1 **Frequency of mosaicism points towards mutation-prone early cleavage cell**
2 **divisions.**

3 *As much as 30% and 50% of de novo mutations (DNM) may occur during the early*
4 *cleavage cell divisions in males and females, respectively, causing frequent*
5 *mosaicism and a high sibling recurrence risk of DNM-dependent diseases.*

6 *Chad Harland[1#], Carole Charlier[1#], Latifa Karim[2], Nadine Cambisano[2], Manon*
7 *Deckers[2], Erik Mullaart[3], Wouter Coppieters[2], Michel Georges[1].*

8 [1] Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University
9 of Liège, Belgium. [2] GIGA Genomics Platform, University of Liège, Belgium. [3] CRV,
10 Arnhem, The Netherlands.

11 [#] *Contributed equally to this work*

12 Correspondence: michel.georges@ulg.ac.be

13

14 **It has recently become possible to directly estimate the germ-line de novo**
15 **mutation (DNM) rate by sequencing the whole genome of father-mother-**
16 **offspring trios, and this has been conducted in human[1-5], chimpanzee[6],**
17 **birds[7] and fish[8]. In these studies DNMs are defined as variants that are**
18 **heterozygous in the offspring while being absent in both parents. They are**
19 **assumed to have occurred in the germ-line of a parent and to have been**
20 **transmitted to the offspring via the sperm or oocyte. This definition**
21 **assumes that detectable mosaïcism in the individual in which the mutation**
22 **occurred is negligible. However, instances of mosaïcism are well-**

23 **documented in humans and other organisms, including ruminants[9,10]. We**
24 **herein take advantage of the unique pedigree structure of cattle to show**
25 **that mosaïcism associated with DNMs is a common occurrence, and that**
26 **this should be taken into account to accurately estimate the mutation rate**
27 **in this and possibly other species. It suggests that early cleavage cell**
28 **divisions are particularly mutation-prone, and that the recurrence risk of**
29 **DNM-dependent disorders in sibs may be higher than generally assumed.**

30 To study the process of DNMs in the cattle germ-line, we sequenced the whole
31 genome of 46 members of three four-generation pedigrees (Figure 1). The
32 source of the sequenced DNA was venous blood for females and sperm for males.
33 Grand-parents, parents and offspring (referred to as probands) were sequenced
34 at average 26-fold depth (min = 23), and grand-offspring at average 21-fold
35 depth (min = 10).

36 We identified 151 candidate DNMs as variants that were (i) detected in a
37 proband, (ii) absent in both parents (and grand-parents when available), (iii)
38 transmitted to at least one grand-offspring, and (iv) not previously reported in
39 unrelated individuals from the 1,000 Bulls project[11] (Suppl. Table 1 and Suppl.
40 Figure 1). We developed amplicons spanning 94 candidate DNMs and sequenced
41 them at average depth of ~1,250 in the 46 animals plus 57 relatives (Figure 1).
42 This confirmed the genuine nature of 91/94 variants, demonstrating the
43 excellent specificity of our pipeline. The three remaining ones were also
44 detected in one of the parents (although not in the grand-parents) in the
45 confirmation, and momentarily ignored.

46 We examined what proportion of DNMs detected in a proband occurred during
47 its development rather than inherited via the sperm or oocyte. An unambiguous
48 distinction between the two types of mutations is that the former will, upon
49 transmission to the next generation, show partial linkage with one of the
50 homologues (i.e. never transmitted with homologue A, sometimes transmitted
51 with homologue B), while the latter will show complete linkage (i.e. never
52 transmitted with homologue A, always transmitted with homologue B). We first
53 examined the pedigree with 11 grand-offspring as it provided the best
54 opportunity to distinguish partial from complete linkage (Suppl. Fig. 2). 21
55 variants showed complete linkage with the paternal homologue, 13 complete
56 linkage with the maternal homologue, and 16 partial linkage with either the
57 paternal or maternal homologue. Thus 21 DNMs appeared to have been
58 inherited from the father, 13 from the mother, while 16 would have occurred
59 early enough during the development of the proband to generate detectable
60 levels of mosaïcism in sperm DNA. Three additional predictions can be made if
61 these 16 mutations occurred during the development of the proband. (i) They
62 should a priori have equal chance to affect the paternal and maternal
63 homologues. Five of the 16 variants affected the paternal and 11 the maternal
64 homologue ($p=0.21$). (ii) The allelic ratio should be inferior to 50% in the
65 proband but equal to 50% in the grand-offspring. The mean allelic ratio was
66 0.28 in the proband, and 0.50 in the grand-offspring, and this difference was
67 highly significant ($p < 10^{-6}$). The corresponding means were 0.47 and 0.48 ($p =$
68 0.40) for the 34 mutations showing complete linkage with either the paternal or
69 maternal homologue (Figure 2, Suppl. Table 1). (iii) The proportion of grand-
70 offspring inheriting the mutation and the strength of linkage should increase

71 with the allelic ratio in the proband, and this was indeed the case ($p_{transmission}$ =
72 0.02; $p_{linkage}$ = 0.008) (Suppl. Fig. 3). We conclude that, in this pedigree, at least
73 32% of the DNMs detected in bulk sperm DNA of the proband using standard
74 criteria did not occur in the germ-line of either the sire or dam but rather during
75 the development of the proband. We will refer to this type of DNMs as Proband-
76 Mosaic (PM), while the DNMs that are absent in the sire and showing complete
77 linkage with the paternal homologue in the grand-offspring will be referred to as
78 Sperm-Non-Mosaic (SNM) (meaning that the sire is not mosaic for the mutation
79 transmitted via his sperm), and those that are absent in the dam yet showing
80 complete linkage with the maternal homologue as Oocyte-Non-Mosaic (ONM)
81 (meaning that the dam is not mosaic for the mutation transmitted via her oocyte).

82 One could argue that the observed high level of mosaicism is due to the fact that
83 we analyzed sperm DNA rather than somatic DNA. We therefore analyzed the
84 three female probands (blood DNA), using the same approach. It is noteworthy
85 that with only five grand-offspring, the probability to detect a mosaic DNM and
86 demonstrate incomplete linkage is reduced (Suppl. Fig. 2). We detected 72
87 mutations transmitted in complete linkage with the paternal homologue, 14 in
88 complete linkage with the maternal homologue, and 12 in partial linkage with
89 either of these. Three of 12 were partially linked to the paternal and nine to the
90 maternal homologue (p=0.14). Their allelic ratio in the probands was 0.29, while
91 being 0.50 in the grand-offspring ($p < 10^{-6}$). The corresponding means were 0.50
92 and 0.49 (p = 0.66) for the 86 mutations showing complete linkage with either
93 the paternal or maternal homologue (Figure 2, Suppl. Table 1). Thus at least
94 12% of DNMs detected in blood DNA of female probands using standard

95 procedures occurred during their development rather than being inherited from
96 the parents, hence being of "PM" type. We did not observe a positive correlation
97 between the level of somatic mosaicism in the female probands and the
98 proportion of grand-offspring inheriting the mutation or the strength of linkage
99 in the grand-offspring ($p_{transmission}$ = 0.85; $p_{linkage}$ = 0.94), suggesting that the
100 degree of mosaicism in the female soma is a poor indicator of the degree of
101 mosaicism in the germ line (Suppl. Fig. 3).

102 If detectable mosaicism for DNMs is common, requiring their absence from the
103 parental DNA (as typically done) will eliminate genuine DNMs. We attempted to
104 recover such events as variants that were (i) absent in the grant-parents, (ii)
105 detected in either sire or dam with a fraction of mutant reads significantly < 50%
106 (Suppl. Table 1), (iii) transmitted to the proband with an allelic ratio of ~50%,
107 (iv) transmitted to at least one grand-offspring with an allelic ratio of ~50%, and
108 (v) not previously reported in unrelated individuals[11]. We detected 61
109 candidates, including the 3/89 variants mentioned above (Suppl. Table 1 and
110 Suppl. Fig. 1). We developed amplicons for 37, and sequenced (average 1,083-
111 fold depth) all 46 individuals plus 57 relatives including ≥ five half-sibs for each
112 proband (Figure 1). For 11 of the 37 tested candidates, the sire's or dam's allelic
113 ratio in the confirmation was ~1:1, and linkage to either the paternal or maternal
114 homologue (in the proband's half-sibs) perfect. These 11 events were thus
115 genuine DNMs, yet were more likely to have occurred in the grand-parents than
116 in the parents. For 25/26 remaining candidates, we observed partial linkage in
117 the half-sibs of the corresponding probands, confirming that these DNMs
118 occurred during the development of the (hence mosaic) parent. For the last

119 1/26, although linkage in the half-sibs of the proband appeared complete, the
120 allelic ratio in the parent remained very significantly skewed when compared to
121 descendants ($p < 10^{-5}$), strongly suggesting that this mutation also occurred
122 during the development of the (hence mosaic) parent. For 20/26 the DNMs were
123 transmitted to at least one of the analyzed half-sibs (Suppl. Table 1 and Suppl. Fig.
124 1). Following this targeted confirmation, we took advantage of whole genome
125 sequence information that became available for 17 of the probands' half-sibs, to
126 trace the inheritance of the 24 non-confirmed candidate variants. We
127 demonstrated partial linkage for 23/24 variants, hence confirming the
128 mosaicism of the corresponding parent. For 15/24 the DNMs were transmitted
129 to at least one of the analyzed half-sibs (Suppl. Table 1 and Suppl. Fig. 1). Taken
130 together, confirmation data provided (i) strong evidence against parental
131 mosaicism for 11/61 variants, (ii) strong evidence in favor of parental mosaicism
132 for 49/61, and (iii) unconfirmed evidence for parental mosaicism for 1/61.
133 Twenty-seven of the 50 (49+1) non-excluded variants were detected in a sire,
134 and 23 in a dam (despite the fact that the sire could not be studied in the 11-
135 grand-offspring pedigree). Twenty-three occurred on the paternal and 27 on the
136 maternal chromosome ($p = 0.34$). The allelic ratio in the parents was 0.13, while
137 being 0.50 in the proband and grand-offspring ($p < 10^{-6}$) (Figure 2, Suppl. Table
138 1). We did observe a significant positive correlation between the level of
139 somatic mosaicism in the sire and the strength of linkage in half-sibs ($p_{transmission}$
140 $= 0.002$), and between the level of somatic mosaicism in the dam and the
141 proportion of half-sibs inheriting the mutation ($p_{linkage} = 0.0008$) (Suppl. Fig. 3).
142 These data suggest that – in cattle – paternal bulk sperm DNA may be mosaic for
143 at least 24% of DNMs present in a sperm cell, while maternal blood DNA may be

144 mosaic for at least 51% of DNMs present in an oocyte. We will refer to these
145 types of mutations as Sperm-Mosaic (SM) and Oocyte-Mosaic (OM), respectively
146 (meaning that the sire/dam is mosaic for the mutation transmitted via the
147 sperm/oocyte).

148 When analyzing the transmission patterns of SM and OM mutations to the half-
149 sibs of the probands, we were struck by the fact that (i) very few half-sibs
150 inherited none of the DNMs detected in the proband, while more than 50%
151 would be expected (23/60; p = 0.05), and (ii) half-sibs sharing multiple DNMs
152 with the proband appeared surprisingly common (Suppl. Fig. 4).

153 In mammals, after fertilization, cleavage, and segregation of (i) the inner cell
154 mass from the trophoblast, (ii) the epiblast from the hypoblast, (iii) the
155 embryonic epiblast from the amniotic ectoderm, a small number of epiblast-
156 derived cells located in the wall of the yolk sac in the vicinity of the allantois are
157 induced to become primordial germ cells (PGC). These migrate to the primitive
158 gonad where they expand and produce >1 million gametogonia. Oogonia initiate
159 meiosis prior to birth in females. Spermatogonia will resume mitotic divisions at
160 puberty allowing (i) the maintenance of a pool of stem cell like spermatogonia,
161 and (ii) sustained spermiogenesis involving ~2 additional mitotic divisions
162 followed by meiosis (Suppl. Fig. 5). We simulated the process of de novo
163 mutagenesis in the male and female germ cell lineages assuming (i) uniform pre-
164 and post-natal mutation rates per cell divisions, and (ii) 40 unrelated PGCs (i.e.
165 sampled at random from the epiblast). Pre- and post-natal mutation rates were
166 adjusted to match the observed number of mutations per gamete (31 in sperm,
167 14 in oocytes). Under these conditions, we virtually never observed the level of

168 mosaicism, nor the sharing between sibs characterizing the real data (Figure 3).
169 We (i) increased the relative mutation rate during the early cell divisions
170 (keeping the mutation rate per gamete constant)(10 and 20-fold increase during
171 the first 2, 4 or 6 cell divisions), (ii) reduced the number of induced PGCs (40 or
172 4), and (iii) varied the relatedness between PGCs (unrelated or related)(Suppl.
173 Fig. 5). Increasing the mutation rate during the very first cell divisions, and
174 increasing the relatedness between induced PGCs in the simulations matched the
175 real data much better (Figure 3). To quantitatively evaluate model fitting we
176 used (i) the proportion of PM, SM and OM mutations with corresponding rate of
177 mosaicism in sperm and soma, and (ii) the proportion of sibs sharing 0, 1, 2, ...
178 DNMs with a proband, to compute the likelihood of the data under different
179 scenarios (see M&M). The data were at least $10^{5.8}$ times more likely when
180 assuming an increased mutation rate during the very first cell divisions than not
181 (whichever the values of the other parameters). Assuming a higher mutation
182 rate during the first cell divisions, the best model with 40 related PGCs was $10^{6.1}$
183 times more likely than the best model with 40 unrelated PGCs, and $10^{18.3}$ times
184 more likely than with 4 PGCs. The data were $10^{3.2}$ more likely when the
185 mutation rate was increased during the 4 or 6 first cell divisions rather than only
186 the first two cell divisions (Table 1 and Supplemental Table 2).

187 We estimated the numbers of the five types of mutations for the four studied
188 probands, accounting for estimated genome coverage (Figure 4A). The
189 estimated number of DNMs per gamete (SNM+SM, ONM+OM) averaged 44.5 for
190 sperm cells and 18.1 for oocytes (male/female ratio of 2.5), corresponding to an
191 average mutation rate of $\sim 1.25 \times 10^{-8}$ per base pair per gamete. These rates are

192 likely to be slightly underestimated as a number (~ 1-5) of SM and OM mutations
193 (with highest degree of mosaicism in the parents) would have remained
194 undetected. The standard approach of ascertaining DNMs (i.e. erroneously
195 considering PM mutations, ignoring SM and OM mutations) would have yielded a
196 mutation rate of 0.97×10^{-8} per bp per gamete, with a 2.2-fold higher mutation
197 rate in bulls than in cows. 195 of the 209 identified DNMs were nucleotide
198 substitutions, 14 small insertion-deletions. Amongst substitutions, the transition
199 (Ti) – transversion (Tv) ratio was unexpectedly low (1.14). This was shown to
200 be due to a significant ~8-fold enrichment of C>A and/or G>T transversions in
201 the mosaic classes of mutations (PM, SM and OM). The non-mosaic classes of
202 mutations (SNM and ONM) were ~30-fold enriched in CpG>TpG transitions as
203 expected, while this signature was less pronounced for mosaic mutations (Figure
204 4B&C). There was no obvious difference between the profile of DNMs in the
205 male and female germ line (data not shown). DNMs were uniformly scattered
206 across the genome (Suppl. Fig. 6).

207 Estimates of the DNM rate per generation from sequencing human families are
208 ~2-fold lower than estimates from primate sequence divergence and possible
209 reasons for this discrepancy have been discussed[12,13]. We wondered whether
210 the rate of mosaicism might affect the rate of nucleotide substitution per
211 generation. We simulated the fixation of DNM per generation in populations of
212 varying effective population size, with constant mutation rate per gamete but
213 varying levels of mosaicism. There was no evidence for an effect of mosaicism on
214 fixation rate (which was always ~ μ) nor on fixation time (which was always
215 ~4Ne)(Suppl. Figure 7).

216 Our work strongly suggests that – in the bovine - early cleavage cell divisions are
217 particularly mutation prone, possibly accounting for as many as ~50% of DNM
218 detected in oocyte, and ~30% of DNM detected in a sperm cell. These findings
219 are consistent with recent evidence reported in humans[14]. The corresponding
220 early mutations were characterized by a distinct signature dominated by C>A
221 and/or G>T transversions, in contrast with DNM occurring during later stages of
222 gametogenesis which are dominated by the expected methyl-CpG to TpG
223 transversions. It is worth nothing that mutations in the proofreading domain of
224 polymerase epsilon, synthesizing the leading strand during DNA replication,
225 cause a mutator phenotype dominated by C>A in tumors[15]. This suggests that
226 the observed early DNM might primarily result from replication errors. The
227 predominance of early C>A transversions causes the overall Ti/Tv ratio to ~1.14,
228 well below expectations. This was unlikely to be artifactual, as the expected
229 Ti/Tv was obtained when applying the same bioinformatics pipeline on
230 simulated DNM in the same pedigrees (M&M). Paradoxically the Ti/Tv ratio is ~2
231 when considering DNA sequence polymorphisms segregating in the domestic
232 cattle population (Suppl. Figure 8). This suggests either sampling variation
233 (meaning that Ti/Tv ratios might differ between families and that we by chance
234 sampled families at the low end) or that the mechanisms underlying the
235 observed excess of C>A (or G>T) transitions emerged recently. It is worth noting
236 in this regard that most analyzed animals were bred by artificial insemination
237 and/or in vitro embryo production.

238 Our results support the notion that the population of induced PGCs is
239 ontogenetically related. This is not unexpected given their physical proximity at

240 the base of the allantois. This may - in combination with a high incidence of
241 early cleavage cell divisions – be medically extremely relevant. Indeed, if
242 applicable to human, it implies that the recurrence risk of DNM-dependent
243 disorders in sibs may be higher than generally assumed[9,13].

244 DNM occurring during the development of an individual, should a priori affect
245 the maternal and paternal chromosome with equal probability. Indeed, this null
246 hypothesis could not be disproved within specific individuals for the PM, OM and
247 SM type of mutations detected in this work. When repeating the test across
248 individuals, however, we note that 47 mosaic mutations occurred on the
249 maternal chromosome versus 29 on the paternal chromosome ($p = 0.05$). This
250 suggests that the maternal and paternal chromosomes might be epigenetically
251 distinct during early development and that this may affect their mutability.

252 Our work points towards the fact that direct estimates of mutation rates from
253 sequencing families may have to be revisited, taken PM, OM and SM status into
254 account, to obtain more accurate estimates of the mutation rate per gamete and
255 per generation which have been raising some questions[12,13]. This may affect
256 both the overall mutation rate as well as its male/female ratio. However, our
257 analyses suggest that the effect is likely to be modest and insufficient to explain
258 the present discrepancy between direct and indirect estimates in human studies.

259

260 **References**

261 1. Roach JC et al. Analysis of Genetic Inheritance in a Family Quartet by Whole-
262 Genome Sequencing. *Science* **328**: 636–639 (2010).

263 2. Conrad DF et al. Variation in genome-wide mutation rates within and between
264 human families. *Nat Genet* **43**: 712–714 (2011.).

265 3. Kong A et al. Rate of de novo mutations and the importance of father/'s age to
266 disease risk. *Nature* **488**: 471–475 (2012).

267 4. Campbell CD et al. Estimating the human mutation rate using autozygosity in a
268 founder population. *Nat Genet* **44**: 1277–1281 (2012).

269 5. Michaelson JJ et al. Whole-Genome Sequencing in Autism Identifies Hot Spots
270 for De Novo Germline Mutation. *Cell* **151**: 1431–1442 (2012).

271 6. Venn O et al. Strong male bias drives germline mutation in chimpanzees.
272 *Science* **344**:1272-1275 (2014).

273 7. Smeds L, Qvarnström A, Ellegren H Direct estimate of the rate of germline
274 mutation in a bird. *Genome Res.* **26**: 1211-1218 (2016).

275 8. Feng C et al. Moderate nucleotide diversity in the Atlantic herring is associated
276 with a low mutation rate. In preparation (2016)

277 9. Campbell IM et al. Somatic mosaicism: implications for disease and
278 transmission genetics. *Trends Genet.* **31**:382-392 (2015).

279 10. Smit M et al. Mosaicism of Solid Gold supports the causality of a noncoding A-
280 to-G transition in the determinism of the callipyge phenotype. *Genetics* **163**:453-
281 456 (2003).

282 11. Daetwyler HD et al. Whole-genome sequencing of 234 bulls facilitates
283 mapping of monogenic and complex traits in cattle. *Nat Genet* **46**:858-865
284 (2014).

285 12. Segurel L, Wyman MJ, Przeworrski M. Determinants of mutation rate
286 variation in the human germline. *Annu Rev Genomics Hum Genet* **15**:47-70 (2014).

287 13. Scally A. Mutation rates and the evolution of germline structure. Philos *Trans*
288 *R Soc Lond B Biol Sci* **371:** 20150137 (2016).

289 14. Rahbari R et al. Timing, rates and spectra of human germline mutation.
290 *Nature Genetics* **48:** 126-133 (2016).

291 15. Shinbrot E et aL Exonuclease mutations in DNA polymerase epsilon reveal
292 replication strand specific mutation patterns and human origins of replication.
293 *Genome Res* **24:** 1740-750 (2014).

294

295 **Acknowledgements**

296 This work was funded by the DAMONA Advanced ERC project to Michel Georges.
297 Carole Charlier is Senior Research Associate from the Fonds de la Recherche
298 Scientifique - FNRS (F.R.S.-FNRS). Chad Harland's PhD fellowship was funded by
299 Livestock Improvement Corporation (Hamilton, New Zealand). We thank Azim
300 Surani for fruitful discussions. We are grateful to Erik Mullaart and CRV
301 (Arnhem the Netherlands) for providing us with the sperm and DNA samples.
302 We used the supercomputing facilities of the Consortium des Equipements de
303 Calcul Intensif en Fédération Wallonie Bruxelles (CECI) funded by the F.R.S-FNRS.

304

305 **Authors contributions**

306 MG, CH, CC: designed the experiments. EM: provided samples. LK, NC, MD, WC:
307 performed the sequencing. CH, MG, CC: analyzed data. MG, CH, CC: wrote the
308 paper.

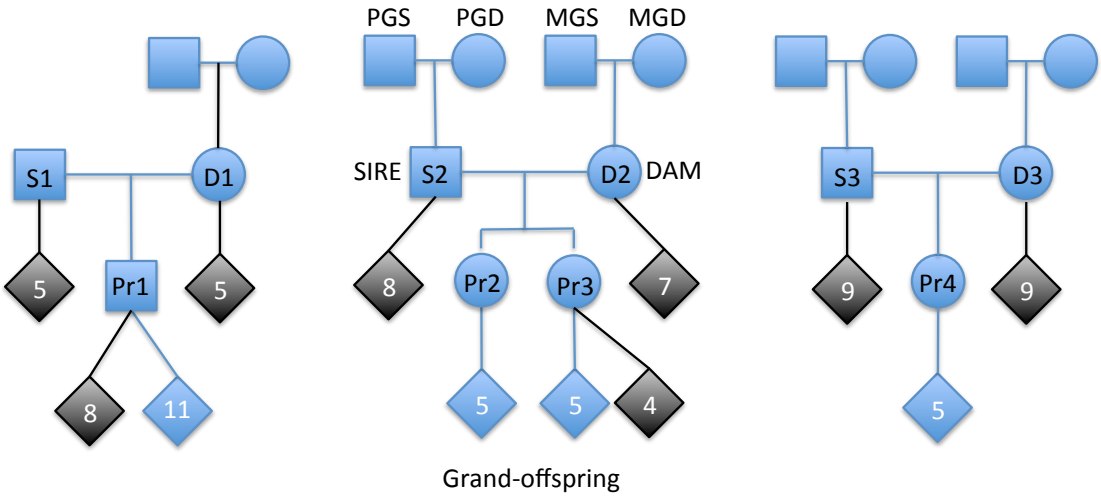
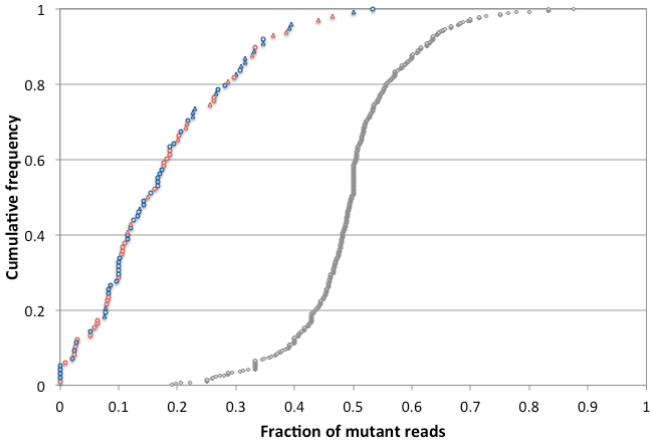
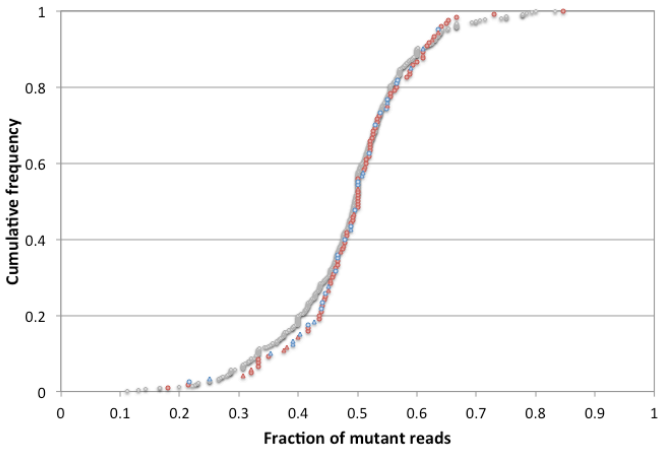


Figure 1: Pedigrees used for the detection of DNMs. PGS: paternal grand-sire, PGD, paternal grand-dam, MGS: maternal grand-sire, MGD: maternal grand-dam, Pr: probands. S: sires. D: dams. Animals in gray were used for confirmation.



314 Figure 2A



315 Figure 2B

316 **Figure 2: Cumulative frequency distribution of the proportion of mutant**
317 **reads. (A) DNMs with detectable mosaicism in the individual (proband or**
318 **parent) in which the mutation occurred:** Blue triangles: P(roband)M(osaic)
319 mutations having occurred during the development of the male proband (Pr1)
320 and for which the proband's sperm is mosaic. Red triangles: P(roband)M(osaic)
321 mutations having occurred during the development of the three female probands
322 (Pr2, Pr3, Pr4) and for which the proband's blood is mosaic. Blue circles:
323 S(perm)M(osaic) mutations having occurred during the development of two
324 sires (S2, S3) and for which the sires' sperm is mosaic. Red circles:
325 O(oocyte)M(osaic) mutations having occurred during the development of three
326 dams (D1, D2, D3) and for which the dams' blood is mosaic. Grey diamonds:
327 same mutations in the descendants of the mosaic animal. **(B) DNMs without**
328 **detectable mosaicism in the individual in which the mutation occurred:**
329 Blue triangles: S(perm)N(on)M(osaic) mutations transmitted by the (non
330 mosaic) sire to the male proband (Pr1) in whom the mutation was detected. Red
331 triangles: S(perm)N(on)M(osaic) mutations transmitted by the (non mosaic) sire
332 to the three female probands (Pr2, Pr3, Pr4) in whom the mutation was detected.
333 Blue circles: O(oocyte)N(on)M(osaic) mutations transmitted by the (non mosaic)
334 dam to the male proband (Pr1) in whom the mutation was detected. Red circles:

335 O(oocyte)N(on)M(osaic) mutations transmitted by the (non mosaic) dam to the
336 three female probands (Pr2, Pr3, Pr4) in whom the mutation was detected.

337

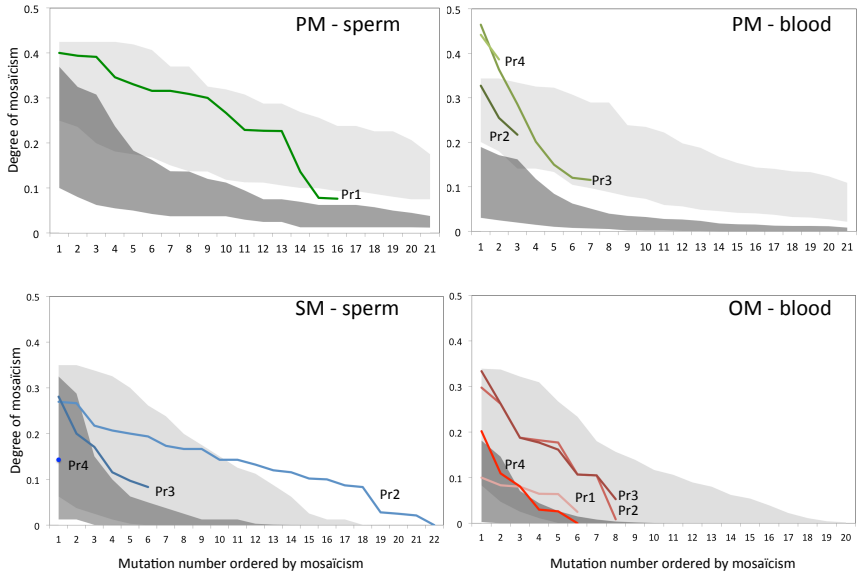


Figure 3A

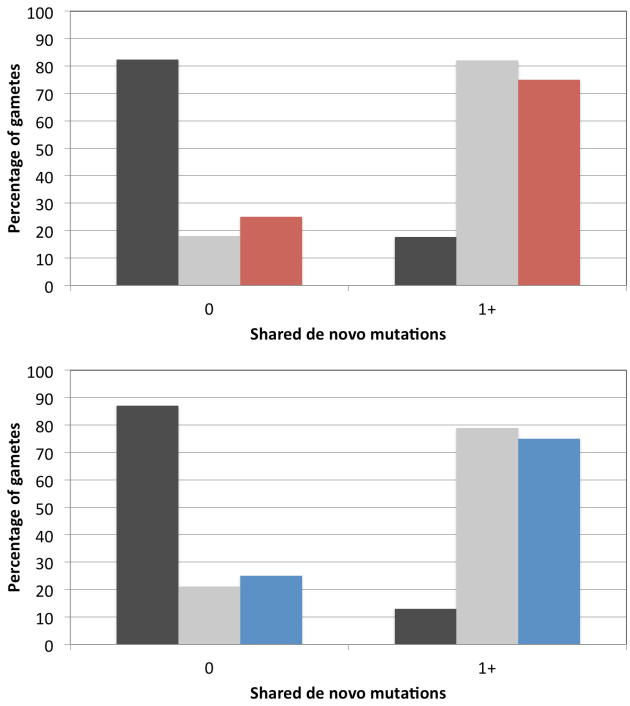


Figure 3B

Figure 3: (A) DNMs mosaic in sperm DNA of the male proband (PM – sperm) or sires (SM – sperm), or in blood DNA of the female probands (PM – blood) or dams (OM – blood) ranked by rate of mosaicism. Colored lines: real data. Pr1-4: proband 1-4. Dark gray shaded area: 95% confidence interval obtained from simulations assuming uniform mutation rate per cell division and 40 unrelated PGCs. Light gray shaded area: 95% confidence interval obtained from simulations assuming 10-fold higher mutation rate during the first 4 cell divisions, and 40 related PGCs (Table 1). **(B)** Distribution of the proportion of half-sibs of the four probands that share 0, or at least 1 (1+) of the DNMs detected in the corresponding proband. Red bars: real observations for DNMs

350 transmitted by the sire (SM+SNM). Blue bars: real observations for DNMs
351 transmitted by the dam (OM+ONM). Dark grey bars: expectation under the null
352 hypothesis of uniform prenatal mutation rate per cell division and 40 unrelated
353 PGCs. Light grey bars: expectation under the best alternative model assuming a
354 10x increased mutation rate during the first 4 cell division and 40 related PGCs
355 (Table 1).

356

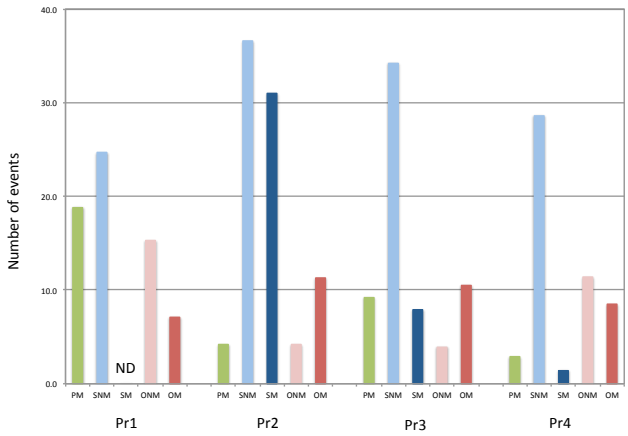


Figure 4A

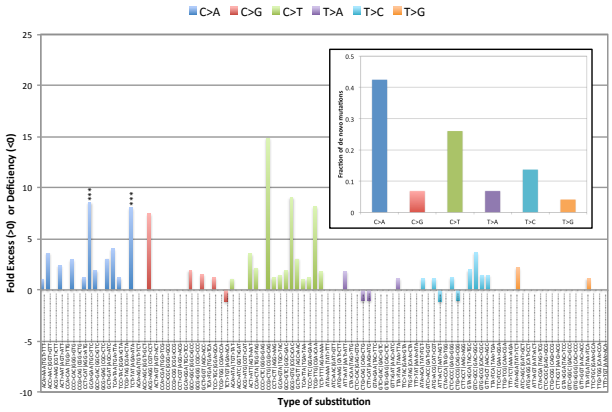


Figure 4C

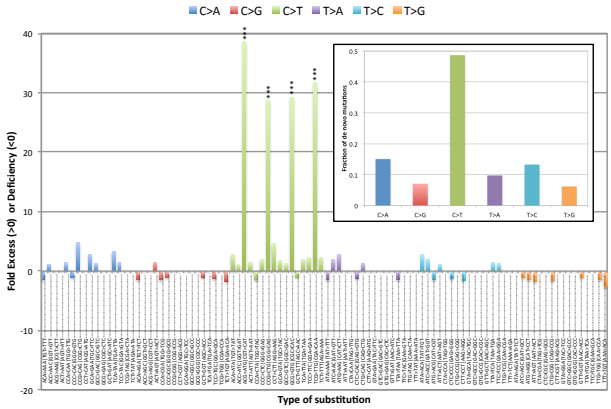


Figure 4B

Figure 4: (A) Types and number of DNMs detected in sperm DNA of male proband Pr1, and in blood DNA of female probands Pr2, Pr3 and Pr4. The overall numbers were estimated based on the observed DNMs, their rate of confirmation, and the estimated degree of genome coverage. **(B)** SNM and ONM (i.e. DNMs assumed to have occurred in the later stages of gametogenesis): fold excess or deficiency over expected for specific nucleotide substitutions accounting for trinucleotide context; ***: p< 0.002 (accounting for multiple

367 testing by Sidak correction). Inset: Proportion of DNMs corresponding to the six
368 possible types of nucleotide substitutions. **(C)** Idem for PM, SM and OM (i.e.
369 DNMs assumed to have occurred in the early stages of gametogenesis).

370

371 **Table 1: Relative likelihood of the observations under different models of**
372 **gametogenesis**

Fold increase mutation rate	During first x divisions	Number of PGC	Related PGCs or not	Log(LR)
10x	**4**	**40**	**T**	0.00
10x	**6**	40	T	-0.16
10x	**2**	40	T	-3.31
20x	2	40	T	-4.90
10x	4	40	**F**	-5.01
10x	2	**4**	F	-18.33
1x	6	40	T	-38.83

373 The first four columns correspond to the parameters that were tested in the model: (i)
374 the fold increase of the mutation rate (1x, 10x, 20x), (ii) during the x first cell divisions
375 (2, 4, 6), (iii) the number of PGCs (4, 40), and (iv) the ontogenetic relatedness of the
376 PGCs (F(alse) or T(rue)). Log(LR) corresponds to the logarithm (base 10) of the
377 likelihood of the data relative to the best model (first line). Parameters are in bold
378 when the corresponding model is the best given that parameter value. We only show
379 results for models that are the best given at least one parameter value. Likelihoods of
380 all models are given in Suppl. Table 2.