeders than non-breeders, though with inter-individual variability.  
236 Strikingly, there was a consistent staining of the hippocampus (and to a lesser extent of the cortex)  
237 specifically in all NMR breeders, but absent in non-breeders (*Figure 2B*), indicating catecholaminergic  
238 innervation of this region. To our knowledge, such a strong hippocampal staining has never been reported  
239 in mice or rats. Importantly, breeders showed significantly more TH expression in periventricular  
240 hypothalamic nuclei A12 and A14 (*Figure 2C*), which are precisely the two dopaminergic cell groups  
241 known to inhibit prolactin (PRL) secretion, a peptide hormone related to lactation and reproduction [48].

242

243 To investigate our prediction that elevated prolactin (hyperprolactinemia) may be a component in the  
244 suppression of reproduction in non-breeding NMRs, we measured plasma PRL in breeding and non-  
245 breeding animals of both sexes in samples taken across thirteen colonies (*Figure 1F*). We found that most  
246 samples from non-breeders had detectable concentrations of PRL, and these often reached very high  
247 levels: NBF (mean  $\pm$  SEM)  $32.64 \pm 6.13$  ng/ml; n=44; range, 0.03-173.57 ng/ml; NBM  $36.77 \pm 9.81$   
248 ng/ml; n=49; range, 0.03-330.30 ng/ml. Among breeders, queens had the expected variance in plasma  
249 PRL concentrations as part of normal ovarian cyclicity, pregnancy and lactation:  $33.02 \pm 12.94$  ng/ml;  
250 n=12; range, 3.60-160.80 ng/ml. Two values were obtained from lactating queens, 21.14 ng/ml (23 days  
251 post-partum, at the end of the period of lactation) and 160.80 ng/ml (seven days post-partum), the latter  
252 being the highest concentration recorded among the breeding female samples. Although the sample size  
253 was low due to the difficulties of identifying them, breeding males had low plasma PRL concentrations,  
254 as predicted from our model of suppression:  $15.91 \pm 6.21$  ng/ml; n=7; range, 3.92-47.92 ng/ml.

255

256 In conclusion, this comparison of gene expression profiles in NMR brains uncovers differences that  
257 follow social status rather than sex, with the Q having a unique gene expression pattern compared to the

258 rest of the colony. Furthermore, we identify dopaminergic pathways as potential key players in NMR  
259 eusocial behavior.

260

261 **Non-breeding female NMRs have pre-pubertal ovary and lack the production of ovarian estrogen**

262

263 To characterize the origin of the differential reproductive behavior of Qs and NBFs, we next focused on  
264 gene expression differences between their ovaries. Mammalian oogenesis is a highly regulated process  
265 that involves several hormonal and molecular pathways. Oogenesis starts in the fetal ovaries with the  
266 development of oogonia from primordial germ cells (PGCs). Each oogonium advances until the first  
267 stages of meiosis (meiosis I) and become arrested at the prophase stage of meiosis I, forming primary  
268 oocytes [49, 50]. After puberty, a few primary oocytes are recruited during each ovarian cycle, and only  
269 one oocyte matures to be ovulated. The maturation process involves the completion of meiosis I, and  
270 generation of secondary oocytes that are again arrested at the metaphase II stage (meiosis II). The second  
271 meiosis will only be finalized after successful fertilization by sperm. Growing oocytes are supported by  
272 surrounding somatic cells (follicular cells, granulosa cells, and theca cells) that produce hormones such as  
273 estrogens, in response to pituitary gonadotrophic hormones (FSH, LH). At birth, the primary oocytes are  
274 embedded into immature primordial follicles, which mature into primary, and secondary follicles. At  
275 puberty, oocytes in meiosis II into tertiary and pre-ovulating Graafian follicle [50]. The ovaries of NMR  
276 NBFs are reported to be at pre-pubertal stage, with mainly primordial follicles, and may be a few  
277 secondary or tertiary follicles [51].

278

279 Based on PCA and hierarchical clustering, Qs (n=3) and NBFs (n=3) ovaries clearly group by status and  
280 cluster separately (*Figure 3A,B and Supplementary Figure 7A*), supporting previous findings reporting  
281 distinct anatomical, endocrine, and physiological differences between the ovaries of Q and NBFs [18, 19,  
282 24]. Irrespective of pregnancy or age, Q ovaries group and cluster together separately from NBFs.  
283 Differential expression analysis between the Qs and NBFs (excluding the pregnant Q ovary), specifically

284 identified 1708 genes, with more genes showing significantly lower expression in the Qs compared to  
285 NBFs (1175 down regulated genes, lower expression in Qs versus 534 up regulated genes higher  
286 expression in Qs compared to NBFs; *Figure 3C, Supplementary Figure 7B, Additional file 25*). DEGs  
287 were globally enriched for biological processes and pathways that are related to regulation of nervous  
288 system development, response to lipid, meiotic nuclear division, gland development, negative regulation  
289 of developmental process, DNA methylation involved in gamete generation and other related terms  
290 (*Figure 3D, Supplementary Figure 7C,D,E, Additional file 26A, 27A, 28A*). Taking into account the  
291 direction of expression changes in Qs versus NBFs, we found that down-regulated genes were enriched  
292 for ontology terms related to gamete generation, central nervous system development, developmental  
293 process involved in reproduction, meiotic cell cycle process, DNA methylation involved in gamete  
294 generation and other related terms (*Supplementary Figure 8A,B,C,D, Additional file 26B, 27B, 28B*). In  
295 contrast, up-regulated genes are mainly enriched for inflammatory response, allograft rejection, leukocyte  
296 differentiation, positive regulation of cell differentiation, response to lipid, regulation of anatomical  
297 structure morphogenesis, regulation of hematopoiesis, cell adhesion and related pathways (*Supplementary*  
298 *Figure 9A,B,C,D, Additional file 26C, 27C, 28C*). These results likely reflect the different cellular  
299 composition of Q and NBF ovaries: NBF ovaries are immature and mainly contain primordial and  
300 primary follicles (*Supplementary Figure 10*) while Q ovaries develop abundant stromal cells to support  
301 oogenesis/folliculogenesis (*Supplementary Figure 10*).

302  
303 To investigate ovarian differences in more detail, we first compared our data with available  
304 transcriptomes of mouse oocytes at different stages of development [52]. We selected two clusters of  
305 genes: a first group that show a decline in expression from primary to small antral follicle stages [52] and  
306 the other group where expression increased at small and large antral [52] (753 genes in total in the two  
307 clusters). However, we did not observe differences in the expression levels of these genes between Q and  
308 NBF ovaries (*Supplementary Figure 11A*). This could be due to the fact that, in NMRs, only a few  
309 follicles progress to the large antral stage at each reproductive cycle, and thus the difference cannot be

310 quantified from whole ovary expression data. As NBF ovaries have been reported to be in an immature  
311 state [18, 19, 24], we next integrated available transcriptomes of whole ovaries from pre-puberal and adult  
312 mice [53]. Surprisingly, Q and NBF ovaries did not significantly differ in the expression levels of genes  
313 that are differentially expressed between pre-pubertal and adult mouse ovaries (*Supplementary Figure*  
314 *11B*). However, by focusing on a selection of oocyte-specific genes (*Zp1*, *Zp2*, *Zp3*, *Figla*, *Nobox*,  
315 *Sohlh1*, *Gja4*, *Pou5fl*, *Oosp1*, *Hlfoo*, *Lhx8*, *Sohlh2*, *Kit*, *Kitl*, *Nlrp14*, *Bmp15*), we observed a consistent  
316 increase in NBF compared to Q ovaries (*Supplementary Figure 11C*). This is in agreement with our  
317 histological assessment (*Supplementary Figure 10*): Q ovaries have more stromal, somatic cells than NBF  
318 ovaries, which are enriched in primary oocytes. Upon further examination of the NMR ovary RNA-seq  
319 data, in agreement with the enrichment of hormone-producing stromal cells in Q ovaries, we found that  
320 the main difference between ovaries of the Qs and NBFs was in the ability to synthesize estrogen. More  
321 specifically, among the top-10 list of DEGs, we found that *Cyp19A1* (*Aromatase*) - a key enzyme in the  
322 synthesis of estrogen- was highly over-expressed in Q compared to NBF ovaries (4-fold change in log2  
323 scale with almost no expression in NBF; *Supplementary Figure 11D*). Aromatase is important for  
324 transformation of androstenedione to estrone and testosterone to estradiol [54, 55]. In addition to  
325 *Cyp19A1*, other genes that are involved in steroid hormone biosynthesis (*Cyp7a1*, *Akr1c18*, *Akr1d1*,  
326 *Ugt1a1*) were also significantly more expressed in Q ovaries.

327

328 As a whole, our RNA-seq analysis suggests that Q ovaries specifically express genes that are required for  
329 estrogen production, and thereby undergoes oogenesis/folliculogenesis, ovulation and other sexual  
330 differentiation processes. In contrast, NBFs seem to have functioning ovaries that are arrested at a pre-  
331 pubertal stage, and do not have the ability to produce estrogen, explaining their failure to ovulate and  
332 subsequent reproductive incompetency.

333

334

335 **Non-breeding male NMRs show signs of defective post meiosis spermatogenesis and sperm  
336 maturation**

337

338 We next wanted to investigate the nature of the gonadal differences between breeding versus non-  
339 breeding male NMRs. Mammalian spermatogenesis begins post-natally at puberty, and continues  
340 throughout life. A pool of spermatogonial stem cells undergo mitotic divisions to produce spermatocytes,  
341 that will undergo meiotic divisions to produce haploid spermatids, which will then undergo further  
342 maturation to become mature spermatozoa [56-58]. This process of spermatogenesis occurs in  
343 seminiferous tubules supported by Sertoli cells and Leydig cells, and coordinated by endocrine signals.  
344 Sertoli cells nourish and provide structural support for the developing sperm cells, upon activation by  
345 follicle-stimulating hormone (FSH) that is secreted by the pituitary gland, which is in turn under the  
346 control of the GnRH secretion from the hypothalamus [58, 59]. Leydig cells that are located outside the  
347 seminiferous tubules produce and secrete testosterone and other androgens that are important for  
348 spermatogenesis, secondary sexual characteristics, development, and testis volume. Leydig cell function  
349 is modulated by the pituitary gonadotrophin luteinizing hormone (LH) [58, 59].

350

351 All male NMRs, independently of their breeding status, have been reported to undergo spermatogenesis  
352 [19, 26, 27]; however, sperm numbers and motility vary between BMs and NBMs [26, 27]. In order to  
353 investigate the molecular pathways that contribute to these differences in spermatogenesis between BMs  
354 and NBMs, we analyzed the gene expression profile of whole testes from BMs and NBMs, using the two  
355 criteria described above in the brain analysis section to define their status (long-term observational  
356 information and testis to body mass ratio). Based on PCA and hierachal clustering, unlike the clear  
357 separation observed for the ovaries of Qs and NBFs, NBM testes do not show a distinct clustering and  
358 grouping based on breeding status, with BM3 clustering with non-breeding animals (*Figure 4A,B*,  
359 *Supplementary Figure 12A*). Nevertheless, differential expression analysis identified 780 genes that  
360 showed significant changes between BMs and NBMs (522 up-regulated and 258 down-regulated genes in

361 BMs compared to NBMs; *Figure 4C, Supplementary Figure 12B, Additional file 29*). Enrichment  
362 analysis for biological processes, molecular functions, and pathways did not highlight obvious enrichment  
363 of spermatogenesis genes, but rather for terms such as: regulation of anatomical structure morphogenesis,  
364 response to steroid hormone, and embryo development (*Figure 4F, Supplementary Figure 12C,D,E*  
365 *Additional file 30A, 31A, 32A*). By focusing on DEGs that show lower expression in breeding compared  
366 to non-breeding testes, we found enrichment for terms that are related to positive regulation of cell-  
367 substrate adhesion, positive regulation of cell-substrate adhesion, processes regulation of histone  
368 acetylation (*Supplementary Figure 13A,B,C,D, Additional file 30B, 31B, 32B*). However, interestingly,  
369 for DEGs that are up-regulated in the testes of BMs we found enrichment of terms related to regulation of  
370 anatomical structure morphogenesis, response to steroid hormone, and multicellular organism  
371 reproduction (*Supplementary Figure 14A,B,C,D, Additional file 30C, 31C, 32C*). In particular, this last  
372 ontology category (GO:0032504) contains a number of genes (45 genes, *Additional file 30D*) involved in  
373 key biological processes related to male reproduction, such as meiotic cell cycle, male genitalia  
374 development, gamete generation, spermatogenesis, and sperm motility (*Lhcgr, Prm1, Prm2, Akap4, Odf3,*  
375 *Akap4, Wfdc2, Fdc2, Txnrd2, Txnrd3, Spata6, Spata22, Rara, Nr2c2, Hsf2bp, Cylc2, Cylc1, Celef3, Ccin,*  
376 *Alkbh5, Adam29, Atp2b4; Additional files 30D*). From this list of 45 genes, *Lhcgr* encodes the receptor for  
377 both LH and choriogonadotropin [60, 61]. Binding of LH to its receptor stimulates testosterone  
378 production by Leydig cells to promote extra-gonadal differentiation and maturation [62, 63]. *Prm1* and  
379 *Prm2* genes encode Protamine proteins that replace histones in later stages of spermatogenesis (sperm  
380 elongation) to allow denser packing of DNA [64-66]. In mouse models, abnormal expression of *Prm1* and  
381 *Prm2* causes a decrease in spermatozoa number, abnormal spermatozoa morphology and motility,  
382 damaged spermatozoa chromatin, and infertility [67-73]. *Odf3* is transcribed more specifically in  
383 spermatids, and is suggested to provide the elastic structure of sperm protecting from damage during  
384 epididymal transport [74, 75]. *Akap4* is transcribed only in the post-meiotic phase of spermatogenesis and  
385 is a cytoskeletal structure present in the principal piece of the sperm flagellum [76]. Targeted disruption  
386 of *Akap4* results in abnormal flagella, and hence motility of spermatozoa [76, 77].

387 To further document what stage of spermatogenesis might be interrupted in NBMs, we imported the list  
388 of spermatogenesis-related gene clusters that show specific expression at several different stages of  
389 mouse spermatogenesis (mitotic, meiotic, post-meiotic, and somatic clusters,) from the Germontline  
390 database (*Additional file 30*) [78]. DEGs belonging to all four clusters tended to show higher expression  
391 in BMs than in NBMs, but this difference was more pronounced for post-meiotic genes (*Figure 4D*).  
392 More specifically, 10% (55 out of the 522) of up-regulated DEGs in BM belonged to the post-meiotic  
393 cluster, while only 2% of down-regulated DEGs in BM (5 out of 258) belonged to this category. This  
394 suggests that the main difference in BM and NBM spermatogenesis is related to genes that are important  
395 in post-meiotic stages. In addition, we plotted the expression level of all genes that belong to these four  
396 clusters. Genes that belong to both the meiotic and post-meiotic stages of spermatogenesis showed  
397 significant up-regulation in BMs compared to NBMs (p-value 0.01 and 0.002 respectively, Wilcoxon  
398 rank sum test; *Figure 4E*), but this trend was more pronounced for the post-meiotic genes (*Figure 4E*).  
399 Finally, to understand the actual nature of the block in spermatogenesis in NBMs, we examined  
400 histological sections of BM and NBM testes. Confirming previous observations [19, 26, 27], all stages of  
401 spermatogenesis could be observed in both BM and NBM animals (*Supplementary Figure 15A, B*),  
402 although it was not possible to obtain a precise quantitative information by this method. Interestingly, a  
403 very clear difference was observed in interstitial (Leydig) cell content, with much higher numbers in the  
404 BM compared to NBM testes (*Supplementary Figure 15C,D*).

405

406 In conclusion, our detailed analysis of NMR testes revealed that although both breeding and non-breeding  
407 males undergo spermatogenesis, non-breeders show signs of impaired post-meiotic sperm maturation at  
408 the transcriptomic level. This may lead to lower sperm numbers and impaired motility that could underlie  
409 their incapacity to breed.

410

411

412

413 **Discussion**

414

415 In this study, we set out to define the molecular pathways that might underlie one of the principal features  
416 of eusociality, namely the extreme socially-induced suppression of reproduction (reproductive skew) in  
417 NMRs. We achieved this by comparison of brain and gonad transcriptomes in the different individuals  
418 that make up the NMR social hierarchy, namely Qs (the only breeding females), NBFs, BMs and NBMs.  
419 The comparative analysis of gene expression in the brain of different NMR social classes revealed that: 1)  
420 the Q has a distinctive gene expression profile when compared to the rest of the colony; 2) both sexes of  
421 non-breeding animals have nearly identical gene expression profiles in the brain; 3) several genes show  
422 significant gene expression differences in comparisons between breeding animals (Qs vs BMs) and  
423 between breeding animals versus non-breeding animals; 4) the Qs and NBFs have clear gene expression  
424 differences in the brain, but this is not reflected in comparisons between BMs and NBMs; 5) the  
425 dopaminergic pathway is identified as a major pathway that may be linked to the differences between  
426 social classes of NMR colonies. More specifically, differences in TH expression (the rate limiting enzyme  
427 for the synthesis of dopamine) were found in the hypothalamic areas of the brain which control prolactin  
428 secretion [48]. Also, the catecholaminergic innervation (potentially dopaminergic) in the hippocampus  
429 was strikingly high in breeders and showed a clear difference between breeders and non-breeders. In mice  
430 or rats, there are sparse catecholaminergic afferents, which arise mainly from the ventral tegmental area  
431 [79] and the locus coeruleus [80]. In NMR breeders, the origin of these afferents and their role in the  
432 different behaviors between castes remains to be investigated. We believe that this potential role of  
433 dopamine in NMR eusociality is a significant discovery that is consistent with findings in other eusocial  
434 animals. Indeed, the dopaminergic pathway is a highly conserved system across vertebrates and some  
435 invertebrates, with many important functions in the nervous system [81-86]. These include a crucial role  
436 in the control of movement and reinforcement learning. Dopamine is also implicated in sexual arousal,  
437 aggression, dominance, sleep, attention, working memory, hormonal regulation, and other functions [82,  
438 87-90]. In eusocial insects and termites, the importance of dopamine in aggression, dominance and social

439 hierarchy has been documented. Brain dopamine levels are elevated in dominant individuals of *Polistes*  
440 paper wasps (*Polistes chinensis*), the worker-totipotent ant (*Harpegnathos saltator*), and the queenless ant  
441 (*Diacamma* sp. from Japan); and a positive correlation was observed between ovarian activity and the  
442 level of dopamine in wasps (*Polistes chinensis*), bumble bees (*Bombus terrestris*), honeybees (*Apis*  
443 *mellifera* L.), ants (*Harpegnathos saltator*) [91-96]. In the honeybee (*Apis mellifera*), tyrosine or royal  
444 jelly-fed workers had higher brain dopamine levels than sucrose fed individuals, resulting in ovarian  
445 development and the inhibition of foraging in the former [97]. Dietary and topical applications of  
446 dopamine in queenless ant subordinate workers have also been shown to induce oocyte growth/activation  
447 [93, 96]. In honey bee (*Apis mellifera*), queen pheromone have been shown to modulate dopamine  
448 signaling pathways in the worker bees [98]. Furthermore, in male honeybees, juvenile hormone (JH),  
449 which regulate development and reproduction, has been shown to increase the levels of dopamine in the  
450 brain [99, 100]. JH and brain dopamine levels have been shown to increase during sexual maturation, and  
451 topical application of a JH analog increases the level of dopamine in the brain of male honeybees [101,  
452 102]. These investigations indicate that the dopaminergic pathway can be crucially involved in the social  
453 system, behavior, and reproductive suppression of many eusocial insects and termites. Our findings point  
454 to the exciting possibility that dopamine-mediated mechanisms are also involved in maintaining  
455 eusociality in the NMRs, suggesting that these pathways may represent a rather universal strategy for  
456 eusociality across the Animal Kingdom.

457 The potential functional significance of the observed differences in dopamine pathways between the Q  
458 and the non-breeders could be at several levels. One possible and well-established mechanism is the  
459 action of dopamine on prolactin regulation (reviewed in [48, 103]). In this pathway, dopamine (also  
460 known as prolactin inhibiting factor), modulates the release of prolactin from the pituitary gland, and  
461 prolactin itself acts on the hypothalamus to regulate the release of GnRH (*Supplementary Figure 16B,C*)  
462 [104]. The plasma prolactin data, measured for the first time in this species, fits our model of elevated  
463 PRL suppressing reproduction in non-breeders, because plasma PRL concentrations measured in spot  
464 samples collected from most non-breeding NMRs often greatly exceeded values which would be

465 considered clinical hyperprolactinemia in humans (*Figure 1F*). For example, levels of circulating PRL  
466 between 3 and 15 mg/l are considered necessary for maintaining normal reproductive function, and levels  
467 below and above are associated with an increased rate of infertility[105] - generally this equates to fasting  
468 concentrations above 25 ng/ml for women and 20 ng/ml for men [106]. Furthermore, plasma PRL in non-  
469 breeders often also exceeded that of lactating queens (*Figure 1F*). We did fail to detect PRL in some  
470 plasma samples from non-breeders. However, plasma PRL concentrations are known to exhibit circadian  
471 variations, and any such pattern or synchrony amongst NMRs remains to be investigated in detail to avoid  
472 sampling at times of low secretion. In mammals, hyperprolactinaemia is well known to be a major cause  
473 of infertility, and plays a role in this context in lactational suppression of reproduction [107]. A higher  
474 level of dopamine in the Q or BM is expected to decrease the release of prolactin. In turn, the absence or  
475 decreased levels of prolactin may result in the normal pulsatile secretion of GnRH from the  
476 hypothalamus, and subsequent release of LH and FSH from the pituitary gland. LH and FSH exert their  
477 effects at the level of the gonad, stimulating sexual differentiation, follicular development,  
478 spermatogenesis and release of the sex hormones estrogen (in females) and testosterone (in males)  
479 (*Supplementary Figure 16B*). The low dopamine in non-breeding animals on the other hand, results in  
480 elevated levels of prolactin, inhibiting the release of GnRH, and thus LH, FSH, estrogen, and testosterone,  
481 and follicular development and spermatogenesis (*Supplementary Figure 16C*). The inhibitory effect of  
482 dopamine on prolactin has been shown in many animals, and specific dopamine neurons (A12 and A14  
483 cell groups, *Supplementary Figure 16A*) are known to regulate the release of prolactin. Interestingly, the  
484 unique presence of cell bodies expressing RFRP-3 in the arcuate nucleus of NMRs, reported by [34] may  
485 be of significance, as it could potentially allow interaction with the A12 dopaminergic cell groups that  
486 regulate PRL section. Despite the differences in expression of genes involved in increasing dopamine in  
487 the brain, we did not observe differences in the GnRH expression between breeding and non-breeding  
488 NMR brains (both males and females). However, this would fit with predictions based on previous NMR  
489 studies that showed similar numbers of immunoreactive GnRH-1 cell bodies among breeders, non-  
490 breeders, males and females [33]. This implies that GnRH is produced by all status groups, but only

491 actually released in breeders, due to a block to its secretion in non-breeders. The neuropeptide kisspeptin  
492 may play a role in this process among females. Kisspeptin is well known to influence the hypothalamo–  
493 pituitary–gonadal axis by direct actions on GnRH-1 neurons [33]. Zhou *et al.* found that breeding females  
494 NMRs had increased numbers of kisspeptin immunoreactive cell bodies in the anterior periventricular  
495 nucleus and rostral periventricular region of the third ventricle (RP3V). This suggests a role for kisspeptin  
496 in the hypogonadotropic state in female NMRs, acting via mechanisms similar to those that underlie  
497 puberty and seasonal breeding in other species. Our observation of small increases in expression of *Kiss-1*  
498 and its receptor (*Kiss-1R*) (the latter being statistically significant) in Qs versus NBFs supports this  
499 hypothesis. Furthermore, elevated prolactin is known to have a suppressing effect on kisspeptin (and  
500 hence GnRH), which would fit a model of suppression involving dopamine pathways. Brown *et al.*  
501 (2014) showed that administration of prolactin to mice caused *Kiss-1* mRNA to be suppressed in the  
502 RP3V (the region with increased kisspeptin immunoreactive cell bodies in Q NMRs). It remains to be  
503 determined how the elevated levels of RFRP-3 (GnIH) in the brain of non-breeders, reported by [34], act  
504 in this mechanism that ultimately inhibits GnRH secretion. Interestingly, elevated prolactin has been  
505 implicated in many studies as a factor mediating both parental and alloparental care, affiliative and other  
506 sociosexual behaviours, in birds and mammals, including rodents and primates [108]. It is thus tempting  
507 to speculate that elevated prolactin in NMRs may play a central role in both cooperative behaviour and  
508 reproductive suppression.

509  
510 The hypothalamic block to reproduction in NMRs ultimately results in the well-documented lack of  
511 gonadal development in both male and female non-breeders [18, 19, 27]. Not surprisingly, given the large  
512 anatomical differences, gene expression profiling of the ovary identified several DEGs between the Qs  
513 and NBFs. One of the important differences in gene expression was the significantly higher expression of  
514 aromatase in Qs ovary (top10 DEGs), the key enzyme in the production of estrogen. This difference in  
515 aromatase expression indicates the ability of the Qs ovary to produce estrogen, and the low production of  
516 ovarian estrogen in NBFs. Aromatase knockout mice display underdeveloped external genitalia and uteri,

517 and precocious depletion of ovarian follicles, and anovulation, with development of the mammary glands  
518 approximately that of prepubertal WT female mice [109-111]. These features of aromatase knockout mice  
519 mimic most of the anatomical and physiological and endocrine differences that were observed between  
520 the Qs and NBFs of NMR [18, 19, 24]. The enrichment analysis on DEGs also highlights the difference  
521 between the Q and NBF ovaries. The small, pre-pubertal ovaries of NBFs mostly contain primordial and  
522 primary follicles that are arrested at the meiosis I prophase. In contrast, the fully functional ovary of the  
523 Qs contains follicles at all stages of development, including some polyovular follicles that may contain up  
524 to three oocytes [112], and larger numbers of stromal cells (*Supplementary Figure 10*). The ovarian  
525 stroma is a diverse mix of cell types that includes theca-interstitial cells, immune cells, blood vessels,  
526 smooth muscle cells, and several types of extracellular matrix proteins [113-118]. The increased stromal  
527 cell content in Qs is reflected in the enrichment of GO terms (such as regulation of leukocyte  
528 differentiation, regulation of hematopoiesis, cell adhesion) for DEGs unregulated in the Qs. The NBF  
529 ovaries, with mostly primary follicles containing oocytes arrested at meiosis I, is reflected in GO terms  
530 (such as DNA methylation involved in gamete generation, sexual reproduction, meiotic nuclear division,  
531 meiotic cell cycle) for DEGs upregulated in the NBFs. Taken together, the expression analyses reported  
532 here is consistent with the observation that the Qs ovary is able to undergo oogenesis/folliculogenesis,  
533 ovulation and other sexual differentiation processes; whereas NBF ovaries are arrested at a pre-pubertal  
534 stage, lack the ability to produce estrogen, ovulate and undergo sexual differentiation (*Supplementary*  
535 *Figure 16D*).

536

537 In the testis, our gene expression profiling of BM and NBMs revealed that there are no obvious  
538 differences, unlike what was observed between Q vs NBF ovaries. This could be partly explained by the  
539 fact that NBFs do not produce mature gametes but Q's do, whereas both BMs and NBMs NMRs undergo  
540 spermatogenesis to produce gametes, and this is reflected in spermatogenesis gene expression profiles  
541 seen in both. Nevertheless, some DEGs were identified that are known to play an important role in male  
542 sexual differentiation, reproduction, and responses to steroids. In particular, the enrichment of GO term

543 involved in male reproduction (meiotic cell cycle, male genitalia development, gamete generation,  
544 spermatogenesis, and sperm motility) for upregulated DEGs in BMs includes genes such as *Prm1*, *Prm2*,  
545 *Odf3* and *Akap4*, which have been shown to have a important role in mouse spermatogenesis. A decrease  
546 in expression of genes such as *Prm1*, *Prm2*, *Odf3* and *Akap4* and related genes in NBMs, might  
547 contribute for the observed reduction in sperm number and motility [28]. This is further complemented by  
548 the low expression of genes that are involved in meiotic and post-meiotic stages of spermatogenesis in  
549 non-breeding animals. Based on these observations, we suggest that the main difference between BMs  
550 and NBMs spermatogenesis is related to post-meiotic and sperm maturation stages (*Supplementary*  
551 *Figure 3E*), where NBMs fail to express critical genes at appropriate level at these stages (post-meiotic  
552 and sperm maturation stages), resulting in low sperm count and impaired motility. Cytological  
553 examination of testicular sections of breeding and non-breeding animals (*Supplementary Figure 15*),  
554 confirms previous observations of a higher interstitial cell content in breeding animals compared to non-  
555 breeding animals. At the gene expression level, we also observed a significant decrease (2 fold,  
556 *Additional file 29*) in the expression of the LH receptor gene (*Lhcgr*) in NBMs compared to BMs. This  
557 fits with predictions based on the reduction in interstitial (Leydig) cells numbers, together with previously  
558 reported observations of lower concentrations of urinary testosterone and plasma LH in NBMs [20].

559

## 560 Conclusion

561

562 Our study reveals gene expression differences among male and female NMR reproductive castes. Our  
563 findings in brain, ovaries and testes provide some of the first insights into the potential molecular  
564 mechanisms that are important in reproduction suppression of NMR. The gene expression differences in  
565 the NMR brains follow social status rather than sex, with the Q having a unique gene expression pattern  
566 compared to the rest of the colony. In highlighting the potential importance of dopamine pathways in the  
567 brain, and a possible role for hyperprolactinemia in mediating suppression, our findings significantly  
568 advance the understanding of the basis of mammalian eusociality and cooperative breeding systems. In

569 the ovaries, Qs express genes that are required for estrogen production, and thereby undergoes  
570 oogenesis/folliculogenesis, ovulation and other sexual differentiation processes. In contrast, NBFs are  
571 arrested at a pre-pubertal stage, and do not have the ability to produce ovarian estrogen, explaining their  
572 failure to ovulate and subsequent reproductive incompetency. In the testis, both BMs and NBM<sup>s</sup> undergo  
573 spermatogenesis, however NBM<sup>s</sup> fail to express genes required for post-meiotic sperm maturation, giving  
574 explanation to lower sperm numbers, impaired motility, and breeding incapacity of NBM<sup>s</sup>. While many  
575 questions remain, the phenotypic plasticity exhibited by the NMR also offers scope for understanding the  
576 dynamics of reproductive activation when non-breeders transition into breeding state (NBF to Qs or NBM  
577 to BM), in particular the role of epigenetics and other regulatory factors affecting the genes associated in  
578 this transformation.

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595 **Material and methods**

596

597 **Animals**

598 NMRs were maintained at Queen Mary University of London in compliance with institutional guidelines.

599 All animals were born in captivity, kept under constant ambient temperature of 28-30°C, and housed in

600 artificial burrow systems composed of interconnected perspex tubing with separate chambers for nesting,

601 food and toilet, simulating natural burrow conditions. They were fed an ad libitum diet of a variety of

602 chopped root vegetables such as sweet potato and turnip. Animal tissues were collected post mortem

603 (*Additional file 1A*) following euthanasia in full accordance with Institutional and National animal care

604 and use guidelines.

605

606 **Blood sampling for hormone assay**

607 Blood samples were obtained from *X NBF*, *Queen BM*, *NBM naked mole-rats* from among 13 captive

608 colonies from QMUL and UP. All blood samples collected at the University of Pretoria were with local

609 ethics committee clearance. Blood samples were collected between 11h00 and 15h00 as follows: The

610 animals were hand held and venous blood samples collected from the hind foot. Approximately 300-

611 500ul of blood was collected into heparinised micro-haematocrit tubes (University of Pretoria samples) or

612 into a heparinised 1 ml syringe from trunk blood following euthanasia (QMUL) prior to tissue collection.

613 The blood was centrifuged at 500g for 15 minutes and the plasma separated from the red cells and stored

614 at -80°C until hormone assay.

615

616 **RNA extraction and sequencing**

617 NMR brain (excluding the olfactory bulb, the cerebellum, the medulla and the pons), ovary, and testis

618 were snap frozen by immersion into liquid nitrogen and subsequently stored at -80°C prior to RNA

619 extraction. Snap frozen tissues were mechanically powdered and mixed to maintain the heterogeneity of

620 the sample. Total RNA was extracted from tissues using Qiagen miRNeasy kit (Qiagen, USA) following

621 the manufacturer's recommendations. The quality of the extracted RNA was controlled using the Agilent  
622 Bioanalyzer (Agilent) and Qubit Fluorometric Quantitation (Thermo Fisher Scientific Inc.). 1 $\mu$ g of high  
623 quality RNA (with RNA Integrity Number (RIN) >7) were used for RNA sequencing. Total RNA, after  
624 polyadenylated RNA purification, was prepared for sequencing using Illumina Truseq library preparation  
625 protocol. For each sample, around 100 (*Additional file 1B*) million raw paired-end sequence reads (101  
626 base pair long) were generated using Illumina HiSeq 2000 sequencing instrument. Data sets are available  
627 from NCBI GEO under accession number #####.

628

## 629 **RNA sequence (RNA-seq) analysis**

630 The quality of the generated RNA sequence was evaluated using FastQC (version 0.11.2) [119]. Sequence  
631 adapters, low quality reads, and overrepresented sequences (e.g. mitochondrial sequences) in the reads  
632 were removed using Trimmomatic (version 0.32) [120], and the quality of the reads was checked again  
633 using FastQC. Sequence that passed the quality assessment were aligned to the NMR genome (hetGla2,  
634 Broad institute, 2012) using tophat2 (version 2.0.6, default parameters) [121], with bowtie2 (version  
635 2.1.0) [122]. For each sample genome guided de-novo transcriptome assembly was performed using  
636 Cufflinks (version 2.2.1, default parameters) [123] and assembled transcript from all samples were  
637 merged using cuffmerge (cufflinks) to generate a master reference transcripts. Merged transcripts (master  
638 transcripts) were annotated to gene name using mouse transcripts. Mouse cDNA were obtained from  
639 Ensembl [124], and a mouse cDNA blast database was generated using blast (version 2.2.25) [125, 126].  
640 Assembled and merged NMR transcripts (master transcript) were searched against mouse cDNA blast  
641 database, and transcripts with e val <= 10<sup>-5</sup>, with a length of >200bp were retained. Transcript  
642 abundance level was generated using master reference transcripts generated by cufflinks and HTSeq  
643 (version 0.5.3p9) [127]. The transcript level quantified using HTSeq was used as an input for further  
644 processing using R software environment for statistical computing and graphics (version 3.2.2). Data  
645 normalization, removal of batch effect and other variant was performed using EDASeq R package  
646 (version 2.2.0) [128] and RUVseq package (Remove Unwanted Variation from RNA-Seq package) [129].

647 In short, read counts were normalized using EDASeq R package, and “in-silico empirical” negative  
648 controls genes, for RUVseq package (RUVg normalization) were obtained by taking least significantly  
649 DEGs based on a first-pass differential expression analysis performed prior to RUVg normalization.  
650 RUVseq package (RUVg normalization) was then performed using the “in-silico empirical” negative  
651 controls genes. Differential expression was performed using edgeR R package (version 3.10.5) [130],  
652 using the negative binomial GLM approach, edgeR normalization on raw counts, and by considering a  
653 design matrix that includes both the covariates of interest and the factors of unwanted variation. DEGs  
654 with false discovery rate (FDR <=0.05, Benjamini-Hochberg multiple testing correction), expression level  
655 > 1 CPM (counts per million), and log fold change >1 were retained and used for further processing, gene  
656 ontology and pathway analysis.

657

#### 658 **Gene ontology and pathway analysis**

659 Gene ontology and pathway analysis were performed using Metascape (metascape.org) [131], Enrichr  
660 [132], and Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, [www.qiagen.com/ingenuity](http://www.qiagen.com/ingenuity)). In  
661 Metascape, for each given gene list of DEGs, pathway and process enrichment analysis was carried out  
662 with the following ontology sources: GO Biological Processes, GO Molecular Functions and KEGG  
663 Pathway. All genes in the genome were used as the enrichment background. Terms with p-value < 0.01,  
664 minimum count 3, and enrichment factor > 1.5 (enrichment factor is the ratio between observed count and  
665 the count expected by chance) are collected and grouped into clusters based on their membership  
666 similarities [131]. P-values are calculated based on accumulative hypergeometric distribution, q-values  
667 are calculated using the Benjamini-Hochberg procedure to account for multiple testing [131]. Kappa  
668 scores were used as the similarity metric when performing hierarchical clustering on the enriched terms  
669 and then sub-trees with similarity > 0.3 are considered a cluster. The most statistically significant term  
670 within a cluster is chosen as the one representing the cluster [131]. To further capture the relationship  
671 among terms, a subset of enriched terms were selected and rendered as a network plot, where terms with  
672 similarity > 0.3 are connected by edges, with the best p-values from each of the 20 clusters [131]. Similar

673 independent enrichment analysis was performed using Enrichr [132], to validate the outcome of  
674 Metascape. The list of DEGs were used as input to Enrichr and the enrichment of GO Biological  
675 Processes, GO Molecular Functions was investigated. The final result was sorted and plotted by using a  
676 combined score. The combined score is a combination of the p-value and z-score calculated by  
677 multiplying the two scores as follows:  $c = \log(p) * z$ , where c is the combined score, p is the p-value  
678 computed using Fisher's exact test, and z is the z-score computed to assess the deviation from the  
679 expected rank [132].

680

#### 681 **Tissue preparation and immunofluorescence**

682 NMRs were deeply anesthetized with 80 mg pentobarbital (400  $\mu$ L Euthatal, IP) and transcardially  
683 perfused with 40 g/L formaldehyde in PBS for 10 min (10 mL/min). Brains were dissected and kept at  
684 4°C for 12h in a PBS solution containing 15% sucrose before being flash frozen in isopentane (1 min at -  
685 30°C) and stored at -80°C. Serial coronal sections (30  $\mu$ m thick) were made with a cryostat (Leica,  
686 France) and mounted onto Superfrost Plus slides. For each animal, the first section containing some  
687 frontal cortex was taken as the origin of the numbering. Sections were stored dry at -80°C until being  
688 processed for immunolabelling.

689 Sections were equilibrated to -20°C before a short additional fixation with 30 g/L formaldehyde in PBS  
690 for 5 min at RT. Formaldehyde was then neutralized with TBS (50 mM Tris, 150 mM NaCl, pH 7.4) for 5  
691 min at 4°C, before two consecutive steps of permeabilization of 5 min each at 4°C, first in PBS  
692 containing 0.5% (vol/vol) Triton X-100, then in PBS with 10% (vol/vol) methanol. The sections were  
693 rinsed with 70% (vol/vol) ethanol (EtOH70) and kept in the same solution for 10 min at RT before being  
694 treated with 0.1% (w/vol) Sudan Black B in EtOH70, for autofluorescence removal. After two quick  
695 rinses with EtOH70 and one PBS wash, 30 min preblocking at 4°C was achieved with the IF buffer ie  
696 PBS, 2% (w/vol) BSA, 0.2% (vol/vol) Tween 20, 50 mM glycine. The anti-TH rabbit polyclonal antibody  
697 (Abcam, ab112) diluted 1/800 in IF buffer was incubated overnight at 4°C. Then the sections were rinsed  
698 3 times 10 minutes with PBS before being incubated with an Alexa488-coupled goat anti-rabbit

699 secondary antibody (Invitrogen A-11034) diluted 1/500 in PBS for 1h at RT. After 3 rinses in PBS and a  
700 30 min DAPI staining step (100 nM in PBS), sections were finally rinsed and mounted in Vectashield  
701 (Vector Laboratories, USA). Image acquisition was carried out at the Cell and Tissue Imaging Platform of  
702 the Genetics and Developmental Biology Department (UMR3215/U934) of Institut Curie. The sections of  
703 the different animals were processed at the same time with the same parameters. Full views of the  
704 sections consisted of mosaics of pictures made with the 5X objective of an upright epifluorescence  
705 microscope (Zeiss). Views of the hippocampus were made with the 10X objective. Additional z-stack  
706 pictures (1  $\mu$ m steps) of the regions of interest were taken for quantification, with the 10X objective of an  
707 upright spinning disk confocal microscope (Roper/Zeiss). For each picture, 5 consecutive confocal optical  
708 sections were summed, then the average intensity of the region of interest was measured, and finally the  
709 measures for both hemispheres were averaged (when applicable).

710

### 711 **Histology of testes and ovaries**

712 Gonadal tissue samples were fixed by immersion in 4% paraformaldehyde in saline within 15 minutes of  
713 collection for a minimum of seven days. After fixation tissue samples were dehydrated and cleared using  
714 standard histological methodology, before embedding in paraffin wax. Sections of 5-8  $\mu$ m were cut and  
715 stained for light microscopy with haematoxylin-eosin. Photomicrographs of sections were captured with a  
716 QIClick<sup>TM</sup> CCD Camera (01-QICLICK-R-F-CLR-12; QImaging) linked to a DMRA2 light microscope  
717 (Leica), using Volocity<sup>®</sup> v.6.3.1 image analysis software (Perkin-Elmer) running on an iMac computer  
718 (27-inch with OS X Yosemite, version 10.10).

719

### 720 **Prolactin assay and validation**

721 Plasma prolactin concentrations were determined using a commercial enzyme-linked immunosorbent  
722 assay (Elabscience<sup>©</sup> Guinea pig prolactin ELISA kit, Catalogue No: E-EL-GP0358) according to the  
723 instructions in the manufacturer's user manual. In brief, 100  $\mu$ l of reference standard and diluted samples  
724 (1/2 to 1/50 in sample diluent) were transferred into coated wells of a 96-well micro-ELISA plate,

725 respectively, and incubated for 90 min at 37°C. Subsequently, all supernatant was removed, and the plate  
726 patted dry. Immediately, 100 ml of biotinylated detection antibody was added, and incubated for 60 min  
727 at 37°C. The plate was washed 3 times, patted dry, and 100 ml of horse radish peroxide (HRP) conjugate  
728 were added and incubated for 30 min at 37°C. Subsequently, the plate was washed 5 times with wash-  
729 buffer, and patted dry. 90 ml of substrate reagent were added, and incubated for 15 min at 37°C. To  
730 terminate the enzymatic reaction, 50 ml of stop solution were added. Optical density was determined at  
731 450 nm and results calculated using a best-fit curve. The sensitivity of the assay was 0.1 ng/ml, the  
732 detection range 0.16-20 ng/ml, and coefficient of variation for repeatability was < 10%.

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751 **Acknowledgments**

752 The authors would like to acknowledge the Cell and Tissue Imaging Platform of the Genetics and  
753 Developmental Biology Department (UMR3215/U934) of *Institut Curie*, member of France-Bioimaging  
754 (supported by ANR-10-INSB-04), particularly Olivier Renaud for help with light microscopy. We thank  
755 all members of the Edith Heard team for stimulating discussions, technical and conceptual advice, notably  
756 Mikael Attia and Ronan Chaligné. We thank Deborah Bourc'his, Nicolas Servant, and Emmanuel Barillot  
757 from Institut Curie for their insightful comments on the manuscript. We would like to acknowledge the  
758 Biological Services Unit staff at QMUL (Richard Rountree, Arthur Gatward, Paul Sroka) for care and  
759 maintenance of animals and Cécile Reyes (Institut Curie, Genomics facility) for tissue fragmentation and  
760 RNA extraction, and Professor Jenny Jarvis (University of Cape Town). We thank the Institut Curie  
761 High-throughput sequencing platform (NGS platform) for generating all the sequencing data.

762

763 **Funding**

764 This work was supported by the International Blaise Pascal Chair to EHB, financed by the Région Ile-de-  
765 France and administered by the Fondation de l'Ecole normale supérieure. The Cell and Tissue Imaging  
766 Platform of the Genetics and Developmental Biology Department was supported by ANR-10-INSB-04.  
767 The Institut Curie High-throughput sequencing platform (NGS platform) was supported by ANR-10-  
768 EQPX-03 and ANR10-INBS-09-08 from the Agence Nationale de la Recherche and by the Canceropôle  
769 (Ile-de-France). L.M-P was funded by an FRM fellowship (SPF20151234950). Funding to EH was from  
770 Labex DEEP (ANR-11-LBX-0044) part of the IDEX Idex PSL (ANR-10-IDEX-0001-02 PSL) and  
771 ABS4NGS (ANR-11-BINF-0001).

772

773 **Availability of data and materials**

774 The datasets generated and/or analysed during the current study are available in the NCBI GEO  
775 repository (<https://www.ncbi.nlm.nih.gov/geo/>), under series number #####

776

777 **Author contributions**

778 EM conceived the study, designed experimental strategies, performed all bioinformatic analyses,  
779 interpreted results, performed experiments and wrote the manuscript. LMP performed and interpreted  
780 brain validation experiments, and participated in writing the manuscript. DG performed RNA extraction.  
781 NCB collected plasma samples and funded hormonal work, while SBG and AG carried out hormone  
782 assays. EHB designed experimental strategies, interpreted results, funded the project. CGF conceived the  
783 study, designed experimental strategies, interpreted results, performed experiments and wrote the  
784 manuscript. EH conceived the study, designed experimental strategies, interpreted results, and wrote the  
785 manuscript.

786

787 **Competing interests**

788 The authors declare no competing interests.

789

790 **Ethical approval**

791 Not applicable

792

793 **Consent for publication**

794 Not applicable

795

796

797

798

799

800

801

802

803 **List of abbreviations**

804	NMR	Naked mole-rats
805	LH	Luteinizing hormone
806	GnRH	Gonadotrophin releasing hormone
807	RNA-seq	RNA-sequencing
808	PCA	Principal component analysis
809	Q	Queen
810	Qs_Tr	Queen technical replicate
811	NBF	Non-breeding females
812	NBF_Tr	Non breeding female technical replicate
813	NBM	Non-breeding males
814	BM	Breeding males
815	CPM	Counts per million
816	DEG	Differentially expressed genes
817	GO	Gene ontology
818	Q genes	Queen genes
819	TH/Th	Tyrosine hydroxylase
820	FSH	Follicle-stimulating hormone
821	PGCs	Primordial germ cells
822	PRL	Prolactin
823	bp	Base pair
824	IF	Immunofluorescence
825	PValue	p-value
826	logCPM	Counts per million in log scale
827	logFC	Log fold change
828	FDR	False discovery rate
829	BH	Benjamini-Hochberg
830		
831		
832		
833		
834		

835 **References**

- 836 1. Szathmary E, Smith JM: The major evolutionary transitions. *Nature* 1995, 374:227-232.
- 837 2. Nowak MA, Tarnita CE, Wilson EO: The evolution of eusociality. *Nature* 2010, 466:1057-1062.
- 838 3. Wilson EO: *The insect societies*. Cambridge, Mass.: Belknap Press of Harvard University Press; 1971.
- 839 4. Michener CD: Comparative Social Behavior of Bees. *Annual Review of Entomology* 1969, 14:299-342
- 840 5. Crespi BJ, Yanega D: The definition of eusociality. *Oxford Journals, Behavioral Ecology* 1995, 6:109-115.
- 841 6. Duffy JE, Morrison CL, Rios R: Multiple origins of eusociality among sponge-dwelling shrimps (Synalpheus). *Evolution* 2000, 54:503-516.
- 842 7. Duffy JE, Macdonald KS: Kin structure, ecology and the evolution of social organization in shrimp: a comparative analysis. *Proc Biol Sci* 2010, 277:575-584.
- 843 8. Jarvis JU: Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science* 1981, 212:571-573.
- 844 9. Friedman DA, Gordon DM: Ant Genetics: Reproductive Physiology, Worker Morphology, and Behavior. *Annu Rev Neurosci* 2016, 39:41-56.
- 845 10. J FE, H K: On Biomass and Trophic Structure of the Central Amazonian Rain Forest Ecosystem *Biotropica* 1973, 5:2-14.
- 846 11. Hölldobler B, Wilson EO: *The ants*. Cambridge, Mass.: Belknap Press of Harvard University Press; 1990.
- 847 12. M OR, CG F: African mole-rats: eusociality, relatedness and ecological constraints. In *Ecology of social evolution*. Edited by J H, J K. Berlin, Germany: Springer; 2008: 205-220
- 848 13. Faulkes CG, Bennett NC, Bruford MW, O'Brien HP, Aguilar GH, Jarvis JU: Ecological constraints drive social evolution in the African mole-rats. *Proc Biol Sci* 1997, 264:1619-1627.
- 849 14. M. JJJ, C. BN: Eusociality has evolved independently in two genera of bathyergid mole-rats — but occurs in no other subterranean mammal. *Behavioral Ecology and Sociobiology* 1993, 33:253-260.
- 850 15. RA. B: The ecology of naked mole-rat colonies: burrowing, food, and limiting factors. In *The Biology of the Naked Mole-Rat*. Edited by PW S, JUM J, RD A. New Jersey: Princeton University Press; 1991: 97-136
- 851 16. EA. L, PW. S: Social organization of naked mole-rat colonies: Evidence for divisions of labor. In *The Biology of the Naked Mole-Rat*. Edited by PW S, JUM J, RD A. New Jersey: Princeton University Press; 1991: 275-336
- 852 17. Buffenstein R: The naked mole-rat: a new long-living model for human aging research. *J Gerontol A Biol Sci Med Sci* 2005, 60:1369-1377.
- 853 18. Faulkes CG, Abbott DH, Jarvis JU: Social suppression of ovarian cyclicity in captive and wild colonies of naked mole-rats, *Heterocephalus glaber*. *J Reprod Fertil* 1990, 88:559-568.
- 854 19. Faulkes CG, Abbott DH, Jarvis JU, Sherriff FE: LH responses of female naked mole-rats, *Heterocephalus glaber*, to single and multiple doses of exogenous GnRH. *J Reprod Fertil* 1990, 89:317-323.
- 855 20. Faulkes CG, Abbott DH: Social control of reproduction in breeding and non-breeding male naked mole-rats (*Heterocephalus glaber*). *J Reprod Fertil* 1991, 93:427-435.
- 856 21. Faulkes CG, Abbott DH: Evidence that primer pheromones do not cause social suppression of reproduction in male and female naked mole-rats (*Heterocephalus glaber*). *J Reprod Fertil* 1993, 99:225-230.
- 857 22. Clarke FM, Faulkes CG: Dominance and queen succession in captive colonies of the eusocial naked mole-rat, *Heterocephalus glaber*. *Proc Biol Sci* 1997, 264:993-1000.
- 858 23. Clarke FM, Faulkes CG: Hormonal and behavioural correlates of male dominance and reproductive status in captive colonies of the naked mole-rat, *Heterocephalus glaber*. *Proc Biol Sci* 1998, 265:1391-1399.
- 859 24. Margulis SW, Saltzman W, Abbott DH: Behavioral and hormonal changes in female naked mole-rats (*Heterocephalus glaber*) following removal of the breeding female from a colony. *Horm Behav* 1995, 29:227-247.
- 860 25. O'Riain MJ, Jarvis JU, Alexander R, Buffenstein R, Peeters C: Morphological castes in a vertebrate. *Proc Natl Acad Sci U S A* 2000, 97:13194-13197.
- 861 26. Dengler-Crish CM, Catania KC: Phenotypic plasticity in female naked mole-rats after removal from reproductive suppression. *J Exp Biol* 2007, 210:4351-4358.
- 862 27. Faulkes CG, Abbott DH, Jarvis JU: Social suppression of reproduction in male naked mole-rats, *Heterocephalus glaber*. *J Reprod Fertil* 1991, 91:593-604.
- 863 28. Faulkes CG, Trowell SN, Jarvis JU, Bennett NC: Investigation of numbers and motility of spermatozoa in reproductively active and socially suppressed males of two eusocial African mole-rats, the naked mole-rat

889 (Heterocephalus glaber) and the Damaraland mole-rat (Cryptomys damarensis). *J Reprod Fertil* 1994,  
890 100:411-416.

891 29. Legan SJ, Karsch FJ: Neuroendocrine regulation of the estrous cycle and seasonal breeding in the ewe. *Biol  
892 Reprod* 1979, 20:74-85.

893 30. Glasier A, McNeilly AS, Baird DT: Induction of ovarian activity by pulsatile infusion of LHRH in women  
894 with lactational amenorrhoea. *Clin Endocrinol (Oxf)* 1986, 74:243-252.

895 31. Abbott DH, Hodges JK, George LM: Social status controls LH secretion and ovulation in female marmoset  
896 monkeys (*Callithrix jacchus*). *J Endocrinol* 1988, 117:329-339.

897 32. Goodman RL, Lehman MN: Kisspeptin neurons from mice to men: similarities and differences.  
898 *Endocrinology* 2012, 153:5105-5118.

899 33. Zhou S, Holmes MM, Forger NG, Goldman BD, Lovern MB, Caraty A, Kallo I, Faulkes CG, Coen CW:  
900 Socially regulated reproductive development: analysis of GnRH-1 and kisspeptin neuronal systems in  
901 cooperatively breeding naked mole-rats (*Heterocephalus glaber*). *J Comp Neurol* 2013, 521:3003-3029.

902 34. Peragine DE, Pokarowski M, Mendoza-Viveros L, Swift-Gallant A, Cheng HM, Bentley GE, Holmes MM:  
903 RFamide-related peptide-3 (RFRP-3) suppresses sexual maturation in a eusocial mammal. *Proc Natl Acad  
904 Sci U S A* 2017, 114:1207-1212.

905 35. Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE: Brain transcriptomic  
906 analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Biol  
907 Sci* 2010, 277:2139-2148.

908 36. Ferreira PG, Patalano S, Chauhan R, Ffrench-Constant R, Gabaldon T, Guigo R, Sumner S: Transcriptome  
909 analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of  
910 alternative phenotypes. *Genome Biol* 2013, 14:R20.

911 37. Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, Toth AL: Genome, transcriptome and  
912 methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in  
913 a social insect. *Mol Ecol* 2016, 25:1769-1784.

914 38. Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkampf M, Roux J, Nygaard S, Glastad KM,  
915 Hagen DE, Viljakainen L, et al: Social insect genomes exhibit dramatic evolution in gene composition and  
916 regulation while preserving regulatory features linked to sociality. *Genome Res* 2013, 23:1235-1247.

917 39. Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat A,  
918 Fischman BJ, et al: Social evolution. Genomic signatures of evolutionary transitions from solitary to group  
919 living. *Science* 2015, 348:1139-1143.

920 40. Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijen CJ,  
921 Hasselmann M, Lozier JD, et al: The genomes of two key bumblebee species with primitive eusocial  
922 organization. *Genome Biol* 2015, 16:76.

923 41. Smith CR, Toth AL, Suarez AV, Robinson GE: Genetic and genomic analyses of the division of labour in  
924 insect societies. *Nat Rev Genet* 2008, 9:735-748.

925 42. Hunt GJ, Gadau JR: Editorial: Advances in Genomics and Epigenomics of Social Insects. *Front Genet*  
926 2016, 7:199.

927 43. Yu C, Li Y, Holmes A, Szafranski K, Faulkes CG, Coen CW, Buffenstein R, Platzer M, de Magalhaes JP,  
928 Church GM: RNA sequencing reveals differential expression of mitochondrial and oxidation reduction  
929 genes in the long-lived naked mole-rat when compared to mice. *PLoS One* 2011, 6:e26729.

930 44. Davies KT, Bennett NC, Tsagkogeorga G, Rossiter SJ, Faulkes CG: Family Wide Molecular Adaptations  
931 to Underground Life in African Mole-Rats Revealed by Phylogenomic Analysis. *Mol Biol Evol* 2015,  
932 32:3089-3107.

933 45. Fang X, Seim I, Huang Z, Gerashchenko MV, Xiong Z, Turanov AA, Zhu Y, Lobanov AV, Fan D, Yim  
934 SH, et al: Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat  
935 genomes. *Cell Rep* 2014, 8:1354-1364.

936 46. Holmes MM, Goldman BD, Goldman SL, Seney ML, Forger NG: Neuroendocrinology and sexual  
937 differentiation in eusocial mammals. *Front Neuroendocrinol* 2009, 30:519-533.

938 47. Bjorklund A, Dunnett SB: Dopamine neuron systems in the brain: an update. *Trends Neurosci* 2007,  
939 30:194-202.

940 48. Ben-Jonathan N, Hnasko R: Dopamine as a prolactin (PRL) inhibitor. *Endocr Rev* 2001, 22:724-763.

941 49. Jung D, Kee K: Insights into female germ cell biology: from in vivo development to in vitro derivations.  
942 *Asian J Androl* 2015, 17:415-420.

943 50. Bukovsky A, Caudle MR, Svetlikova M, Wimalasena J, Ayala ME, Dominguez R: Oogenesis in adult  
944 mammals, including humans: a review. *Endocrine* 2005, 26:301-316.

945 51. Reiss M, Sants H: *Behaviour and Social Organisation*. 1987.

946 52. Pan H, O'Brien M J, Wigglesworth K, Eppig JJ, Schultz RM: Transcript profiling during mouse oocyte  
947 development and the effect of gonadotropin priming and development in vitro. *Dev Biol* 2005, 286:493-  
948 506.

949 53. Pan L, Gong W, Zhou Y, Li X, Yu J, Hu S: A comprehensive transcriptomic analysis of infant and adult  
950 mouse ovary. *Genomics Proteomics Bioinformatics* 2014, 12:239-248.

951 54. Simpson ER, Zhao Y, Agarwal VR, Michael MD, Bulun SE, Hinshelwood MM, Graham-Lorenz S, Sun  
952 T, Fisher CR, Qin K, Mendelson CR: Aromatase expression in health and disease. *Recent Prog Horm Res*  
953 1997, 52:185-213; discussion 213-184.

954 55. Simpson ER, Mahendroo MS, Means GD, Kilgore MW, Hinshelwood MM, Graham-Lorenz S, Amarneh  
955 B, Ito Y, Fisher CR, Michael MD, et al.: Aromatase cytochrome P450, the enzyme responsible for estrogen  
956 biosynthesis. *Endocr Rev* 1994, 15:342-355.

957 56. Griswold MD: Spermatogenesis: The Commitment to Meiosis. *Physiol Rev* 2016, 96:1-17.

958 57. Griswold MD, Oatley JM: Concise review: Defining characteristics of mammalian spermatogenic stem  
959 cells. *Stem Cells* 2013, 31:8-11.

960 58. J W, JB S, CM L: Mammalian spermatogenesis. *Func Dev Embryol* 2007, 1:99-117.

961 59. Ramaswamy S, Weinbauer GF: Endocrine control of spermatogenesis: Role of FSH and LH/ testosterone. *Spermatogenesis* 2014, 4:e996025.

962 60. Themmen APN, Huhtaniemi IT: Mutations of gonadotropins and gonadotropin receptors: elucidating the  
963 physiology and pathophysiology of pituitary-gonadal function. *Endocr Rev* 2000, 21:551-583.

964 61. Gospodarowicz D: Properties of the luteinizing hormone receptor of isolated bovine corpus luteum plasma  
965 membranes. *J Biol Chem* 1973, 248:5042-5049.

966 62. Rivero-Muller A, Potorac I, Pintiaux A, Daly AF, Thiry A, Rydlewski C, Nisolle M, Parent AS,  
967 Huhtaniemi I, Beckers A: A novel inactivating mutation of the LH/chorionic gonadotrophin receptor with  
968 impaired membrane trafficking leading to Leydig cell hypoplasia type 1. *Eur J Endocrinol* 2015, 172:K27-  
969 36.

970 63. Ascoli M, Fanelli F, Segaloff DL: The lutropin/choriogonadotropin receptor, a 2002 perspective. *Endocr  
971 Rev* 2002, 23:141-174.

972 64. Sassone-Corsi P: Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science*  
973 2002, 296:2176-2178.

974 65. Steger K: Transcriptional and translational regulation of gene expression in haploid spermatids. *Anat  
975 Embryol (Berl)* 1999, 199:471-487.

976 66. Balhorn R: The protamine family of sperm nuclear proteins. *Genome Biol* 2007, 8:227.

977 67. Steger K, Failing K, Klonisch T, Behre HM, Manning M, Weidner W, Hertle L, Bergmann M, Kliesch S:  
978 Round spermatids from infertile men exhibit decreased protamine-1 and -2 mRNA. *Hum Reprod* 2001,  
979 16:709-716.

980 68. Bjorndahl L, Kvist U: Human sperm chromatin stabilization: a proposed model including zinc bridges. *Mol  
981 Hum Reprod* 2010, 16:23-29.

982 69. Carrell DT, Emery BR, Hammoud S: Altered protamine expression and diminished spermatogenesis: what  
983 is the link? *Hum Reprod Update* 2007, 13:313-327.

984 70. Luke L, Vicens A, Tourmente M, Roldan ER: Evolution of protamine genes and changes in sperm head  
985 phenotype in rodents. *Biol Reprod* 2014, 90:67.

986 71. Imken L, Rouba H, El Houate B, Louanjli N, Barakat A, Chafik A, McElreavey K: Mutations in the  
987 protamine locus: association with spermatogenic failure? *Mol Hum Reprod* 2009, 15:733-738.

988 72. Ravel C, Chantot-Bastaraud S, El Houate B, Berthaut I, Verstraete L, De Larouziere V, Lourenco D,  
989 Dumaine A, Antoine JM, Mandelbaum J, et al: Mutations in the protamine 1 gene associated with male  
990 infertility. *Mol Hum Reprod* 2007, 13:461-464.

991 73. Clark AG, Civetta A: Evolutionary biology. Protamine wars. *Nature* 2000, 403:261, 263.

992 74. Petersen C, Aumuller G, Bahrami M, Hoyer-Fender S: Molecular cloning of Odf3 encoding a novel coiled-  
993 coil protein of sperm tail outer dense fibers. *Mol Reprod Dev* 2002, 61:102-112.

994 75. Baltz JM, Williams PO, Cone RA: Dense fibers protect mammalian sperm against damage. *Biol Reprod*  
995 1990, 43:485-491.

996 76. Miki K, Willis WD, Brown PR, Goulding EH, Fulcher KD, Eddy EM: Targeted disruption of the Akap4  
997 gene causes defects in sperm flagellum and motility. *Dev Biol* 2002, 248:331-342.

998 77. Moretti E, Scapigliati G, Pasarelli NA, Baccetti B, Collodel G: Localization of AKAP4 and tubulin  
999 proteins in sperm with reduced motility. *Asian J Androl* 2007, 9:641-649.

1000

1001 78. Lardenois A, Gattiker A, Collin O, Chalmel F, Primig M: GermOnline 4.0 is a genomics gateway for  
1002 germline development, meiosis and the mitotic cell cycle. *Database (Oxford)* 2010, 2010:baq030.  
1003 79. Gasbarri A, Sulli A, Packard MG: The dopaminergic mesencephalic projections to the hippocampal  
1004 formation in the rat. *Prog Neuropsychopharmacol Biol Psychiatry* 1997, 21:1-22.  
1005 80. Kempadoo KA, Mosharov EV, Choi SJ, Sulzer D, Kandel ER: Dopamine release from the locus coeruleus  
1006 to the dorsal hippocampus promotes spatial learning and memory. *Proc Natl Acad Sci U S A* 2016,  
1007 113:14835-14840.  
1008 81. Waddell S: Dopamine reveals neural circuit mechanisms of fly memory. *Trends Neurosci* 2010, 33:457-  
1009 464.  
1010 82. Wicker-Thomas C, Hamann M: Interaction of dopamine, female pheromones, locomotion and sex behavior  
1011 in *Drosophila melanogaster*. *J Insect Physiol* 2008, 54:1423-1431.  
1012 83. Cottrell GA: Occurrence of dopamine and noradrenaline in the nervous tissue of some invertebrate species.  
1013 *Br J Pharmacol Chemother* 1967, 29:63-69.  
1014 84. Kindt KS, Quast KB, Giles AC, De S, Hendrey D, Nicastro I, Rankin CH, Schafer WR: Dopamine  
1015 mediates context-dependent modulation of sensory plasticity in *C. elegans*. *Neuron* 2007, 55:662-676.  
1016 85. Carlsson A: Thirty years of dopamine research. *Adv Neurol* 1993, 60:1-10.  
1017 86. Schultz W: Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *J  
1018 Neurophysiol* 1986, 56:1439-1461.  
1019 87. Joshua M, Adler A, Bergman H: The dynamics of dopamine in control of motor behavior. *Curr Opin  
1020 Neurobiol* 2009, 19:615-620.  
1021 88. Dayan P, Balleine BW: Reward, motivation, and reinforcement learning. *Neuron* 2002, 36:285-298.  
1022 89. Wise RA: Dopamine, learning and motivation. *Nat Rev Neurosci* 2004, 5:483-494.  
1023 90. Heinz A, Schlagenhauf F: Dopaminergic dysfunction in schizophrenia: salience attribution revisited.  
1024 *Schizophr Bull* 2010, 36:472-485.  
1025 91. Sasaki K, Yamasaki K, Nagao T: Neuro-endocrine correlates of ovarian development and egg-laying  
1026 behaviors in the primitively eusocial wasp (*Polistes chinensis*). *J Insect Physiol* 2007, 53:940-949.  
1027 92. Penick CA, Brent CS, Dolezal K, Liebig J: Neurohormonal changes associated with ritualized combat and  
1028 the formation of a reproductive hierarchy in the ant *Harpegnathos saltator*. *J Exp Biol* 2014, 217:1496-  
1029 1503.  
1030 93. Okada Y, Sasaki K, Miyazaki S, Shimoji H, Tsuji K, Miura T: Social dominance and reproductive  
1031 differentiation mediated by dopaminergic signaling in a queenless ant. *J Exp Biol* 2015, 218:1091-1098.  
1032 94. Bloch G, Simon T, Robinson GE, Hefetz A: Brain biogenic amines and reproductive dominance in bumble  
1033 bees (*Bombus terrestris*). *J Comp Physiol A* 2000, 186:261-268.  
1034 95. JW. H, Woodring J: Elevated brain dopamine levels associated with ovary development in queenless  
1035 worker honeybees (*Apis mellifera L.*) *Comparative Biochemistry and Physiology* 1995, 111C:271-279.  
1036 96. T. D, Simoes ZLP, Bitondi MMG: Dietary dopamine causes ovary activation in queenless *Apis mellifera*  
1037 workers. *Apidologie, Springer Verlag* 2003, 34:281-289.  
1038 97. Matsuyama S, Nagao T, Sasaki K: Consumption of tyrosine in royal jelly increases brain levels of  
1039 dopamine and tyramine and promotes transition from normal to reproductive workers in queenless honey  
1040 bee colonies. *Gen Comp Endocrinol* 2015, 211:1-8.  
1041 98. Beggs KT, Glendining KA, Marechal NM, Vergoz V, Nakamura I, Slessor KN, Mercer AR: Queen  
1042 pheromone modulates brain dopamine function in worker honey bees. *Proc Natl Acad Sci U S A* 2007,  
1043 104:2460-2464.  
1044 99. Akasaka S, Sasaki K, Harano K, Nagao T: Dopamine enhances locomotor activity for mating in male  
1045 honeybees (*Apis mellifera L.*). *J Insect Physiol* 2010, 56:1160-1166.  
1046 100. Mezawa R, Akasaka S, Nagao T, Sasaki K: Neuroendocrine mechanisms underlying regulation of mating  
1047 flight behaviors in male honey bees (*Apis mellifera L.*). *Gen Comp Endocrinol* 2013, 186:108-115.  
1048 101. S. DOT, A. R. W. E: Reactivation of juvenile hormone synthesis in adult drones of the honey bee, *Apis  
1049 mellifera carnica*. *Experientia, Springer* 1995, 51:945-952.  
1050 102. Giray T, Robinson GE: Common endocrine and genetic mechanisms of behavioral development in male  
1051 and worker honey bees and the evolution of division of labor. *Proc Natl Acad Sci U S A* 1996, 93:11718-  
1052 11722.  
1053 103. Fitzgerald P, Dinan TG: Prolactin and dopamine: what is the connection? A review article. *J  
1054 Psychopharmacol* 2008, 22:12-19.  
1055 104. Brown RS, Herbison AE, Grattan DR: Effects of Prolactin and Lactation on A15 Dopamine Neurones in  
1056 the Rostral Preoptic Area of Female Mice. *J Neuroendocrinol* 2015, 27:708-717.

1057 105. Kauppila A, Martikainen H, Puistola U, Reinila M, L R: Hypoprolactinemia and ovarian function. *Fertility*  
1058 and *Sterility* 1988  
1059 , 49:437–441.

1060 106. Majumdar A, Mangal NS: Hyperprolactinemia. *J Hum Reprod Sci* 2013, 6:168-175.

1061 107. Brown RS, Herbison AE, Grattan DR: Prolactin regulation of kisspeptin neurones in the mouse brain and  
1062 its role in the lactation-induced suppression of kisspeptin expression. *J Neuroendocrinol* 2014, 26:898-908.

1063 108. Snowdon CT, Ziegler TE: Variation in prolactin is related to variation in sexual behavior and contact  
1064 affiliation. *PLoS One* 2015, 10:e0120650.

1065 109. Fisher CR, Graves KH, Parlow AF, Simpson ER: Characterization of mice deficient in aromatase (ArKO)  
1066 because of targeted disruption of the cyp19 gene. *Proc Natl Acad Sci U S A* 1998, 95:6965-6970.

1067 110. Toda K, Takeda K, Okada T, Akira S, Saibara T, Kaname T, Yamamura K, Onishi S, Shizuta Y: Targeted  
1068 disruption of the aromatase P450 gene (Cyp19) in mice and their ovarian and uterine responses to 17beta-  
1069 oestradiol. *J Endocrinol* 2001, 170:99-111.

1070 111. Hewitt SC, Winuthayanon W, Korach KS: What's new in estrogen receptor action in the female  
1071 reproductive tract. *J Mol Endocrinol* 2016, 56:R55-71.

1072 112. R. L: What is an hystricomorph? In *The Biology of hystricomorph rodents*. Edited by IW R, Weir BJ e.  
1073 London: Zoological Society of London; 1974: 7-20

1074 113. Kent J, Ryle M: Histochemical studies on three gonadotrophin-responsive enzymes in the infantile mouse  
1075 ovary. *J Reprod Fertil* 1975, 42:519-536.

1076 114. Paranko J, Pelliniemi LJ: Differentiation of smooth muscle cells in the fetal rat testis and ovary:  
1077 localization of alkaline phosphatase, smooth muscle myosin, F-actin, and desmin. *Cell Tissue Res* 1992,  
1078 268:521-530.

1079 115. Brannstrom M, Mayrhofer G, Robertson SA: Localization of leukocyte subsets in the rat ovary during the  
1080 periovulatory period. *Biol Reprod* 1993, 48:277-286.

1081 116. Best CL, Pudney J, Welch WR, Burger N, Hill JA: Localization and characterization of white blood cell  
1082 populations within the human ovary throughout the menstrual cycle and menopause. *Hum Reprod* 1996,  
1083 11:790-797.

1084 117. Magoffin DA: The ovarian androgen-producing cells: a 2001 perspective. *Rev Endocr Metab Disord* 2002,  
1085 3:47-53.

1086 118. Berkholz CB, Lai BE, Woodruff TK, Shea LD: Distribution of extracellular matrix proteins type I  
1087 collagen, type IV collagen, fibronectin, and laminin in mouse folliculogenesis. *Histochem Cell Biol* 2006,  
1088 126:583-592.

1089 119. S. A: FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics* 2010.

1090 120. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data.  
1091 *Bioinformatics* 2014, 30:2114-2120.

1092 121. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: TopHat2: accurate alignment of  
1093 transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013, 14:R36.

1094 122. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012, 9:357-359.

1095 123. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter  
1096 L: Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform  
1097 switching during cell differentiation. *Nat Biotechnol* 2010, 28:511-515.

1098 124. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P,  
1099 Fitzgerald S, Gil L, et al: Ensembl 2016. *Nucleic Acids Res* 2016, 44:D710-716.

1100 125. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: BLAST+:  
1101 architecture and applications. *BMC Bioinformatics* 2009, 10:421.

1102 126. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol*  
1103 1990, 215:403-410.

1104 127. Anders S, Pyl PT, Huber W: HTSeq--a Python framework to work with high-throughput sequencing data.  
1105 *Bioinformatics* 2015, 31:166-169.

1106 128. Risso D, Schwartz K, Sherlock G, Dudoit S: GC-content normalization for RNA-Seq data. *BMC*  
1107 *Bioinformatics* 2011, 12:480.

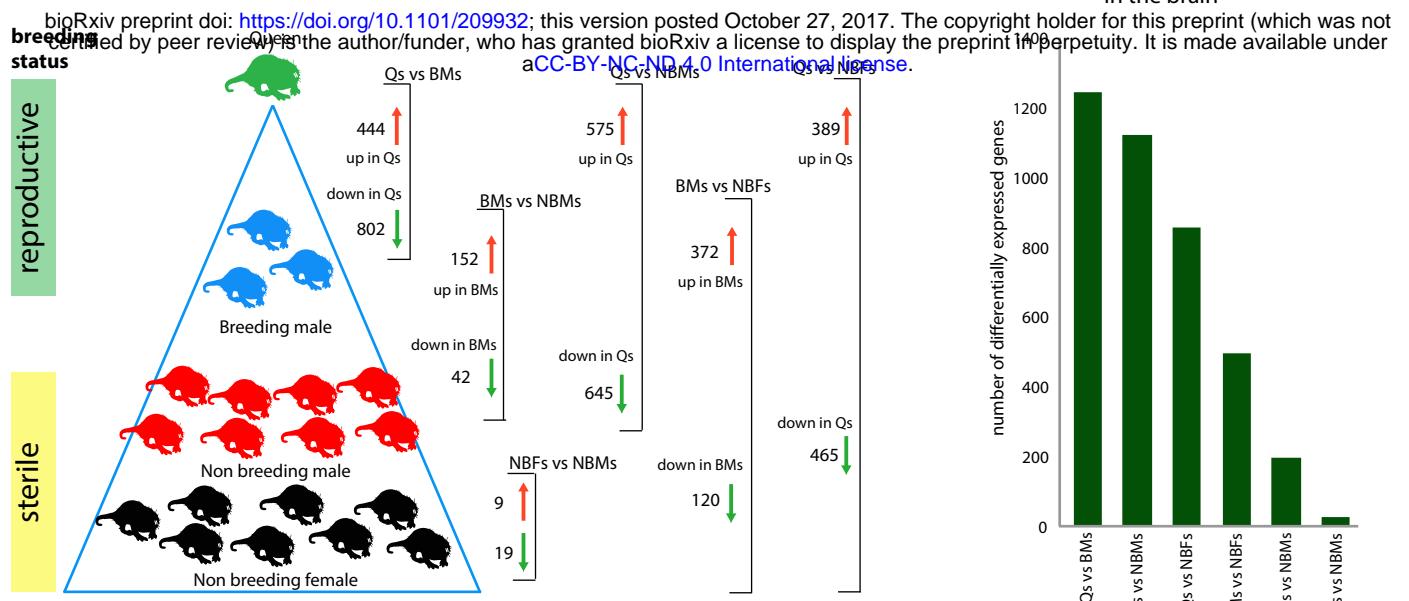
1108 129. Risso D, Ngai J, Speed TP, Dudoit S: Normalization of RNA-seq data using factor analysis of control  
1109 genes or samples. *Nat Biotechnol* 2014, 32:896-902.

1110 130. Robinson MD, McCarthy DJ, Smyth GK: edgeR: a Bioconductor package for differential expression  
1111 analysis of digital gene expression data. *Bioinformatics* 2010, 26:139-140.

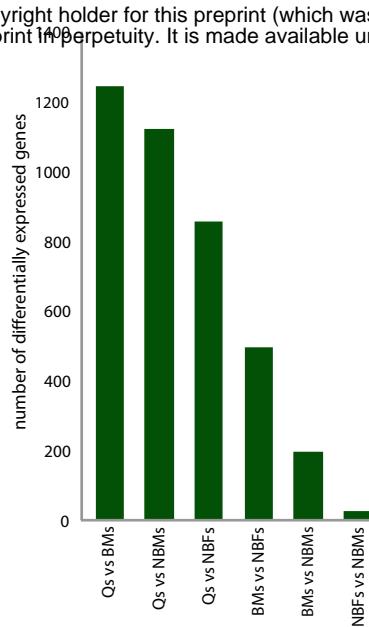
1112 131. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA, Moulton HM, DeJesus P, Che J,  
1113 Mulder LC, et al: Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4  
1114 in Virus Budding. *Cell Host Microbe* 2015, 18:723-735.  
1115 132. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik  
1116 KM, Lachmann A, et al: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.  
1117 *Nucleic Acids Res* 2016, 44:W90-97.  
1118

# Figure 1

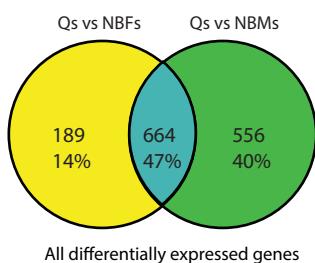
## A Gene expression differences in the brain



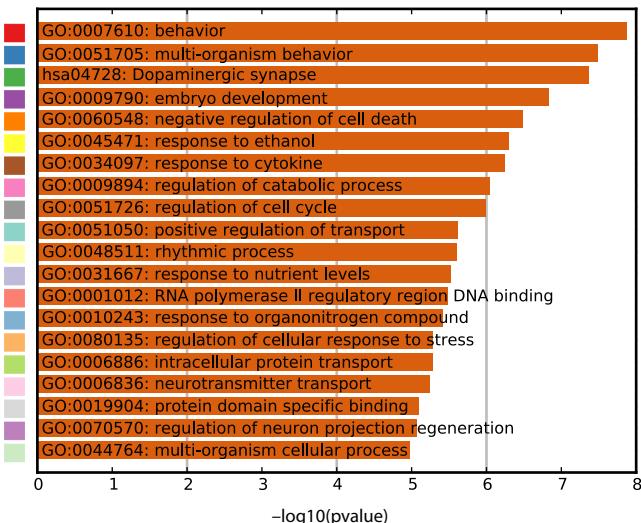
## B Total number of differentially expressed genes in the brain



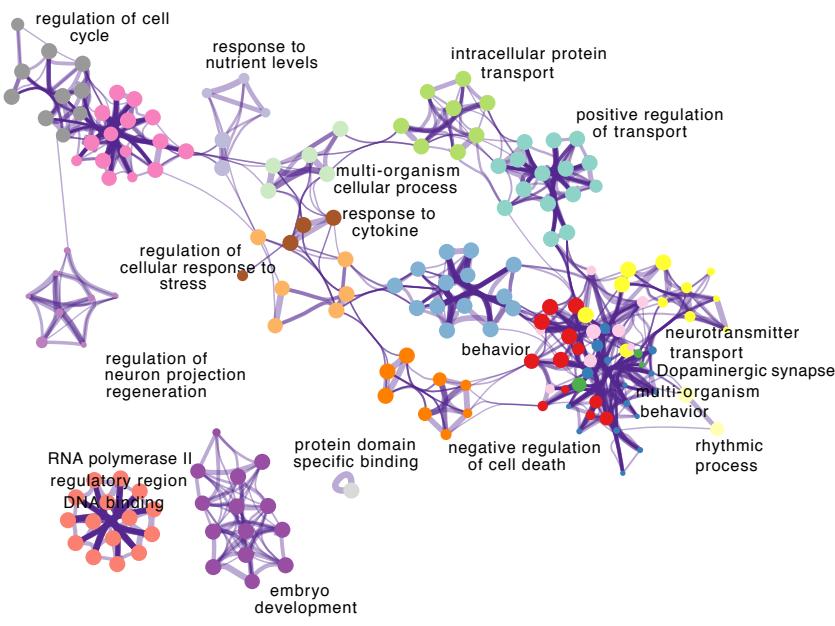
## C Venn diagram of DEGs between brains of Qs vs NBFs and Qs vs NBM



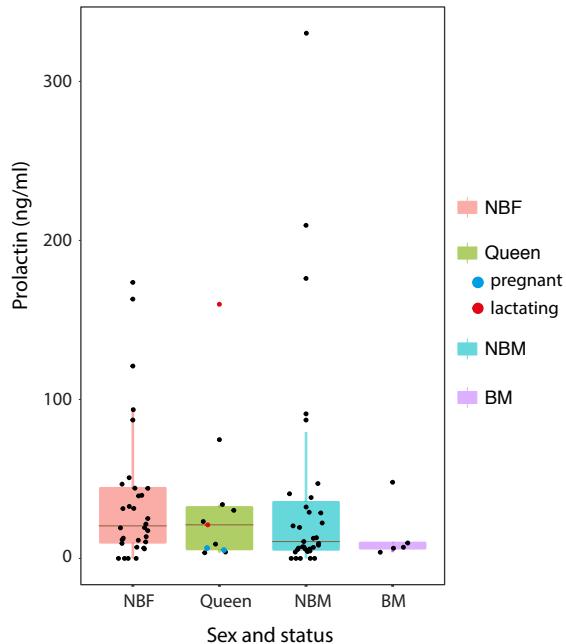
## D Heatmap of enriched terms for queen genes (brain)

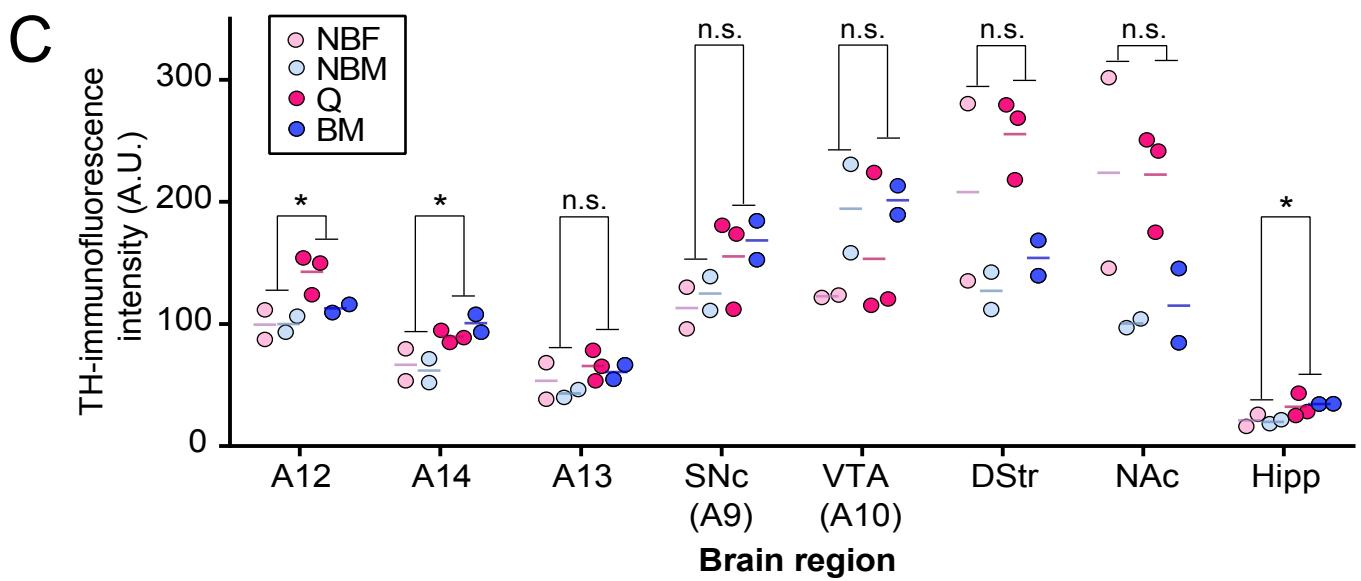
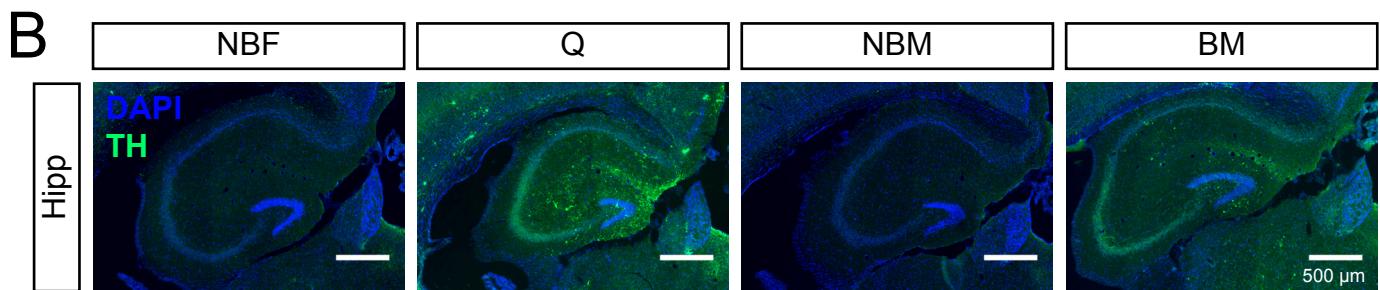
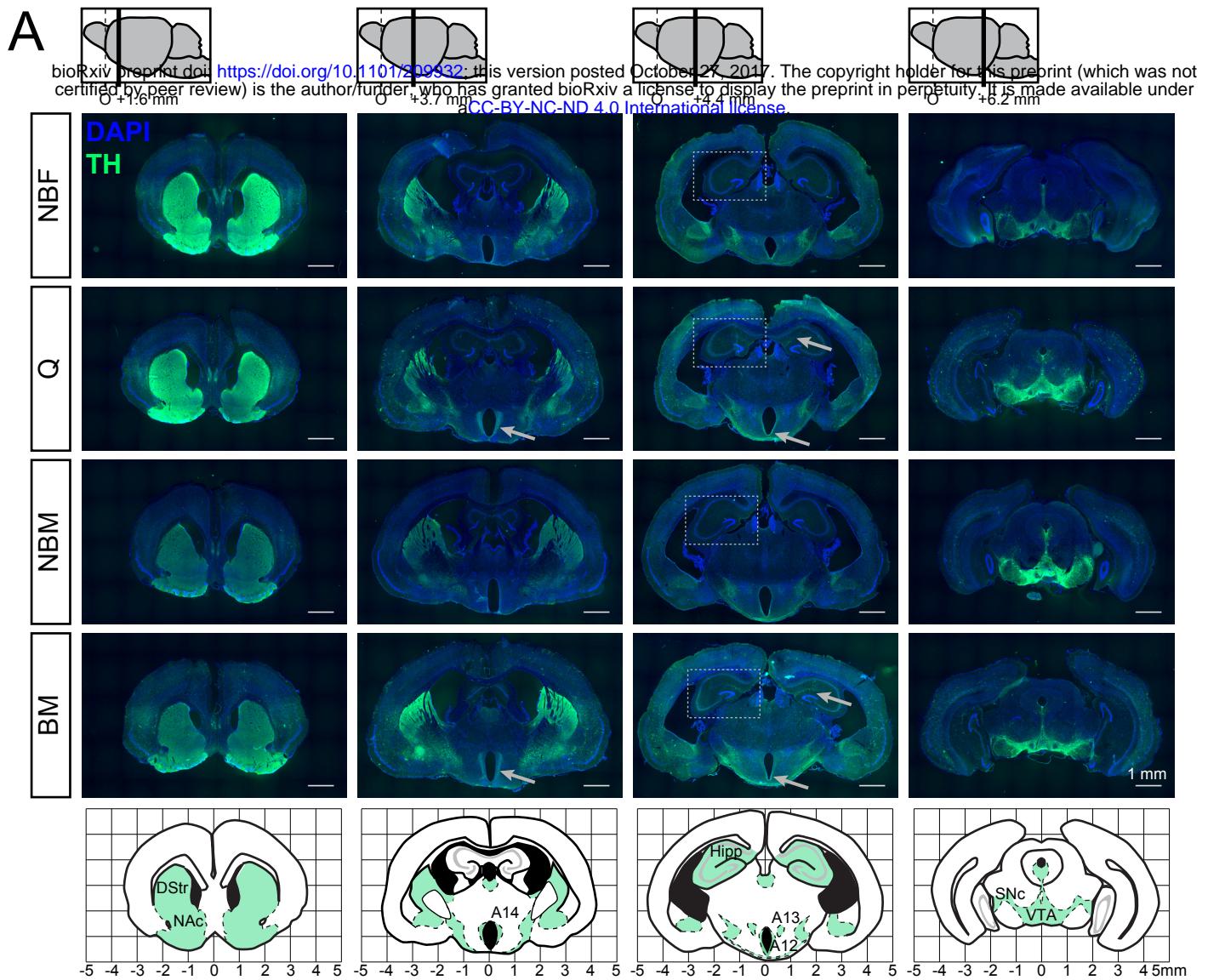


## E Network of enriched terms for queen genes (brain)



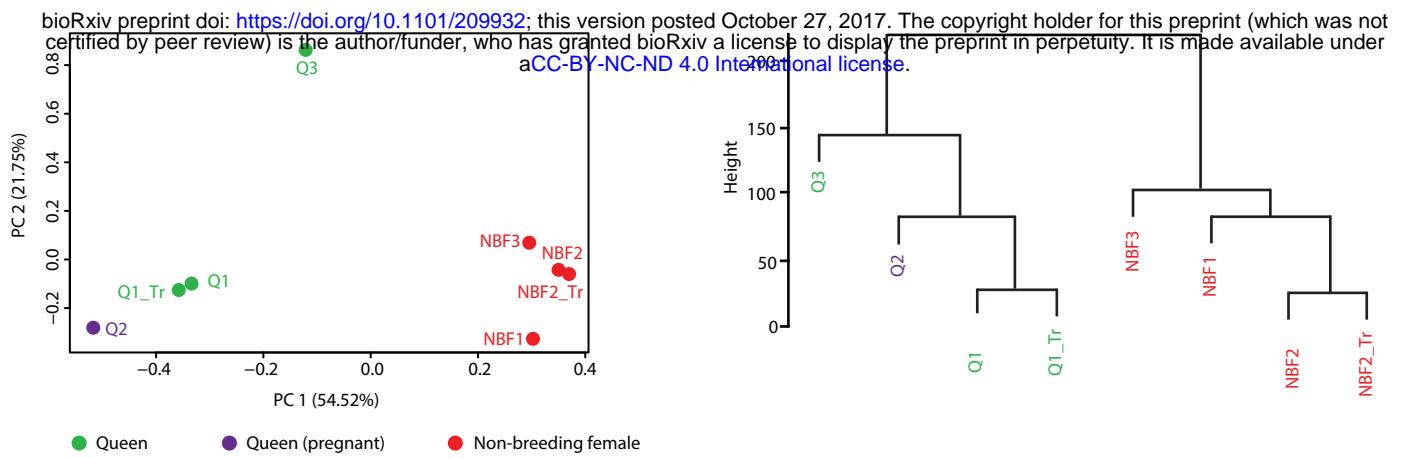
## F Plasma prolactin level in breeding and non-breeding NMRs



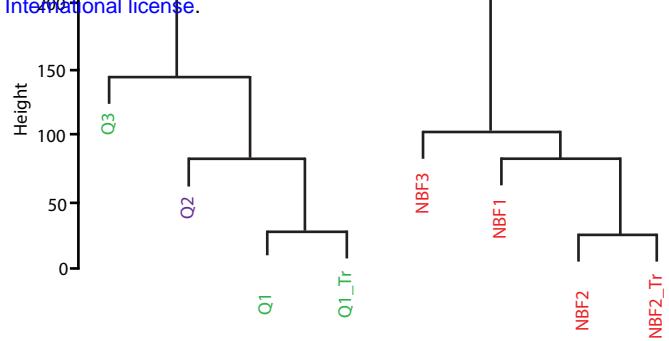


# Figure 3

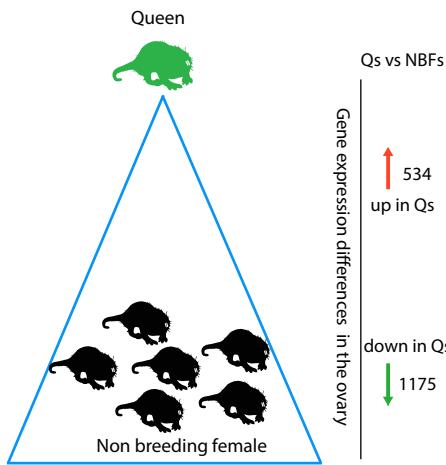
## A PCA plot showing the clustering of Q and NBF ovaries



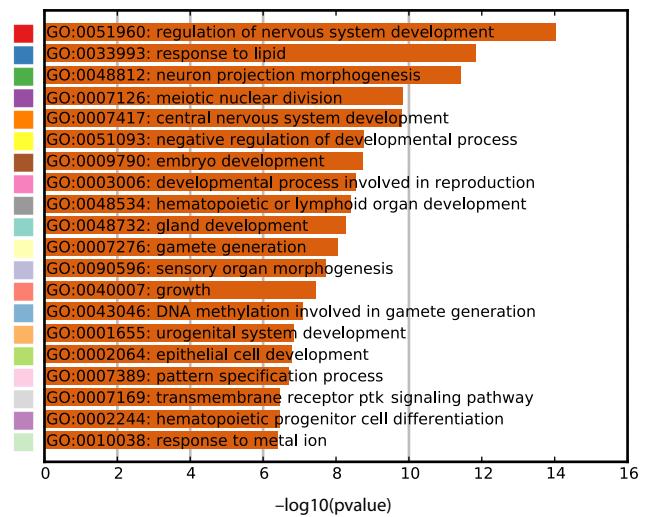
## B Hierarchical clustering of Q and NBF ovaries



## C DEGs between Q and NBF ovaries

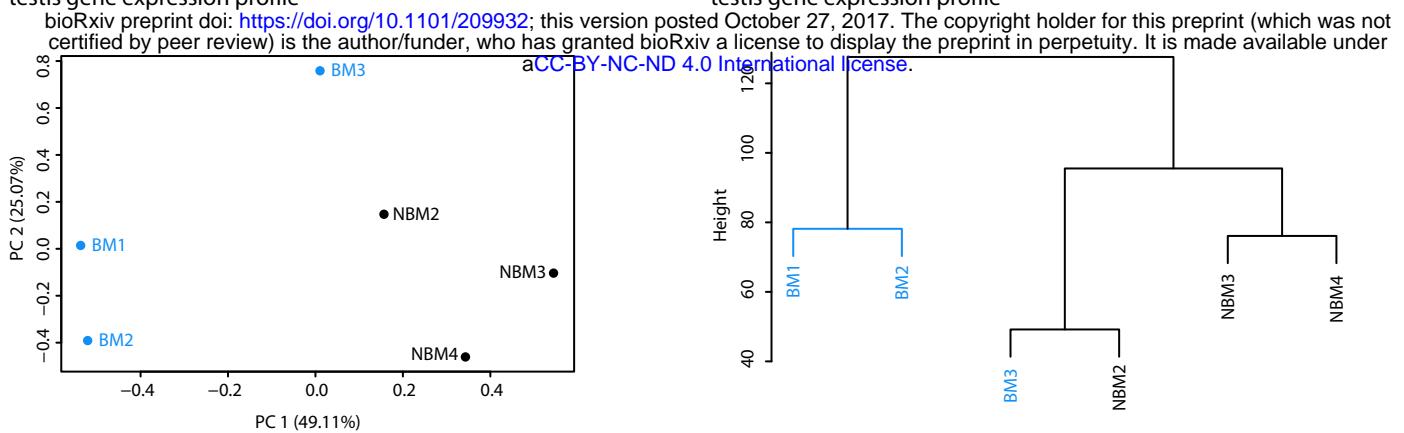


## D Heatmap of enriched terms for DEGs between Q and NBF ovaries

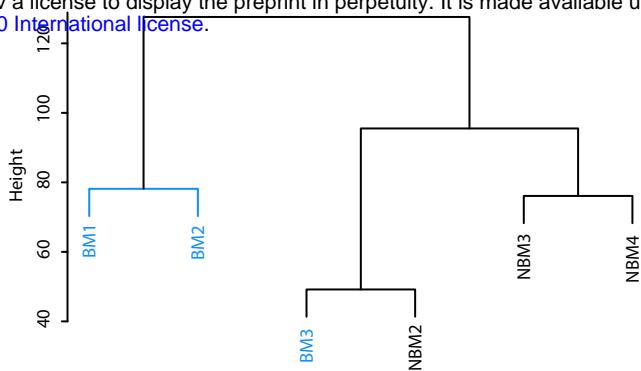


# Figure 4

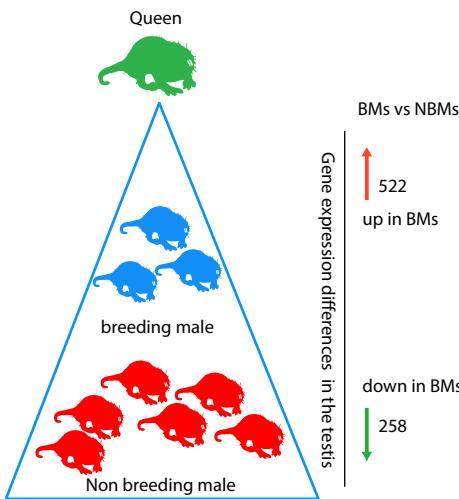
## A PCA plot showing the clustering of BMs and NBMs testis gene expression profile



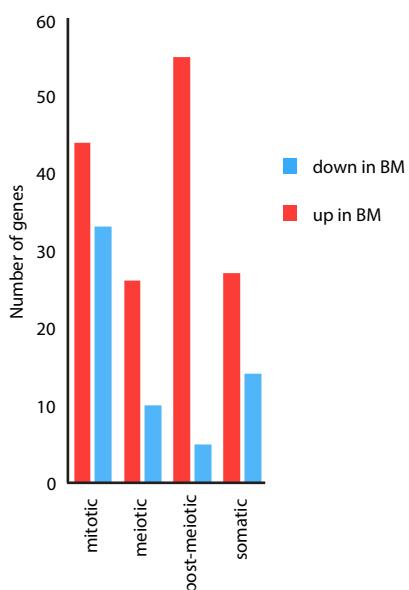
## B Hierarchical clustering of BMs and NBMs testis gene expression profile



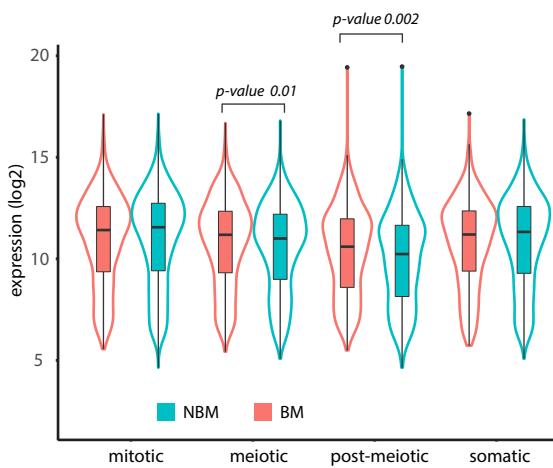
## C DEGs between BMs and NBMs testis



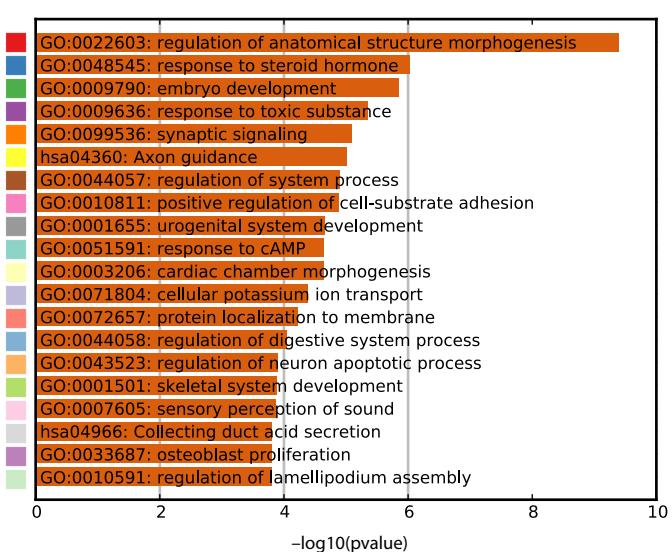
## D Number of DEGs that belong to different stages of spermatogenesis



## E Violin and box plots showing expression level of NMR genes that belong to different clusters of mouse spermatogenesis



## F Heatmap of enriched terms for DEGs between BMs and NBMs testis



1 **Figure legends**

2

3 **Figure 1: Gene expression profile of naked mole rat colony members.** **A)** Number of DEGs between  
4 the different members of a naked mole rat colony brains. DEGs (FDR <0.05, log fold change >1) were  
5 divided into up regulated (higher expression in one group) and down regulated (lower expression in one  
6 group) and the final count is plotted. **B)** Total number of DEGs in each comparison (similar to *Figure 1A*,  
7 but total number in each comparison plotted). **C)** Venn diagram showing DEGs that are common in the  
8 comparison between Qs vs NBFs and Qs vs NBMs (Q genes, 661 genes in total), and genes that show  
9 specific differential expression between Qs vs NBFs (192 genes) and Qs vs NBMs (560 genes). **D)**  
10 Heatmap of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark Gene Sets, KEGG  
11 Pathway) for genes that show differential expression in the comparison between Qs vs NBFs and Qs vs  
12 NBMs (Q genes). Significance of enrichment is indicated on the x-axis in  $-\log_{10}(p\text{-value})$ . The color code  
13 on the y-axis is used to show the clustering and relation of the significantly enriched networks/terms  
14 (shown in *Figure 1E*). Detailed list of enriched terms and families can be found in *Additional file 3*. **E)**  
15 Network of enriched terms (*Figure 1D*) colored by cluster ID, indicating the relationship between the  
16 different enriched terms (cluster ID color correspond to the color code shown in *Figure 1D* y-axis). Nodes  
17 that share the same cluster are close to each other. Each circle node represents a term and the size of the  
18 circle is proportional to the number of genes that fall into that term, and the identity of the cluster is  
19 indicated by its color (nodes of the same color belong to the same cluster). Similar terms are linked by an  
20 edge (the thickness of the edge represents the similarity score). One term from each cluster is selected to  
21 have its term description shown as label. **F)** Plasma prolactin concentrations in breeding and non-breeding  
22 NMRs in samples taken across thirteen colonies. Abbreviations: Q, queen; Q1\_Tr, technical replicate for  
23 Q1; NBF, non-breeding female; NBM, non-breeding male; BM, breeding male; DEGs, differentially  
24 expressed genes.

25

26 **Figure 2: Tyrosine hydroxylase (TH) immunostaining of NMR brain sections. A)** Representative  
27 immunostaining of whole coronal sections for the four reproductive castes (NBF, non-breeding female;  
28 NBM, non-breeding male; Q, Queen; BM, breeding male). The top panel indicates the coordinates of the  
29 coronal sections relative to O, the origin of slice numbering. Grey arrows point to the regions that show  
30 differential staining (quantified in C). Dashed boxes indicate the hippocampal region. The bottom panel  
31 represents the regions where TH-staining has been observed. (Labels refer to regions quantified in C:  
32 DStr, dorsal striatum; Hipp, hippocampus; NAc, nucleus accumbens; SNc, substantia nigra compacta;  
33 VTA, ventral tegmental area). **B)** TH immunofluorescence in the hippocampus. **C)** Intensity of the TH  
34 staining in the different brain regions. Non-parametric Mann-Whitney test, \*p < 0.05; n.s., non-  
35 significant.

36  
37 **Figure 3) Gene expression profiles of Q and NBF ovaries. A)** Principal component analysis (PCA) plot  
38 showing the clustering of Q and NBF ovaries based on global expression profile. **B)** Hierarchical  
39 clustering of ovary samples using euclidean distance matrix computation and ward.D2 agglomeration  
40 method. **C)** number of DEGs between Q and NBF ovaries. **D)** Heatmap of enriched terms (Canonical  
41 Pathways, GO Biological Processes, Hallmark Gene Sets, KEGG Pathway) for all DEGs colored by p-  
42 values. Significance of enrichment is indicated on the x-axis in  $-\log_{10}(p\text{-value})$ . The color code on the y-  
43 axis is used to show the clustering and relation of this networks (shown in *Supplementary Figure 7C*).  
44 More information about the enriched terms can be found in *Additional file 26A*. (Q1\_Tr and NBF2\_Tr  
45 indicate technical replicates for Q1 and NBF2 respectively)

46  
47 **Figure 4: Gene expression profile of breeding and non-breeding males testes. A)** Principal  
48 component analysis (PCA) plot showing the clustering of BMs and NBMs testis based on global gene  
49 expression profile. **B)** Hierarchical clustering of testis samples using euclidean distance matrix  
50 computation and ward.D2 agglomeration method. **C)** Number of DEGs between breeding and non-  
51 breeding animals. **D)** Bar plot showing the number of NMR DEGs (BMs vs NBMs testes) that belong to

52 different stages of mouse spermatogenesis. Gene clusters that are expressed at different stages of mouse  
53 spermatogenesis were obtained from Germonline, and mapped to the DEGs that were identified in the  
54 comparison between breeding and non-breeding NMR testes. The DEGs which map to different stages of  
55 spermatogenesis were divided into up or down regulated, (if they show higher or lower expression in BM  
56 respectively) and plotted. **E)** Violin and box plots showing the average expression level of all BMs and  
57 NBMs genes (NMR testes) that were mapped to genes which belong to the different clusters (4 clusters)  
58 of mouse spermatogenesis. Gene clusters that are expressed at different stages of mouse spermatogenesis  
59 were obtained from Germonline, mapped to NMR BMs and NBMs testis expression data, and plotted (x-  
60 axis, cluster name; y-axis: average expression level in  $\log_2$  scale). **F)** Heatmap of enriched terms  
61 (Canonical Pathways, GO Biological Processes, Hallmark Gene Sets, KEGG Pathway) for all DEGs  
62 colored by p-values (generated using Metascape). Significance of enrichment is indicated on the x-axis in  
63  $-\log_{10}(p\text{-value})$ . The color code on the y-axis is used to show the clustering and relation of this network  
64 (shown in *Supplementary Figure 12C*). Detailed list of enriched terms and genes that belong to these  
65 terms is provided in *Additional file 30A*.

66

67

68

69

70

71

72

73

74

75

76

77

78 **Supplementary Figure legends**

79

80 **Supplementary Figure 1: A)** Principal component analysis (PCA) plots showing the clustering of  
81 different NMR colony members based on global brain gene expression profile. **B)** Cluster dendrogram  
82 showing hierarchical clustering of the NMR brain gene expression. Hierarchical clustering was generated  
83 using euclidean distance matrix computation and ward.D2 agglomeration method. **C)** Enriched biological  
84 process terms for DEGs that are common in the comparison between Qs vs NBMs and Qs vs NBMs (Q  
85 genes) generated using Enrichr. The length of the bar represents the significance of that specific gene-set  
86 or term, and the color intensity provides additional information about the significance (the brighter the  
87 color, the more significant that term is). Statistical information used to generate the graph including the p-  
88 value and other enriched terms are provided in *Additional file 18*. **D)** Enriched molecular functions terms  
89 for DEGs that are common in the comparison between Qs vs NBMs and Qs vs NBMs (Q genes)  
90 generated using Enrichr. The length of the bar represents the significance of that specific gene-set or term,  
91 and the color intensity provides additional information about the significance (the brighter the color, the  
92 more significant that term is). Statistical information used to generate the graph including the p-value and  
93 other enriched terms are provided in *Additional file 19*.

94

95 **Supplementary Figure 2: Gene expression differences between the brains of Qs and NBFs. A)**  
96 Principal component analysis (PCA) plot showing the clustering of Qs and NBFs brains based global  
97 gene expression profile. **B)** Cluster heatmap of Qs and NBFs brain gene expression. Sample distance was  
98 calculated euclidean distance matrix computation and cluster agglomeration was done using ward.D2  
99 method. Heatmap color indicate the euclidean distance between samples indicated in the heatmap color  
100 key. **C)** Volcano plot showing significance versus fold-change. The log fold change in expression is  
101 indicated on the x-axis and significance  $-\log_{10}(p\text{-value})$  is indicated on the y-axis. DEGs are indicated by  
102 red color. **D)** Heatmap of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark Gene  
103 Sets, KEGG Pathway) for DEGs between Qs and NBFs (generated using Metascape). Significance of

104 enrichment is indicated on the x-axis in  $-\log_{10}(p\text{-value})$ . The color code on the y-axis is used to show the  
105 clustering and relation of these networks (shown in *Supplementary Figure 2E*). Detailed list of enriched  
106 terms can be found at *Additional file 3*. **E**) Network of enriched terms (*Supplementary Figure 2D*) colored  
107 by cluster ID, indicating the relationship between the different enriched terms (cluster ID color  
108 correspond to the color code shown in *Supplementary Figure 2D* y-axis). Nodes that share the same  
109 cluster are close to each other. Each circle node represents a term and the size of the circle is proportional  
110 to the number of genes that fall into that term, and the identity of the cluster is indicated by its color  
111 (nodes of the same color belong to the same cluster). Similar terms are linked by an edge (the thickness of  
112 the edge represents the similarity score). One term from each cluster is selected to have its term  
113 description shown as label. **E**) Enriched biological process terms for DEGs that are common in the  
114 comparison between Qs vs NBFs generated using Enrichr. The length of the bar represents the  
115 significance of that specific gene-set or term, and the color intensity provides additional information  
116 about the significance (the brighter the color, the more significant that term is). Statistical information  
117 used to generate the graph including the p-value and other enriched terms are provided in *Additional file*  
118 **4**. **G**) Enriched molecular functions terms in for DEGs that are common in the comparison between Qs vs  
119 NBFs generated using Enrichr. The length of the bar represents the significance of that specific gene-set  
120 or term, and the color intensity provides additional information about the significance (the brighter the  
121 color, the more significant that term is). Statistical information used to generate the graph including the p-  
122 value and other enriched terms are provided in *Additional file 5*.

123

124 **Supplementary Figure 3: Gene expression differences between the brains of BMs NBMs. A)** Scatter  
125 plot of testis size to body weight. The size of the circles in the scatter plot are proportional to testis to  
126 body weight ratio. BMs are depicted in blue and NBMs in black. **B)** Principal component analysis (PCA)  
127 plot showing the clustering of BMs and NBMs based global brain gene expression profile. **C)** Cluster  
128 heatmap of BMs and NBMs brain gene expression. Sample distance was calculated euclidean distance  
129 matrix computation and cluster agglomeration was done using ward.D2 method. Heatmap colors indicate

130 the euclidean distance between samples indicated in the heatmap color key. **D)** Volcano plot showing the  
131 significance versus fold-change. The log fold change in expression is indicated on the x-axis and  
132 significance in  $-\log_{10}(p\text{-value})$  is indicated on the y-axis. DEGs are indicated by the red color. **E)** Heatmap  
133 of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark Gene Sets, KEGG Pathway)  
134 for DEGs between BMs and NBMs (generated using Metascape). Significance of enrichment is indicated  
135 on the x-axis in  $-\log_{10}(p\text{-value})$ . The color code on the y-axis is used to show the clustering and relation  
136 of this networks (shown in *Supplementary Figure 3F*). Detailed list of enriched terms can be found at  
137 *Additional file 7*. **F)** Network of enriched terms (*Supplementary Figure 3E*) colored by cluster ID,  
138 indicating the relationship between the different enriched terms (cluster ID color correspond to the color  
139 code shown in *Supplementary Figure 3E y-axis*). Nodes that share the same cluster are close to each  
140 other. Each circle node represents a term and the size of the circle is proportional to the number of genes  
141 that fall into that term, and the identity of the cluster is indicated by its color (nodes of the same color  
142 belong to the same cluster). Similar terms are linked by an edge (the thickness of the edge represents the  
143 similarity score). One term from each cluster is selected to have its term description shown as label. **G)**  
144 Enriched biological process terms for DEGs that are common in the comparison between BMs vs NBMs  
145 generated using Enrichr. The length of the bar represents the significance of that specific gene-set or term,  
146 and the color intensity provides additional information about the significance (the brighter the color, the  
147 more significant that term is). Statistical information used to generate the graph including the p-value and  
148 other enriched terms are provided in *Additional file 8*. **H)** Enriched molecular functions terms in for  
149 DEGs that are common in the comparison between BMs vs NBMs generated using Enrichr. The length of  
150 the bar represents the significance of that specific gene-set or term, and the color intensity provides  
151 additional information about the significance (the brighter the color, the more significant that term is).  
152 Statistical information used to generate the graph including the p-value and other enriched terms are  
153 provided in *Additional file 9*.

154

155 **Supplementary Figure 4: Gene expression differences between the brains of Qs and BMs. A)**

156 Principal component analysis (PCA) plot showing the clustering of Qs and BMs based global brain gene  
157 expression profile. **B)** Cluster heatmap of Qs and BMs brain gene expression. Sample distance was  
158 calculated euclidean distance matrix computation and cluster agglomeration was done using ward.D2  
159 method. Heatmap color indicates the euclidean distance between samples indicated in the heatmap color  
160 key. **C)** Volcano plot showing significance versus fold-change. The log fold change in expression is  
161 indicated on the x-axis and significance in  $-\log_{10}(p\text{-value})$  is indicated on the y-axis. DEGs are indicated by  
162 the red color. **D)** Heatmap of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark  
163 Gene Sets, KEGG Pathway) for DEGs between Qs and BMs (generated using Metascape). Significance  
164 of enrichment is indicated on the x- in  $-\log_{10}(p\text{-value})$ . The color code on the y-axis is used to show the  
165 clustering and relation of this networks (shown in *Supplementary Figure 4E*). Detailed list of enriched  
166 terms can be found at *Additional file 11*. **E)** Network of enriched terms (*Supplementary Figure 4D*)  
167 colored by cluster ID, indicating the relationship between the different enriched terms (cluster ID color  
168 correspond to the color code shown in *Supplementary Figure 4D* y-axis). Nodes that share the same  
169 cluster are close to each other. Each circle node represents a term and the size of the circle is proportional  
170 to the number of genes that fall into that term, and the identity of the cluster is indicated by its color  
171 (nodes of the same color belong to the same cluster). Similar terms are linked by an edge (the thickness of  
172 the edge represents the similarity score). One term from each cluster is selected to have its term  
173 description shown as label. **F)** Enriched biological process terms for DEGs that are common in the  
174 comparison between Qs vs BMs generated using Enrichr. The length of the bar represents the significance  
175 of that specific gene-set or term, and the color intensity provides additional information about the  
176 significance (the brighter the color, the more significant that term is). Statistical information used to  
177 generate the graph including the p-value and other enriched terms are provided in *Additional file 12*. **G)**  
178 Enriched molecular functions terms for DEGs that are common in the comparison between Qs vs BMs  
179 generated using Enrichr. The length of the bar represents the significance of that specific gene-set or term,  
180 and the color intensity provides additional information about the significance (the brighter the color, the

181 more significant that term is). Statistical information used to generate the graph including the p-value and  
182 other enriched terms are provided in *Additional file 13*.

183

184 **Supplementary Figure 5: Gene expression differences between the brains of NBFs and NBMs. A)**  
185 Cluster heatmap of NBFs and NBMs gene brain gene expression. Sample distance was calculated  
186 euclidean distance matrix computation and cluster agglomeration was done using ward.D2 method.  
187 Heatmap color indicate the euclidean distance between samples indicated in the heatmap color key. **B)**  
188 Volcano plot showing the significance versus fold-change. The log fold change in expression is indicated  
189 on the x-axis and significance in  $-\log_{10}(p\text{-value})$  is indicated on the y-axis. DEGs are indicated by the red  
190 color.

191

192 **Supplementary Figure 6: Venn diagrams showing the number of DEGs in the brain that were**  
193 **identified in comparisons among different sex/status groups, and the relationship between them. A)**  
194 Venn diagram showing all DEGs that are common in the comparison between Qs vs NBFs, Qs vs NBMs,  
195 and Qs vs BMs. **B)** Venn diagram showing DEGs that also shows higher expression in the comparison  
196 between Qs vs NBFs, Qs vs NBMs, and Qs vs BMs. **C)** Venn diagram showing DEGs that also show  
197 lower expression in the comparison between Qs vs NBFs, Qs vs NBMs, and Qs vs BMs. **D)** Venn  
198 diagram showing DEGs that also show higher expression in the comparison between Qs vs NBFs and Qs  
199 vs NBMs. **E)** Venn diagram showing DEGs that also show lower expression in the comparison between  
200 Qs vs NBFs and Qs vs NBMs. **F)** Venn diagram showing all DEGs that are common in the comparison  
201 between BMs vs NBFs and BMs vs NBMs. **G)** Venn diagram showing DEGs that also show higher  
202 expression in the comparison between BMs vs NBFs and BMs vs NBMs. **H)** Venn diagram showing  
203 DEGs that also show lower expression in the comparison between BMs vs NBFs and BMs vs NBMs.

204

205 **Supplementary Figures 7: Gene expression profile of Q and NBF ovaries. A)** Cluster heatmap of  
206 ovary samples. Sample distance was calculated euclidean distance matrix computation and cluster

207 agglomeration was done using ward.D2 method. Heatmap color indicates the euclidean distance between  
208 samples indicated in the heatmap color key. **B)** Volcano plot showing the significance versus fold-change.  
209 The log fold change in expression is indicated on the x-axis and significance in  $-\log_{10}(p\text{-value})$  is  
210 indicated on the y-axis. DEGs are indicated by the red color. **C)** Network of enriched terms colored by  
211 cluster ID, indicating the relationship between the different enriched terms (*Figure 3D*). Nodes that share  
212 the same cluster are close to each other. Each circle node represents a term and the size of the circle is  
213 proportional to the number of genes that fall into that term, and the identity of the cluster is indicated by  
214 its color (nodes of the same color belong to the same cluster). Similar terms are linked by an edge (the  
215 thickness of the edge represents the similarity score). One term from each cluster is selected to have its  
216 term description shown as label. **D)** Enriched biological processes for DEGs between Qs vs NBFs ovary  
217 generated using Enrichr. The length of the bar represents the significance of that specific gene-set or term,  
218 and the color intensity provides additional information about the significance (the brighter the color, the  
219 more significant that term is). Statistical information used to generate the graph including the p-value and  
220 other enriched terms are provided in *Additional file 27a*. **E)** Enriched molecular functions for DEGs  
221 between Qs vs NBFs ovary generated using Enrichr. The length of the bar represents the significance of  
222 that specific gene-set or term, and the color intensity provides additional information about the  
223 significance (the brighter the color, the more significant that term is). Statistical information used to  
224 generate the graph including the p-value and other enriched terms are provided in *Additional file 28A*.

225

226 **Supplementary Figure 8: Gene enrichment for DEGs that also show lower expression in Q ovaries.**  
227 **A)** Heatmap of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark Gene Sets,  
228 KEGG Pathway) for DEGs that show lower expression in Q ovary (generated using Metascape). The  
229 color of the bar graph is proportion to the p-values. Significance of enrichment is indicated on the x-axis  
230 in  $-\log_{10}(p\text{-value})$ . The color code on the y-axis is used to show the clustering and relation of these  
231 networks (shown in *Supplementary Figure 8B*). More information about the enriched terms can be found  
232 in *Additional file 26B*. **B)** Network of enriched terms colored by cluster ID, indicating the relationship

233 between the different enriched terms (*Supplementary Figure 8A*). Nodes that share the same cluster are  
234 close to each other. Each circle node represents a term and the size of the circle is proportional to the  
235 number of genes that fall into that term, and the identity of the cluster is indicated by its color (nodes of  
236 the same color belong to the same cluster). Similar terms are linked by an edge (the thickness of the edge  
237 represents the similarity score). One term from each cluster is selected to have its term description shown  
238 as label. **C)** Enriched biological processes for DEGs between Qs and NBFs ovary and also show lower  
239 expression in the Qs generated using Enrichr. The length of the bar represents the significance of that  
240 specific gene-set or term, and the color intensity provides additional information about the significance  
241 (the brighter the color, the more significant that term is). Statistical information used to generate the graph  
242 including the p-value and other enriched terms are provided in *Additional file 27B*. **D)** Enriched molecular  
243 functions for DEGs between Qs and NBFs ovary and also show lower expression in the Qs generated  
244 using Enrichr. The length of the bar represents the significance of that specific gene-set or term, and the  
245 color intensity provides additional information about the significance (the brighter the color, the more  
246 significant that term is). Statistical information used to generate the graph including the p-value and other  
247 enriched terms are provided in *Additional file 28B*.

248

249 **Supplementary Figure 9: Gene enrichment for DEGs that also show higher expression in Q ovaries.**  
250 **A)** Heatmap of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark Gene Sets,  
251 KEGG Pathway) for DEGs that show higher expression in Qs ovary (generated using Metascape). The  
252 color of the bar graph is proportion to the p-values. Significance of enrichment is indicated on the x axis  
253 in  $-\log_{10}(p\text{-value})$ . The color code on the y-axis is used to show the clustering and relation of this  
254 networks (shown in *Supplementary Figure 9B*). More information about the enriched terms can be found  
255 in *Additional file 26C*. **B)** Network of enriched terms colored by cluster ID, indicating the relationship  
256 between the different enriched terms (*Supplementary Figure 9A*). Nodes that share the same cluster are  
257 close to each other. Each circle node represents a term and the size of the circle is proportional to the  
258 number of genes that fall into that term, and the identity of the cluster is indicated by its color (nodes of

259 the same color belong to the same cluster). Similar terms are linked by an edge (the thickness of the edge  
260 represents the similarity score). One term from each cluster is selected to have its term description shown  
261 as label. **C)** Enriched biological processes for DEGs between Qs and NBFs ovary and also show higher  
262 expression in the Qs generated using Enrichr. The length of the bar represents the significance of that  
263 specific gene-set or term, and the color intensity provides additional information about the significance  
264 (the brighter the color, the more significant that term is). Statistical information used to generate the graph  
265 including the p-value and other enriched terms are provided in *Additional file 27C*. **D)** Enriched molecular  
266 functions for DEGs between Qs and NBFs ovary and also show higher expression in the Q generated  
267 using Enrichr. The length of the bar represents the significance of that specific gene-set or term, and the  
268 color intensity provides additional information about the significance (the brighter the color, the more  
269 significant that term is). Statistical information used to generate the graph including the p-value and other  
270 enriched terms are provided in *Additional file 28C*.

271

272 **Supplementary Figure 10: NMR Q and NBF ovary histology.** Representative sections (at the same  
273 magnification) through the ovary of non-breeding (**A**) and Queen (**B**) NMRs: S stroma, O oocyte, P  
274 primordial follicle, 1° primary follicle, 2° secondary follicle, 2°\* secondary follicle with two oocytes, 3°  
275 tertiary follicle.

276

277 **Supplementary Figure 11: Gene expression levels of NMR genes at different stages of mouse  
278 oogenesis/folliculogenesis.** **A)** Violin plot combined with boxplot showing the expression level of NMR  
279 Qs and NBFs (ovary RNA-seq data) for clusters that show stage specific expression during mouse  
280 oogenesis/folliculogenesis. Mouse gene cluster that show a decrease in expression from the primary to  
281 small antral stage (cluster 1) and another gene cluster (cluster 5: where gene expression increased at small  
282 antral and large antral follicle stages) were obtained from (50) (753 genes in total in the two  
283 clusters). NMR Qs and NBFs ovary expression level was mapped to these mouse folliculogenesis genes  
284 names (cluster 1 and cluster 5), and the expression level of these genes in the Q and NBF NMR ovaries

285 was plotted (x-axis: clusters, y-axis: expression level in  $\log_2$ ). **B)** Violin plot in combination with boxplot  
286 showing the expression level of Qs and NBFs genes (NMR ovary) for DEGs that show up and down  
287 regulation in mouse infant vs adult whole ovary comparison (51). From 7021 DEGs that show up or down  
288 regulation in mouse infant vs adult whole ovary gene expression comparison (51) 4494 genes that show  
289  $>2$  fold change (up or down) were extracted, and their expression in Qs and NBFs (NMR ovary) was  
290 plotted. x-axis: differential expression group (down, DEGs that show lower expression in adult mouse  
291 ovary compared to infant; and up, for genes that show higher expression in adult mouse ovary compared  
292 to infant); y-axis expression level in  $\log_2$  scale. **C)** Bar plot showing the expression level of mouse oocyte  
293 specific genes in NMR Qs and NBFs ovary RNA-seq data. Mouse oocyte specific genes were obtained  
294 from [ref], and their average expression level in Q and NBF NMR ovaries (RNA-seq data) was plotted (x-  
295 axis: gene name, y-axis: expression level in  $\log_2$  scale). **D)** Expression of *Cyp19a1* (*Aromatase*) in Q and  
296 NBF NMR ovaries (x-axis, sample name; y-axis, expression level in counts per million (CPM)).

297 **Supplementary Figure 12: Gene expression profile of BMs and NBMs testis, and enriched terms for**  
298 **DEGs. A)** Cluster heatmap of testis samples. Sample distance was calculated euclidean distance matrix  
299 computation and cluster agglomeration was done using ward.D2 method. Heatmap colors indicate the  
300 euclidean distance between samples shown in the heatmap color key. **B)** Volcano plot showing the  
301 significance versus fold-change. The log fold change in expression is indicated on the x-axis and  
302 significance in  $-\log_{10}(p\text{-value})$  is indicated on the y-axis. DEGs are indicated by the red color. **C)** Network  
303 of enriched terms colored by cluster ID, indicating the relationship between the different enriched terms  
304 (*Figure 4F*). Nodes that share the same cluster are close to each other. Each circle node represents a term  
305 and the size of the circle is proportional to the number of genes that fall into that term, and the identity of  
306 the cluster is indicated by its color (nodes of the same color belong to the same cluster). Similar terms are  
307 linked by an edge (the thickness of the edge represents the similarity score). One term from each cluster is  
308 selected to have its term description shown as label. **D)** Enriched biological processes (generated using  
309 Enrichr). The length of the bar represents the significance of that specific gene-set or term, and the color  
310 intensity provides additional information about the significance (the brighter the color, the more

311 significant that term is). Statistical information used to generate the graph including the p-value and other  
312 enriched terms are provided in *Additional file 31A*. **E)** Enriched molecular functions (generated using  
313 Enrichr). The length of the bar represents the significance of that specific gene-set or term, and the color  
314 intensity provides additional information about the significance (the brighter the color, the more  
315 significant that term is). Statistical information used to generate the graph including the p-value and other  
316 enriched terms are provided in *Additional file 32A*.

317

318 **Supplementary Figure 13: Gene enrichment for DEGs that also show lower expression in BM**  
319 **testes. A)** Heatmap of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark Gene  
320 Sets, KEGG Pathway) for DEGs that show lower expression in BMs (generated using Metascape). The  
321 color code on the y-axis is used to show the clustering and relation of enriched networks (*Supplementary*  
322 *Figure 13B*). Detailed list of enriched terms and genes that belong to these terms is provided in *Additional*  
323 *file 30B*. **B)** Network of enriched terms found in above colored by cluster ID, indicating the relationship  
324 between the different enriched terms (*Supplementary Figure 13A*). Gene enrichment was generated using  
325 DEGs that show higher expression in BMs compared to NBMs. Nodes that share the same cluster are  
326 close to each other. Nodes that share the same cluster are close to each other. Each term is represented by  
327 a circle node, where its size is proportional to the number of input genes fall into that term, and its color  
328 represent its cluster identity (i.e., nodes of the same color belong to the same cluster). Similar terms are  
329 linked by an edge (the thickness of the edge represents the similarity score). One term from each cluster is  
330 selected to have its term description shown as label. **C)** Enriched biological processes (generated using  
331 Enrichr). The length of the bar represents the significance of that specific gene-set or term, and the color  
332 intensity provides additional information about the significance (the brighter the color, the more  
333 significant that term is). Statistical information used to generate the graph including the p-value and other  
334 enriched terms are provided in *Additional file 31B*. **D)** Enriched molecular functions (generated using  
335 Enrichr). The length of the bar represents the significance of that specific gene-set or term, and the color  
336 intensity provides additional information about the significance (the brighter the color, the more

337 significant that term is). Statistical information used to generate the graph including the p-value and other  
338 enriched terms are provided in *Additional file 32B*.

339

340 **Supplementary Figure 14: Gene enrichment for DEGs that also show higher expression in breeding**  
341 **male testes. A)** Heatmap of enriched terms (Canonical Pathways, GO Biological Processes, Hallmark  
342 Gene Sets, KEGG Pathway) for DEGs that show higher expression in BMs (generated using Metascape).  
343 Enriched terms significance is colored by p-values. The color code on the y-axis is used to show the  
344 clustering and relation of enriched networks (*Supplementary Figure 14B*). Detailed list of enriched terms  
345 and genes that belong to these terms is provided in *Additional file 30C*. **B)** Network of enriched terms  
346 found in above colored by cluster ID, indicating the relationship between the different enriched terms  
347 (*Supplementary Figure 14A*). Gene enrichment was generated using genes that show higher expression in  
348 breeding males compared to NBMs. Nodes that share the same cluster are close to each other. Nodes that  
349 share the same cluster are close to each other. Each term is represented by a circle node, where its size is  
350 proportional to the number of input genes fall into that term, and its color represent its cluster identity  
351 (i.e., nodes of the same color belong to the same cluster). Similar terms are linked by an edge (the  
352 thickness of the edge represents the similarity score). One term from each cluster is selected to have its  
353 term description shown as label. **C)** Enriched biological processes. The length of the bar represents the  
354 significance of that specific gene-set or term, and the color intensity provides additional information  
355 about the significance (the brighter the color, the more significant that term is). Statistical information  
356 used to generate the graph including the p-value and other enriched terms are provided in *Additional file*  
357 *31C*. **D)** Enriched molecular functions. The length of the bar represents the significance of that specific  
358 gene-set or term, and the color intensity provides additional information about the significance (the  
359 brighter the color, the more significant that term is). Statistical information used to generate the graph  
360 including the p-value and other enriched terms are provided in *Additional file 32C*.

361

362 **Supplementary Figure 15: NMR BM and NBM testis histology.** Representative sections through the  
363 testes of adult (**A,C**) breeding and (**B,D**) non-breeding male NMRs showing spermatogenesis; (**A**) and  
364 (**B**) are higher magnification showing examples of the cell types at different stages:  $S_A$  spermatogonia  
365 Type A,  $S_B$  spermatogonia Type B,  $S_1$  primary spermatocyte,  $S_3$  spermatid,  $S_4$  spermatocyte,  $S_t$  Sertoli  
366 cell ( $S_2$  secondary spermatocytes not identified). Sections (**C**) and (**D**) are lower magnification (from the  
367 respective individuals) and show the marked differences in the relative amount of interstitial tissue  
368 (mainly Leydig cells) to seminiferous tubules (bounded by the dotted lines).

369

370 **Supplementary Figure 16: A)** Potentially dopaminergic neuron cell groups (A8-A16), expressing the  
371 enzyme tyrosine hydroxylase but not the dopamine beta-hydroxylase, with their principal projections in  
372 the adult rodent brain (adapted from (45)). **B)** and **C)** A model describing the possible pathways though  
373 which a dopamine could regulate reproductive division of labor; **B)** in breeding animals (Qs and BMs),  
374 increased dopamine production in the hypothalamus suppresses the production of prolactin in from the  
375 pituitary. In the absence (or low amounts) of prolactin, GnRH is released from the hypothalamus, which  
376 acts on anterior pituitary to facilitate the release of FSH and LH. FSH and LH act on the gonads of  
377 breeding animals (Qs and BMs), to bring about normal gonadal development, gametogenesis and the  
378 release of sex hormones (estrogen and testosterone). **C)** in non-breeding naked mole-rats animals, lower  
379 levels of dopamine release result in increased production of prolactin by the anterior pituitary, which in  
380 turn suppresses the release of GnRH from the hypothalamus. In the absence of normal GnRH secretion,  
381 the anterior pituitary produces low levels of FSH and LH, resulting in suppression of normal gonadal  
382 development, gametogenesis and the release of sex hormones. **D)** A model showing the possible arrest  
383 point for NBF NMRs (indicated by the color graduation in the horizontal bar). The Q ovary can complete  
384 all the stages of oogenesis, however NBF fail at the stage where they produce estrogen (and are pre-  
385 pubertal in appearance). **E)** A model showing the possible arrest point where defects in spermatogenesis  
386 occur in non-breeding NMRs (indicated by the color graduation in the horizontal bar). Breeding males

387 have the ability to complete all stages of spermatogenesis; however, non-breeding animals fail at post-

388 meiotic stages and have further deficiencies in sperm maturation.

## Additional file legends

### Additional file 1: Animals/samples used and details of their subsequent analysis.

Tabs in the table indicate:

**A) Sacrificed animals (NMRs) for the experiments:** Table headers indicate animal/sample number in colony, social group (colony) where the animal belonged, year of birth, age (in years), body weight (in gram), whether used for sequencing (RNA-seq) or Immunofluorescence (IF), and the name used in the manuscript.

**B) RNA-seq information.** Table headers indicate: Tissue, name used to identify sample in manuscript (Name in manuscript), RNA-seq sample number, total sequence obtained (number of reads), length of each read in bp (sequence length), together with percentage GC content of total number of reads.

### Additional file 2: DEGs between brains of Qs and NBFs.

Tabs in the table indicate:

**All DEGs Qs vs NBFs:** all DEGs between brains of Qs and NBFs.

**DEG up in Q (Qs vs NBFs):** DEGs with higher expression in Qs brains.

**DEG down in Qs (Qs vs NBFs):** DEGs with lower expression in Qs brains.

In all tabs the table headers indicate: the naked mole rat gene id from our transcript assembly (NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); p-value (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

### Additional file 3: Result of enrichment analysis on DEGs between brains of Qs and NBFs using Metascape.

The GroupID, ID for clustering of related terms, the enriched terms (Term), the description of the enriched terms (Description), the p-value in log10 scale (p-values calculated based on

accumulative hypergeometric distribution); q value in log10 scale (q-values are calculated

using the Benjamini-Hochberg procedure to account for multiple testing), list of Symbols of upload hits in this term (Symbols).

**Additional file 4:** Result of enrichment analysis on DEGs between brains of Qs and NBFs using Enrichr (GO Biological Process).

Enrichment analysis was performed as described in material and methods. The table headers indicate: Term; enriched GO Biological Process; p-value, computed by Fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 5:** Result of enrichment analysis on DEGs between brains of Qs and NBFs using Enrichr (GO molecular functions).

The table headers indicate: Term; enriched GO molecular functions; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 6:** DEGs between brains of BMs and NBMs.

Tabs in the table indicate:

**All DEGs (BMs vs NBMs) tab:** all DEGs between brains of BMs vs NBMs.

**DEG\_up\_in\_BMs (BMs vs NBMs) tab:** DEGs with higher expression in BMs brains.

**DEG\_down\_in\_BMs (BMs vs NBMs) tab:** DEGs with lower expression in BMs brains.

In all cases the table headers indicate: the naked mole rat gene id from our transcript assembly

(NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); pvalue (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

**Additional file 7:** Result of enrichment analysis on DEGs between brains of BMs and NBMs using Metascape.

The table headers indicate: the GroupID, ID for clustering of related terms, the enriched terms (Term), the description of the enriched terms (Description), the p-value in log10 scale (p-values calculated based on accumulative hypergeometric distribution), q-value in log10 scale (q-values are calculated using the Benjamini-Hochberg procedure to account for multiple testing), list of Symbols of upload hits in this term (Symbols).

**Additional file 8:** Result of enrichment analysis on DEGs between brains of BMs and NBMs using Enrichr (GO Biological Process).

Enrichment analysis was performed as described in material and methods. The table headers indicate: Term; enriched GO Biological Process; p-value, computed by fisher's exact test; adjusted P-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 9:** Result of enrichment analysis on DEGs between brains of BMs and NBMs using Enrichr (GO molecular functions).

The table headers indicate: Term; enriched GO molecular functions; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg

method for correction for multiple hypotheses testing, the rank score or (z-score), computed

using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 10:** DEGs between brains of Qs and BMs.

Tabs in the table indicate:

**All\_DEGs (Qs\_vs\_BMs) tab:** all DEGs between brains of Qs and BMs.

**DEG\_and\_up\_in\_Q (Qs\_vs\_BMs) tab:** DEGs with higher expression in Qs brain.

**DEG\_and\_down\_in\_Q (Qs\_vs\_BMs) tab:** DEGs with lower expression in Qs brain.

In all cases, the table headers indicate: the naked mole rat gene id from our transcript assembly (NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); p-value (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

**Additional file 11:** Result of enrichment analysis on DEGs between brains of Qs and BMs using Metascape.

The table headers indicate: the GroupID, ID for clustering of related terms, the enriched terms (Term), the description of the enriched terms (Description), the p-value in log10 scale (p-values calculated based on accumulative hypergeometric distribution), q-value in log10 scale (q-values are calculated using the Benjamini-Hochberg procedure to account for multiple testing), list of Symbols of upload hits in this term (Symbols).

**Additional file 12:** Result of enrichment analysis on DEGs between brains of Qs and BMs using Enrichr (GO Biological Process).

The table headers indicate: Term, enriched GO Biological Process; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 13:** Result of enrichment analysis on DEGs between brains of Qs and BMs using Enrichr (GO molecular functions).

The table headers indicate: Term; enriched GO molecular functions; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 14:** DEGs between brains of NBFs and NBMs.

Tabs in the table indicate:

**All\_DEGs (NBFs vs NBMs) tab:** all DEGs brains of NBFs and NBMs.

**DEGs up in NBF (NBFs vs NBMs) tab:** DEGs with higher expression in NBFs brain.

**DEGs down in NBF (NBFs vs NBMs) tab:** DEGs with lower expression in NBFs brain.

In all tabs the table headers indicate: the naked mole rat gene id from our transcript assembly (NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); p-value (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

**Additional file 15:** DEGs between brains of Qs and NBMs.

Tabs in the table indicate:

**All\_DEGs (Qs\_vs\_NBMs) tab:** all DEGs between brains of Qs and NBMs.

**DEG\_up\_in\_Q (Qs\_vs\_NBMs) tab:** DEGs with higher expression in Qs brain.

**DEG\_down\_in\_Q (Qs\_vs\_NBMs) tab:** DEGs with lower expression in Qs brain.

In all cases the table headers indicate: the naked mole rat gene id from our transcript assembly (NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); p-value (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

**Additional file 16:** Defining queen specific genes (Q genes). The list of DEGs in the comparison between Qs vs NBFs, Qs vs NBMs, Qs vs BMs were mapped to each other to identify queen specific genes.

The different tables in the excel sheet are:

**Q vs NBF NBM BM all tab:** DEGs between Qs vs NBFs, Qs vs NBMs and Qs vs BMs.

**Q vs NBF NBM all tab:** all DEGs between Qs vs NBFs and Qs vs NBMs.

**Q vs NBF NBM BM down in Q tab:** genes that are differentially expressed in the comparison (Qs vs NBFs, Qs vs NBMs, and Qs vs BMs) and also show lower expression in Q (down regulation in Q)

**Q vs NBF NBM down in Q tab:** genes that are differentially expressed in the comparison (Qs vs NBFs, Qs vs NBMs) and also show lower expression in Q (down regulation in Q)

**Q vs NBF NBM BM up in Q tab:** genes that are differentially expressed in the comparison (Qs vs NBFs, Qs vs NBMs, and Qs vs BMs) and also show higher expression in Q (up regulation in Q)

**Q vs NBF NBM up in Q tab:** genes that are differentially expressed in the comparison (Qs vs NBFs, Qs vs NBMs) and also show higher expression in Q (up regulation in Q)

**Additional file 17:** Result of enrichment analysis of Q genes using Metascape.

The table headers indicate: GroupID, ID for clustering of related terms, the enriched terms (Term), the description of the enriched terms (Description), the p-value in log10 scale (p-values calculated based on accumulative hypergeometric distribution), q value in log10 scale (q-values are calculated using the Benjamini-Hochberg procedure to account for multiple testing), list of Symbols of upload hits in this term (Symbols).

**Additional file 18:** Result of enrichment analysis of Q genes using Enrichr (GO Biological Process).

The table headers indicate: the Term; enriched GO Biological Process; P-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 19:** Result of enrichment analysis of Q genes using Enrichr (GO molecular functions).

The table headers indicate: the Term; enriched GO molecular functions; p-value, computed by fisher's exact test; adjusted P-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 20:** DEGs between brains of BMs and NBFs.

Tabs in the table indicate:

**All DEGs BMs vs NBFs tab:** all DEGs brains of BMs and NBFs.

**DEGs Up in BM (BMs vs NBFs) tab:** DEGs with higher expression in BMs brain.

**DEGs down in BM (BMs vs NBFs) tab:** DEGs with lower expression in BMs brain.

In all cases the table headers indicate: the naked mole rat gene id from our transcript assembly (NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); p-value (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

**Additional file 21:** Defining BM specific genes (BM genes). The list of DEGs in the comparison between BMs vs NBFs, BMs vs NBMs, were mapped to each other to identify breeding male specific genes.

**Additional file 22:** Result of enrichment analysis of BM genes using Metascape.

Enrichment analysis was performed as described in material and methods. The output of the enrichment analysis is presented here. The table headers indicate: The GroupID, ID for clustering of related terms, the enriched terms (Term), the description of the enriched terms (Description), the p-value in log10 scale (p-values calculated based on accumulative hypergeometric distribution), q value in log10 scale (q-values are calculated using the Benjamini-Hochberg procedure to account for multiple testing), list of Symbols of upload hits in this term (Symbols).

**Additional file 23:** Result of enrichment analysis of BM genes using Enrichr (GO Biological Process).

The table headers indicate: the Term; enriched GO Biological Process; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from

multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 24:** Result of enrichment analysis of BM genes using Enrichr (GO molecular functions).

The table headers indicate: the Term; enriched GO molecular functions; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

**Additional file 25:** DEGs between Qs and NBFs ovary.

Tabs in the table indicate:

***All\_DEGs (Qs\_vs\_NBFs) tab:*** all DEGs between ovaries of Qs and NBFs.

***DEG\_and\_down\_in\_Q (Qs\_vs\_NBFs) tab:*** DEGs with lower expression in Qs ovary.

***DEG\_and\_up\_in\_Q (Qs\_vs\_NBFs) tab:*** DEGs with higher expression in Qs ovary.

In all tabs the table headers indicate: the naked mole rat gene id from our transcript assembly (NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); p-value (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

**Additional file 26:** Result of enrichment analysis on DEGs between Qs and NBFs ovary using Metascape. The table headers indicate: The GroupID, ID for clustering of related terms, the enriched terms (Term), the description of the enriched terms (Description), the p-value in log10 scale (p-values calculated based on accumulative hypergeometric distribution), q value

in log10 scale (q-values are calculated using the Benjamini-Hochberg procedure to account

for multiple testing), list of symbols of upload hits in this term (Symbols).

The tabs in the table show:

**A) all DEGs:** enriched pathways, Biological Process, and molecular functions for all DEGs between Qs and NBFs ovary.

**B) DEGs up in Q:** enriched pathways, Biological Process, and molecular functions for DEGs between Qs and NBFs ovary, and up regulated in Q compared to NBF.

**C) DEGs down in Q:** enriched pathways, Biological Process, and molecular functions for DEGs between Qs and NBFs ovary, and down regulated in Q compared to NBF.

**Additional file 27:** Result of enrichment analysis on DEGs between Qs and NBFs ovary using Enrichr (GO Biological Process).

The table headers indicate: the Term; enriched GO Biological Process; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

The tabs in the table show:

**A) All DEGs:** enriched Biological Process for all DEGs between Qs and NBFs ovary.

**B) DEGs down in Q:** enriched Biological Process for DEGs between Q and NBFs ovary, and down regulated in Qs compared to NBFs.

**C) DEGs up in Q:** enriched Biological Process for DEGs between Qs and NBFs ovary, and up regulated in Qs compared to NBFs.

**Additional file 28:** Result of enrichment analysis on DEGs between Qs and NBFs ovary using Enrichr (GO molecular functions).

The table headers indicate: the Term, enriched GO molecular functions; p-value, computed

by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

The tabs in the table show:

- A) All DEGs:** enriched molecular functions for all DEGs between Qs and NBFs ovary.
- B) DEGs down in Qs:** enriched molecular functions for DEGs between Qs and NBFs ovary, and down regulated in Qs compared to NBFs.
- C) DEGs up in Qs:** enriched Biological Process for DEGs between Qs and NBFs ovary, and up regulated in Qs compared to NBFs.

**Additional file 29:** DEGs between BMs and NBMs testis.

Tabs in the table indicate:

**All DEGs (BMs vs NBMs) tab:** all DEGs between BMs and NBMs testis.

**DEG down in BM (BM vs NBMs) tab:** DEGs and higher expression in BMs testis.

**DEG up in BM (BM vs NBMs) tab:** DEGs and lower expression in BMs testis.

**Germonline genes:** Gene list obtained from Germonline ([germonline.org](http://germonline.org))

In all tabs the table headers indicate: the naked mole rat gene id from our transcript assembly (NMR\_geneID); the log fold change (logFC); counts per million in log scale (logCPM); pvalue (PValue); false discovery rate using Benjamini-Hochberg (FDR); annotation of NMR\_geneID to mouse Ensembl gene ID (Ensembl\_geneID); and annotation of NMR\_geneID to mouse gene name (gene\_name).

**Additional file 30:** Result of enrichment analysis on DEGs between BMs and NBMs testis using Metascape.

The table headers indicate: the GroupID, ID for clustering of related terms, the enriched terms

(Term), the description of the enriched terms (Description), the p-value in log10 scale (p-values calculated based on accumulative hypergeometric distribution), q-value in log10 scale (q-values are calculated using the Benjamini-Hochberg procedure to account for multiple testing), list of Symbols of upload hits in this term (Symbols).

The tabs in the table show:

**A) All DEGs:** enriched pathways, Biological Process, and molecular functions for all DEGs between BMs and NBM testis.

**B) DEGs down in\_BM:** enriched pathways, Biological Process, and molecular functions for DEGs between BMs and NBMs testis, and up regulated in BMs compared to NBMs.

**C) DEGs up in BM:** enriched pathways, Biological Process, and molecular functions for DEGs between BMs and NBMs testis, and up regulated in BMs compared to NBMs.

**D) reprod &steroid respon:** selected genes that belong to terms, multicellular organism reproduction, response to steroid hormone, multicellular organism reproduction and response to steroid hormone.

**Additional file 31:** Result of enrichment analysis on differentially expressed between BMs and NBMs testis using Enrichr (GO Biological Process).

The table headers indicate: Term; enriched GO Biological Process; p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

The tabs in the table show:

**A) All\_deg:** GO Biological Process enrichment for all DEGs between BMs and NBMs testis.

**B) DEGs up in BM:** GO Biological Process enrichment for DEGs between BMs and NBMs testis, and up regulated in BMs compared to NBMs.

**C) DEGs down in BM:** GO Biological Process enrichment for DEGs between BMs and

NBMs testis, and down regulated in BMs compared to NBMs.

**Additional file 32:** Result of enrichment analysis on DEGs between BMs and NBMs testis using Enrichr (GO molecular functions). The table headers indicate: Term; enriched GO molecular functions; p-value, p-value, computed by fisher's exact test; adjusted p-value, an adjusted p-value using the Benjamini-Hochberg method for correction for multiple hypotheses testing; the rank score or (z-score), computed using a modification to Fisher's exact test in which a z-score was computed for deviation from an expected rank; combined score, a combination of the p-value and z-score calculated by multiplying the two scores ( $c = \log(p) * z$ ); Genes, belonging to the enriched category.

The tabs in the table show:

**A) All DEG:** GO molecular functions enrichment for DEGs between BMs and NBMs testis.

**B) DEG down in BM:** GO molecular functions enrichment for DEGs between BMs and NBMs testis, and down regulated in BMs compared to NBMs.

**C) DEG up in BM:** GO molecular functions enrichment for DEGs between BMs and NBMs testis, and up regulated in BMs compared to NBMs.