1 ***In vitro*** **fertilization does not increase the incidence of *de novo* copy number alterations in**

2 **fetal and placental lineages**

3

4 Masoud Zamani Esteki[1,2,3,16,17*], Triin Viltrop[4,16], Olga Tšuiko[3,4,16], Airi Tiirats[5], Mariann Koel[6], Margit Nõukas[7,8],

5 Olga Žilina[7], Katre Teearu[7], Heidi Marjonen[9], Hanna Kahila[10], Jeroen Meekels[1], Viveca Söderström-Anttila[11],

6 Anne-Maria Suikkari[11], Aila Tiitinen[10], Reedik Mägi[8], Sulev Kõks[12,13], Nina Kaminen-Ahola[9], Ants Kurg[7], Thierry

7 Voet[3,14,17*], Joris Robert Vermeesch[3,17*] & Andres Salumets[5,8,10,15,17*]

8

9 Affiliations

10 [1] Department of Clinical Genetics, Maastricht University Medical Centre+, Maastricht, The Netherlands

11 [2] Department of Genetics and Cell Biology, GROW School for Oncology and Developmental Biology, Maastricht University,

12 Maastricht, The Netherlands

13 [3] Centre for Human Genetics, University Hospital Leuven, Department of Human Genetics, KU Leuven, Leuven, Belgium

14 [4] Department of Biomedicine, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

15 [5] Department of Obstetrics and Gynaecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia

16 [6] Department of Cell Biology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

17 [7] Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

18 [8] Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia

19 [9] Department of Medical and Clinical Genetics, Medicum, University of Helsinki, Helsinki, Finland

20 [10] Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

21 [11] Väestöliitto Fertility Clinic, Helsinki, Finland

22 [12] Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Murdoch, Western Australia

23 [13] The Perron Institute for Neurological and Translational Science, Nedlands, Western Australia

24 [14] Sanger-EBI Single Cell Genomics Centre, Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

25 [15] Competence Centre on Health Technologies, Tartu, Estonia

26 [16] These authors contributed equally

27 [17] Co-senior authors

28

29 * Correspondence should be addressed to: M.Z.E. (masoud.zamaniesteki@mumc.nl); Th.V. (thierry.voet@kuleuven.be); J.R.V.

30 (joris.vermeesch@kuleuven.be) and A.S. (andres.salumets@ccht.ee)

31 **Although chromosomal instability (CIN) is a common phenomenon in cleavage-stage**

32 **embryogenesis following *in vitro* fertilization (IVF)[1-3], its rate in naturally conceived human**

33 **embryos is unknown. CIN leads to mosaic embryos that contain a combination of**

34 **genetically normal and abnormal cells and is significantly higher in *in vitro* produced**

35 **preimplantation embryos as compared to in vivo-conceived preimplantation embryos[4].**

36 **Even though embryos with CIN-derived complex aneuploidies may arrest between the**

37 **cleavage- and blastocyst-stage of embryogenesis[5,6], a high number of embryos containing**

38 **abnormal cells can pass this strong selection barrier[7,8]. However, prevalence and extent of**

39 **CIN during prenatal development and at birth, following IVF treatment, is not well**

40 **understood. Here we profiled the genomic landscape of fetal and placental tissues**

41 **postpartum from IVF and naturally conceived children to investigate the prevalence and**

42 **persistence of large genetic aberrations that have likely arisen from IVF-related CIN. We**

43 **demonstrate that CIN is not preserved at later stages of prenatal development, and that *de***

44 ***novo* numerical aberrations or large structural DNA imbalances occur at similar rates in**

45 **IVF and naturally conceived live-born neonates. Our findings affirm that human IVF**

46 **treatment has no detrimental effect on the chromosomal constitution of fetal and placental**

47 **lineages.**

48

49 CIN is a common event in human cleavage-stage embryogenesis that results in a high frequency

50 of chromosomal mosaicism in IVF embryos[1-3] and likely provides an explanation for the low

51 success rate per embryo transfer of IVF[9]. Strikingly, most IVF embryos are mixtures of euploid

52 and aneuploid cells and only a low fraction of IVF embryos contain a normal karyotype in all

53 blastomeres[1,10]. Between the cleavage- and blastocyst-stage of embryogenesis, CIN may cause

54   complex aneuploidies and the arrest of many embryos[5,6]. Even after this substantial selection,

55   genetically aberrant cells can be preserved at high frequency in the developing blastocysts, such

56   that about 40% of blastocysts have been reported to be uniformly euploid[7,8]. Despite the high

57   prevalence of human IVF embryos with aberrant chromosomal configurations, its clinical

58   implications are poorly understood. Recently, we established *Bos taurus* as a model organism to

59   study CIN at cleavage stage embryogenesis[11] and showed that CIN in *in vitro*-produced embryos

60   is significantly higher in prevalence when compared to *in vivo*-conceived embryos[4], rising the

61   alarming concern about the suboptimal *in vitro* environment on genetic integrity of embryos.

62   However, the fate of aberrant cells after cleavage-stage embryogenesis is not clear and it remains

63   impossible to investigate naturally conceived human embryos *in vivo*.

64   Cell lineages containing genetic aberrations in mosaic cleavage-stage embryos have been

65   speculated to survive in the trophectoderm, having no or little influence on the composition of

66   the inner cell mass and its subsequent fetal development[12]. This possibility is in accordance with

67   the confined placental mosaicism found in approximately 1–2% of ongoing natural pregnancies[13]

68   and the low prevalence of genetic mosaicism in postnatal and adult DNA samples[14].

69   Furthermore, euploid blastomeres may outgrow blastomeres with chromosomal aberrations[15],

70   allowing normal embryonic development and birth. This view of preferential selection of normal

71   cells is supported by recent studies reporting the birth of healthy babies after intrauterine transfer

72   of genetically mosaic IVF embryos[16,17]. In addition, a recent report described lineage-specific

73   fate in chimeric mouse embryos. This report demonstrated that different mechanisms act on

74   aneuploid cells in fetal or placenta lineages, such that apoptotic depletion of aneuploid cells

75   occurs in fetal lineage, whereas aneuploid cells in placenta persist but slow down their

76   divisions[18].

77   To investigate the persistence of IVF-related CIN, we profiled *de novo* DNA copy

78   number alterations in DNA samples extracted from cord blood cells and placental tissue, which

79   are derived from embryonic and extra-embryonic cell lineages, respectively, of both IVF and

80   naturally conceived neonates using haplarithmisis[3] and molecular karyotyping. Importantly, the

81   transferred embryos of IVF newborns were not screened nor counter selected for chromosomal

82   aneuploidy status before intrauterine transfer. Following live birth, we scrutinized genome-wide

83   single nucleotide polymorphism (SNP) profiles in DNA from the father, mother, placenta and

84   neonate umbilical cord blood of 111 families (quartets; **Fig. 1a**), of which 49 and 62 quartets

85   were from IVF and natural pregnancies, respectively. In the IVF cohort, the infertile patients

86   underwent fresh or frozen IVF ($n = 32$) or intracytoplasmic sperm injection (ICSI, $n = 17$)

87   embryo transfer, with about 4 out of 5 embryos transferred on days 2/3 (n=38, 77.5%) and the

88   remaining on days 5/6 at blastocyst stage (n=11, 22.5%). These quartets were further subdivided

89   based on birthweight at delivery, with 48 IVF and 50 natural conception families categorized as

90   'appropriate for gestational age' (AGA), and one IVF and 12 natural conception families as

91   'small for gestational age' (SGA) (**Supplementary Tables 1** and **2**). DNA copy number variants

92   (CNVs) were classified by whether they were found exclusively in a single sample or a few to all

93   samples within a quartet. This approach empowered our analysis and allowed us to differentiate

94   inherited CNVs (present in either or both parents, umbilical cord blood and placenta), *de novo*

95   CNVs (present in placenta and umbilical cord blood only) and placental *de novo* CNVs (present

96   in placenta only). CNVs were detected by the combined use of logR values, B Allele Frequency

97   (BAF) and parent-of-origin haplotypes (haplarithms). The latter analysis allowed us to account

98   for maternal DNA contribution in placental tissue (**Fig. 1b, Extended Data Fig. 1**), to identify

99   the parental origin of *de novo* CNVs (**Fig. 1c** and **Table 1**) and to more accurately estimate the

100   level of mosaicism of CNVs within the DNA sample (**Fig. 2a, b** and **Table 1**).

101   We detected *de novo* large CNVs (>100 kb) in placental genomes of twelve families; of

102   the total fifteen detected aberrations four were also present in newborns and eleven were

103   exclusively present in the placenta. We did not detect any CNVs present only in umbilical cord

104   blood DNA, which would indicate a genomic alteration in the neonate only. The mosaic ($n = 12$)

105   *de novo* CNVs were 5–70% mosaic (**Table 1**) and were together with *de novo* non-mosaic CNVs

106   ($n = 3$) randomly located across the genome (**Fig. 1d**). Our analysis showed comparable

107   prevalence ($P = 0.327$ odds ratio test or $P = 0.775$ Pearson's chi square test) of *de novo* CNVs in

108   both cohorts with AGA: six of the 50 families with naturally conceived neonates (12.0%) and in

109   three of the 48 families with IVF neonates (6.3%) (**Table 1**). When IVF with fresh and frozen

110   embryo transfers was compared to ICSI fresh/frozen embryo transfers similar proportions of

111   mosaic *de novo* CNV cases were found (6.3%, 2/32 and 5.9%, 1/17, respectively, $P = 0.96$ odds

112   ratio test), indicating that the mode of fertilization (IVF vs ICSI) and embryo freezing are

113   unlikely to influence the rate of large CNVs in fetal and placental lineages.

114   We found three partial trisomies with different levels of mosaicism in family P172 from

115   the naturally conceived cohort and a mosaic full-chromosome trisomy in family P106 from the

116   IVF cohort (**Fig. 2a, b, Table 1** and **Extended Data Figs. 2** and **3**), showing mosaicism of 5–

117   25%. We obtained biopsies from two other spatially distinct locations, applied haplarithmisis and

118   confirmed the results using droplet digital PCR (ddPCR) (**Fig. 2c** and **Supplementary Table 3**).

119   The partial mosaic trisomies on Chr 6 (22.4 Mb), Chr 9 (5.8 Mb) and Chr 21 (22.0 Mb) of P172

120   were not present in the other placental biopsies (**Fig. 2c** and **Extended Data Fig. 2**). In contrast,

121   the full Chr 2 mosaic trisomy was present in all the P106 placental biopsies (**Fig. 2c** and

122 **Extended Data Fig. 3**). This result indicates that the mosaic Chr 2 trisomy arose as CIN in the

123 early cleavage-stage embryo, but segregated into the placental lineage only (**Fig. 3** and **Table 1**),

124 because it was present across different parts of the placenta with the same degree of mosaicism.

125 The presence of the low-degree mosaic Chr 2 trisomy reinforces an earlier observation in mouse

126 embryos, which suggested a progressive reduction of aneuploid cell lineage in placenta[18] because

127 only 10% of the cells carried the anomaly. However, the segmental trisomies in family P172

128 were present in only one of the placental locations, indicating that these genomic alterations

129 either emerged later in gestation through local clonal expansion of aberrant cells in the placenta

130 or appeared early in development, contributed to the trophectoderm and were largely overgrown

131 by normal cells in placental tissue. The presence of multiple *de novo* DNA copy number variants

132 has recently been shown to lead to genomic disorders[19]. The finding that multiple CNVs are

133 present with a different degree of mosaicism, albeit in placenta genome only, is indicative that

134 they occurred in a cumulative fashion[20].

135   To further investigate CNV heterogeneity across the placenta and reaffirm the (non-)

136 mosaic nature of the detected *de novo* CNVs, we further analysed 20 DNA samples derived from

137 spatially different locations of all but two placentas (F351 and F1021) with *de novo* CNVs

138 (**Table 1**). As expected, *de novo* non-mosaic CNVs found in both placenta and cord blood were

139 consistently present in all the biopsies, whereas *de novo* mosaic CNVs were not present in the

140 other biopsies (**Extended Data Fig. 4**). Importantly, the full Chr 2 mosaic trisomy was found in

141 all the biopsies (**Extended Data Fig. 3**), indicating its early embryonic origin (see above).

142   We identified two clinically significant non-mosaic CNVs in the SGA cohort, with the

143 genetic aberrations found both in umbilical cord and placental DNA. P153 carried a *de novo*

144 deletion on Chr 7q11.23. Postpartum follow-up revealed clinical features of Williams–Beuren

145   syndrome[21]. P177 carried a paternally inherited duplication on Chr 15 with features of Chr

146   15q13.3 duplication syndrome[22,23]. The low birth weight of these neonates might be indicative of

147   the underlying chromosomal aberration; therefore, the SGA neonates should be subjected to

148   submicroscopic chromosomal analysis.

149   Although the prevalence of *de novo* CNVs in SGA (25.0%, 3/12) and AGA (12.0%, 6/50)

150   groups of the cohort of natural conceptions is not significantly different ($P = 0.44$, odds ratio

151   test), current sample size for SGA ($n = 12$), provides limited statistical power (<0.50). A study

152   with increased SGA sample size ($n > 44$, **Methods**) is required to evaluate the effect of

153   birthweight on the prevalence of *de novo* CNVs. Furthermore, our data indicates that the current

154   sample size (n=50 per AGA cohort) provides limited statistical power to detect CNVs with small

155   effect size (w=0.1-0.2) between the two cohorts. We estimated that future studies with smaller

156   effect size require a data set of ~10-fold larger[24].

157   The human placenta is characterized by deep invasion of trophoblasts into the

158   endometrium. Trophoblasts have similar properties to cancer cells, including rapid proliferation,

159   migration and invasiveness. As CIN is a hallmark of both tumorigenesis and early

160   embryogenesis, a recent study hypothesized that the placenta resembles tumor tissue by

161   demonstrating an extensive load of *de novo* genomic gains, harboring excessively expressed

162   genes required for the invasiveness of trophoblasts[25]. Our results challenge this view; as we

163   found that placental *de novo* CNVs were present in only 12 of the 111 families (10.8%) (**Fig. 1d**

164   and **Table 1**), and placental heterogeneity was found in 85.7% (6/7) of the studied 'placenta

165   only' *de novo* mosaic CNVs. To determine whether genes located in the *de novo* CNV regions (*n*

166   = 475) had specific placental biology-related functions, we analyzed gene expression in full-term

167   placentas following live birth (**Methods**). Only 18.1% (*n* = 86) of these genes had moderate or

168   high expression according to the whole RNA sequencing of placental tissue (**Fig. 2d** and

169   **Supplementary Table 4**), and genes from regions with *de novo* gains (18.0%, 77/428) and

170   losses (19.1%, 9/47, $P = 0.84$ odds ratio test) demonstrated similar proportion of moderate-to-

171   high expression genes in placental tissue. Moreover, enrichment analyses of genes located in *de*

172   *novo* CNV regions did not reveal any molecular pathway related to placental biology. All this

173   indicates that these new CNVs are likely sporadic and do not play any role in placental biology.

174   In this study, our main outcome measure was the extent and prevalence of *de novo* large

175   CNVs (>100 kb) in live-born IVF-conceived neonates when compared to naturally conceived

176   ones. Although we have not studied the embryonic loss with or without aneuploidies in IVF- or

177   naturally conceived pregnancies, we speculate that when all cells of day-3 (cleavage-stage)

178   embryos are investigated for the presence of (segmental) aneuploidies, on average 80% of the

179   embryos present with at least one aneuploid cell (range 70%-90%)[1,3,10]. The IVF success rate per

180   embryo transfer of day-3 embryos is about 35% in Helsinki and Tartu IVF centres. Assuming

181   that all euploid day-3 embryos survive, they would constitute at most 20% of the total number of

182   embryos transferred resulting in pregnancy/delivery, but indicating that at least 15% of all

183   embryos did contain some aneuploid cells. This would amount to 60% and 40% of the live births

184   resulting from respectively euploid and mosaic euploid/aneuploid cleavage-stage embryos.

185   Conversely, about 20% (15/80%) of mosaic embryos survive and result in healthy live births,

186   whereas 80% (65/80%) are selected against. Since it is unlikely that all euploid embryos make it

187   to term, the latter is a lower estimate.

188   Mosaicism rates in human IVF embryos are still debated and vary across studies

189   primarily because of the definition of mosaicism used, the technology applied, and the variance

190   in assisted reproduction protocols used. First, important differences in the definition of a mosaic

191   embryo exist between studies. Embryos having both euploid and aneuploid cells were defined as

192   mosaic in 1993[26], but since, an arbitrary classification of embryos as mosaic or diploid

193   influenced the mosaicism rate reported by different studies. For instance, some studies[27-29]

194   considered an embryo mosaic only if more than 50% of blastomeres within the embryo were

195   aneuploid. The reason for considering an embryo having less than 50% abnormal cells as diploid

196   was that they are likely viable and a low percentage of abnormal cells in those embryos was

197   believed clinically irrelevant. Second, since the discovery of aneuploidies in human

198   preimplantation embryos[30,31], different technologies have been used to profile the chromosomal

199   constitution of preimplantation embryos. Depending on the technology used the rate of

200   mosaicism was reported to range from 15%[32] to >90%[1,3]. However, in a meta-analysis of all

201   technologies included and carried out on 815 embryos from different developmental stages, 73%

202   were classified as mosaic[10]. The high variability occurs, because different technologies have

203   different sensitivities and specificities for profiling the ploidy state of a selected chromosomes

204   (e.g. FISH[32]) or the entire genome (microarray[1,3] or next-generation sequencing[3,33]). The above,

205   and the number of cells per embryo analyzed with those technologies, can affect the reported

206   embryo mosaicism rates in those studies, with the studies analysing all the cells from a single

207   embryo showing the higher rate of mosaicism[10]. In current study, we applied the same wet- and

208   dry-lab technologies as in our preimplantation embryo studies[3,34,35] and further validated our

209   approach using sensitive droplet digital PCR and standard copy number analysis methods[36,37],

210   allowing us to rule out technology-driven bias. Third, the mosaicism rate is reported to be

211   influenced by the fertility clinic where the IVF procedure is carried out[38]; one key component

212   that can explain differences in the reported mosaicism rate is embryo culture conditions[39].

213   However, in this study, the IVF procedure, including culture conditions, were similar in the

214   participating IVF clinics.

215   Taken together, our findings suggest that the high level of chromosomal mosaicism in

216   IVF cleavage-stage embryos is not preserved at later stages of prenatal development. Although

217   we found a low-level mosaic trisomy in one IVF placental genome, it was not present in the fetal

218   lineage of the same pregnancy, suggesting a proliferative defect of trisomic cells. The sporadic

219   *de novo* genomic aberrations found in approximately 10.8% of pregnancies were scattered across

220   the genome and represented random and rare events of embryonic CIN, without any observed

221   functional consequence on placental biology or fetal health. Thus, our findings confirm that the

222   IVF procedure has no detrimental effect on large genomic rearrangements in fetal and placental

223   lineages of live-born neonates, which helps to remove a major health concern for IVF neonates.

224

238

AUTHOR CONTRIBUTIONS

240   M.Z.E., Tr.V., A.K., Th.V., J.R.V. and A.S. conceived the study and designed the experiments.

241   M.Z.E. Tr.V., O.T., J.M., Th.V., J.R.V. and A.S. analysed and interpreted the data. Tr.V., O.T.,

242   A.T., H.M., H.K., V.S-A., A-M.S., A.T., N.K-A. and S.K. carried out sample collection. O.T.

243   and M.Z.E. performed ddPCR assays. M.K. carried out RNA sequencing analysis. M.N., K.T.,

244   O.Z. and R.M. performed PennCNV and QuantiSNP analyses. M.Z.E. drafted the initial version

245   of the manuscript. M.Z.E., Tr.V., A.K., Th.V., J.R.V. and A.S. wrote and edited the manuscript.

246   M.Z.E., Th.V., J.R.V. and A.S. jointly supervised this study. All the authors read and approved

247   the manuscript for submission.

248

COMPETING FINANCIAL INTERESTS

250   M.Z.E., J.R.V., and Th.V. are co-inventors on a patent applications ZL913096-

251   PCT/EP2014/068315-WO/2015/028576 ''Haplotyping and copy-number typing using

252   polymorphic variant allelic frequencies''.

253

References

255   1.   Vanneste, E., *et al.* Chromosome instability is common in human cleavage-stage
256   embryos. *Nat Med* **15**, 577-583 (2009).
257   2.   Chavez, S.L., *et al.* Dynamic blastomere behaviour reflects human embryo ploidy by the
258   four-cell stage. *Nature communications* **3**, 1251 (2012).
259   3.   Zamani Esteki, M., *et al.* Concurrent Whole-Genome Haplotyping and Copy-Number
260   Profiling of Single Cells. *Am J Hum Genet* (2015).
261   4.   Tsuiko, O., *et al.* Genome stability of bovine in vivo-conceived cleavage-stage embryos
262   is higher compared to in vitro-produced embryos. *Hum Reprod* **32**, 2348-2357 (2017).

263 5.   McCoy, R.C.*, et al.* Evidence of Selection against Complex Mitotic-Origin Aneuploidy
264 during Preimplantation Development. *PLoS Genet* **11**, e1005601 (2015).
265 6.   McCoy, R.C.*, et al.* Common variants spanning PLK4 are associated with mitotic-origin
266 aneuploidy in human embryos. *Science* **348**, 235-238 (2015).
267 7.   Fragouli, E.*, et al.* Cytogenetic analysis of human blastocysts with the use of FISH, CGH
268 and aCGH: scientific data and technical evaluation. *Hum Reprod* **26**, 480-490 (2011).
269 8.   Popovic, M.*, et al.* Chromosomal mosaicism in human blastocysts: the ultimate challenge
270 of preimplantation genetic testing? *Hum Reprod* **33**, 1342-1354 (2018).
271 9.   Vanneste, E.*, et al.* What next for preimplantation genetic screening? High mitotic
272 chromosome instability rate provides the biological basis for the low success rate. *Hum Reprod*
273 **24**, 2679-2682 (2009).
274 10.   van Echten-Arends, J.*, et al.* Chromosomal mosaicism in human preimplantation
275 embryos: a systematic review. *Hum Reprod Update* **17**, 620-627 (2011).
276 11.   Destouni, A.*, et al.* Zygotes segregate entire parental genomes in distinct blastomere
277 lineages causing cleavage-stage chimerism and mixoploidy. *Genome Res* **26**, 567-578 (2016).
278 12.   Ledbetter, D.H. Chaos in the embryo. *Nature medicine* **15**, 490-491 (2009).
279 13.   Kalousek, D.K. & Vekemans, M. Confined placental mosaicism. *Journal of medical*
280 *genetics* **33**, 529-533 (1996).
281 14.   Biesecker, L.G. & Spinner, N.B. A genomic view of mosaicism and human disease. *Nat*
282 *Rev Genet* **14**, 307-320 (2013).
283 15.   Santos, M.A.*, et al.* The fate of the mosaic embryo: chromosomal constitution and
284 development of Day 4, 5 and 8 human embryos. *Hum Reprod* **25**, 1916-1926 (2010).
285 16.   Greco, E., Minasi, M.G. & Fiorentino, F. Healthy Babies after Intrauterine Transfer of
286 Mosaic Aneuploid Blastocysts. *N Engl J Med* **373**, 2089-2090 (2015).
287 17.   Dimitriadou, E.*, et al.* Principles guiding embryo selection following genome-wide
288 haplotyping of preimplantation embryos. *Hum Reprod* **32**, 687-697 (2017).
289 18.   Bolton, H.*, et al.* Mouse model of chromosome mosaicism reveals lineage-specific
290 depletion of aneuploid cells and normal developmental potential. *Nat Commun* **7**, 11165 (2016).
291 19.   Liu, P.*, et al.* An Organismal CNV Mutator Phenotype Restricted to Early Human
292 Development. *Cell* **168**, 830-842 e837 (2017).
293 20.   Voet, T. & Vermeesch, J.R. Mutational Processes Shaping the Genome in Early Human
294 Embryos. *Cell* **168**, 751-753 (2017).
295 21.   Merla, G., Brunetti-Pierri, N., Micale, L. & Fusco, C. Copy number variants at Williams-
296 Beuren syndrome 7q11.23 region. *Human genetics* **128**, 3-26 (2010).
297 22.   Szafranski, P.*, et al.* Structures and molecular mechanisms for common 15q13.3
298 microduplications involving CHRNA7: benign or pathological? *Human mutation* **31**, 840-850
299 (2010).
300 23.   van Bon, B.W.*, et al.* Further delineation of the 15q13 microdeletion and duplication
301 syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. *Journal of*
302 *medical genetics* **46**, 511-523 (2009).
303 24.   Chow SC, S.J., Wang H. *Sample Size Calculation in Clinical Research.*, (New York:
304 Marcel Dekker, 2003).
305 25.   Kasak, L., Rull, K., Vaas, P., Teesalu, P. & Laan, M. Extensive load of somatic CNVs in
306 the human placenta. *Sci Rep* **5**, 8342 (2015).

307   26.   Delhanty, J.D*., et al.* Detection of aneuploidy and chromosomal mosaicism in human
308   embryos during preimplantation sex determination by fluorescent in situ hybridisation, (FISH).
309   *Hum Mol Genet* **2**, 1183-1185 (1993).
310   27.   Munne, S*., et al.* Assessment of numeric abnormalities of X, Y, 18, and 16 chromosomes
311   in preimplantation human embryos before transfer. *Am J Obstet Gynecol* **172**, 1191-1199;
312   discussion 1199-1201 (1995).
313   28.   Ziebe, S*., et al.* FISH analysis for chromosomes 13, 16, 18, 21, 22, X and Y in all
314   blastomeres of IVF pre-embryos from 144 randomly selected donated human oocytes and impact
315   on pre-embryo morphology. *Hum Reprod* **18**, 2575-2581 (2003).
316   29.   Baart, E.B*., et al.* Preimplantation genetic screening reveals a high incidence of
317   aneuploidy and mosaicism in embryos from young women undergoing IVF. *Hum Reprod* **21**,
318   223-233 (2006).
319   30.   Steptoe, P.C. & Edwards, R.G. Birth after the reimplantation of a human embryo. *Lancet*
320   **2**, 366 (1978).
321   31.   Angell, R.R., Aitken, R.J., van Look, P.F., Lumsden, M.A. & Templeton, A.A.
322   Chromosome abnormalities in human embryos after in vitro fertilization. *Nature* **303**, 336-338
323   (1983).
324   32.   Harper, J.C*., et al.* Mosaicism of autosomes and sex chromosomes in morphologically
325   normal, monospermic preimplantation human embryos. *Prenat Diagn* **15**, 41-49 (1995).
326   33.   Voet, T*., et al.* Single-cell paired-end genome sequencing reveals structural variation per
327   cell cycle. *Nucleic Acids Res* **41**, 6119-6138 (2013).
328   34.   Dimitriadou, E., Zamani Esteki, M., Vermeesch, JR. Copy number variation analysis by
329   array analysis of single cells following whole genome amplification. in *Methods in Molecular*
330   *Biology.* (ed. Kroneis, T.) (Springer, New York, 2015).
331   35.   Destouni, A*., et al.* Genome-wide haplotyping embryos developing from 0PN and 1PN
332   zygotes increases transferrable embryos in PGT-M. *Hum Reprod* (2018).
333   36.   Wang, K*., et al.* PennCNV: an integrated hidden Markov model designed for high-
334   resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*
335   **17**, 1665-1674 (2007).
336   37.   Colella, S*., et al.* QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and
337   accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* **35**, 2013-
338   2025 (2007).
339   38.   Munne, S*., et al.* Euploidy rates in donor egg cycles significantly differ between fertility
340   centers. *Hum Reprod* **32**, 743-749 (2017).
341   39.   Munne, S*., et al.* Detailed investigation into the cytogenetic constitution and pregnancy
342   outcome of replacing mosaic blastocysts detected with the use of high-resolution next-generation
343   sequencing. *Fertil Steril* **108**, 62-71 e68 (2017).

344 **Figure Note**

345 **Figure legends**

346 **Figure 1 | Haplarithmisis reveals genetic mosaicism with parent-of-origin information. (a)**

347 Schematic representation of a quartet. (**b**) Haplarithmisis can determine the maternal DNA

348 contribution to the bulk placental DNA sample (see **Methods**). (**c**) A maternally inherited CNV

349 (red arrows) and placental *de novo* CNV with paternal origin found on Chr 5 of P070 (purple

350 arrow). B allele fraction (BAF) of a SNP is the signal intensity of allele B over signal intensities

351 of both expected alleles for that SNP ($BAF = (\frac{B}{A+B})$), and logR (relative copy number) of that

352 SNP is the base 2 logarithm of the summed normalized signal intensities of both alleles over

353 expected signal intensity values ($logR = \frac{R_{Sample}}{R_{expexted}}$). (**d**) Overview of *de novo* CNVs found in this

354 study (see also **Source data**). G and L denote gains and losses, respectively, as shown in **Table**

355 **1**.

356

357 **Figure 2 | Mosaic *de novo* CNVs and overlap with placental transcriptome.** Haplarithms

358 revealed *de novo* mosaic (**a**) partial trisomies in P172 placenta of paternal origin (see also

359 **Extended Data Fig. 2**) and (**b**) full-chromosome trisomy in P106 placenta of maternal origin

360 (see also **Extended Data Fig. 3**). These aberrations are indicated by purple arrows. (**c**)

361 Validation of detected mosaic *de novo* CNVs by ddPCR (each circle and error bar indicate mean

362 and standard deviation, respectively, of four independent measurements). (**d**) Placental

363 expression of genes detected in *de novo* CNV regions, expressed as reads per kilobase per

364 million mapped reads (RPKM).

365

Figure 3 | Schematic representation of plausible occurrence and segregation of *de novo* CNVs into fetal or placental lineages found in this study. *De novo* CNVs can occur before conception in the germ cells or during conception (I), during the cleavage divisions of early development (II-IV) or later during development, e.g. blastocyst-stage embryogenesis (V-VI). Note that, for simplicity in the CIN-derived mosaic CNVs, we merely showed CNV burst in 2-cell stage embryos (II-IV types) and blastocyst-stage embryos (V-VI types) (see also **Table 1**). Red circles indicate cells carrying *de novo* CNV.

*Table 1 | **De novo CNVs detected by haplarithmisis.***

Cohort	Family	Genome coordinates	Length (bp)	SNP-probes (#)	O:T[a]	CN[b]	Origin[c]	Mosaicism (%)[d]	Biopsy (cnv:bps)[e]
Naturally conceived. Appropriate birth weight for gestational age	**P017**	Chr16:78670954-78867839	196885	37	OP:III/V	L1	Maternal	65%	1:3
	P073	Chr7:13238496-13348832	110336	26	OP:III/V	L2	Paternal	60%	1:3
	P080	Chr3:174684649-174831994	147345	32	OP:III/V	L3	Maternal	60%	1:3
	P142	Chr14:47624206-47746980	122774	16	OP:III/V	L4	Paternal	70%	1:3
	P162	Chr14:64554517-64953726	399209	189	OP:III/V	L5	Paternal	40%	1:3
	F351	Chr22:23250737-23644794	394057	99	B:I	L6	Maternal	-	NA[f]
	F351	Chr4:91709484-91910883	201399	24	OP:III/V	L7	Maternal	70%	NA[f]
Naturally conceived. Small birth weight for gestational age	**P070**	Chr5:103670162-104157752	487590	45	OP:III/V	G1	Paternal	65%	1:3
	P153	Chr7:72305671-74115258	1809587	359	B:I	L8	Maternal	-	3:3
	P172	Chr21:26058246-48084247	22026001	5796	OP:III/V	G2	Paternal	25%	1:5
	P172	Chr6:204072-22569968	22365896	4106	OP:III/V	G3	Paternal	10%	1:5
	P172	Chr9:135277529-141031439	5753910	2915	OP:III/V	G4	Paternal	5%	1:5
IVF. Appropriate birth weight for gestational age	**P091**	Chr16:21605180-21966869	361689	41	B:I	L9	Maternal	-	3:3
	P106	Chr2:1-243048760	243015748	44591	OP:III/V	G5	Maternal	10%	5:5
	F1021	Chr17:2832092-3070476	238384	71	B:II	G6	Maternal	40%	NA[f]

375   [a] O:T demonstrates occurrence and type. Occurrence can be B = *de novo* copy number alteration occurred in both placenta and
376   umbilical cord DNA samples, OP = *de novo* copy number alteration occurred in placental DNA sample only; and type can be in
377   one of the stratified CNV types illustrated in **Fig. 3**. [b] CN = copy number; L = Loss; G = Gain; [c] parent of origin is determined
378   based on the parental allele which is affected, e.g. maternal loss or gain means that maternal allele is deleted or duplicated,
379   respectively; for non-mosaic deletions present in both placenta and cord blood, we used parent-of-origin module of siCHILD to
380   determine parental origin; [d] percentage of abnormal cells; [e] cnv:bps demonstrates number of CNVs ('cnv') found in spatially
381   different placenta biopsies ('bps'; see also **Extended Data Figs. 2**, **3** and **4**); [f] more biopsies from these placentas were not
382   available.

METHODS

Ethical approval. The study was approved by the local Ethics Committee of the University of Tartu (213/T-21) and the Ethics Committee of Helsinki University Central Hospital (285/13/03/03/2013). All couples signed the informed consent form.

Selection of participating families. In total, 111 couples participated in this study (**Supplementary Table 1** and **2**), including 62 couples who delivered neonates after spontaneous conception (natural cohort) and 49 couples who delivered after IVF (IVF cohort). No prior genetic predisposition was identified in any of the couples participating in this study. Clinical data and biological material were collected from singleton pregnancies at birth from vaginal delivery or caesarean section (**Supplementary Table 2**). For each family, we isolated DNA from the father, mother, as well as the placenta and neonate umbilical cord blood following live birth (quartet DNA sample set). The natural cohort was divided into two subgroups based on the neonate birthweight: (i) AGA neonates, with a birthweight between the 10th and 90th percentile ($n = 50$ families), and (ii) SGA neonates, with a birthweight below the 10th percentile ($n = 12$ families). The IVF cohort consisted of 48 families with AGA neonates and one family with an SGA neonate. Birthweight percentiles were calculated by using data from the Estonian Medical Birth Registry[40] and the Finnish National Institute for Health and Welfare[41].

Sample collection and DNA extraction. Full-thickness placental blocks of around 1 cm were taken from placenta and stored at -80 °C for subsequent DNA extraction. To investigate placenta CNV heterogeneity, we analysed 3-5 biopsies from spatially distinct locations of 10 placentas

17

406   (**Table 1** and **Extended Data Figs. 2**, **3** and **4**). All samples were collected by the same medical

407   personnel. Maturity and health of term placenta were confirmed by histological examination. A

408   total of 9 mL of peripheral venous blood from parents and 4 mL of umbilical cord blood were

409   collected into ethylenediaminetetraacetic acid-containing tubes (BD Vacutainer, Becton,

410   Dickinson and Co.). In Estonia, DNA was isolated from blood and placental tissue by using the

411   PureLink Genomic DNA Kit (Invitrogen, Life Technologies). DNA samples were eluted with

412   200 µL of PureLink Genomic Elution Buffer (10 mM Tris-HCl, pH 9.0, 0.1 mM EDTA) and

413   stored at -20 °C until analysis. Finnish parental DNA from blood samples was extracted by

414   NucleoSpin Blood XL Kit (Macherey-Nagel) and DNA from umbilical cord blood by

415   NucleoSpin Tissue Kit (Macherey-Nagel). NanoDrop 2000 (Thermo Fisher Scientific Inc.) was

416   used to assess DNA concentration and purity.

417

418   **SNP genotyping.** For SNP genotyping, the Infinium PsychArray-24 v1.1 (Illumina; GEO:

419   GPL22819) and PsychArray-24 v1.3 BeadChip SNP arrays (Illumina; GEO: GPL26750) were

420   used, which contains approximately 590,000 SNP markers with a median marker spacing of 5.5

421   kb. Genotype calls, SNP B-allele frequency values and logR values of all samples were

422   computed by the Illumina GenomeStudio software.

423

424   **Haplarithmisis.** We applied haplarithmisis[3] on each quartet DNA sample. Discrete SNP calls

425   derived from umbilical cord blood were used as seeds to phase the parental genotypes. Parental

426   haplarithms were used to infer DNA copy number state, parent-of-origin and level of mosaicism

427   of *de novo* copy number alterations in placental/newborn's DNA. Level of mosaicism was

428   calculated as described[42].

429   Haplarithmisis can be used to estimate contributions of maternal and fetal genomes in

430   placental DNA samples. We carried out a simulation analysis by blending maternal and fetal

431   alleles with different proportions (from $1\%_{Mother} : 99\%_{Child}$ to $99\%_{Mother} : 1\%_{Child}$) and applied

432   haplarithmisis (**Extended Data Fig. 5**). This process allowed us to determine accurately the

433   maternal and fetal contributions to placental tissue (**Supplementary Table 5** and **Extended**

434   **Data Fig. 1**).

435

436   **Copy number profiling.** We applied both PennCNV[36] (with 0.2% false positive rate and 86%

437   sensitivity for constitutional CNVs) and QuantiSNP[37] (with 0.1% false positive rate for

438   constitutional CNVs) algorithms as well as the copy number module of siCHILD[3] on the raw

439   logR values with minor modifications. Raw logR-values were smoothed by using a moving

440   average window of five consecutive SNP probes, wave-corrected for GC% bias by a Lowess fit

441   and normalized to a trimmed mean of normal diploid chromosomes. Normalized logR-values

442   were segmented by piecewise-constant fitting[43] (gamma = 14). A copy number alteration was

443   considered reliable if more than six consecutive SNP probes supported the copy number change,

444   its length was more than 33 kb (six times 5.5 kb median distance between two probes) and it was

445   supported by raw or segmented parent-of-origin specific haplotypes, coined haplarithms. As a

446   sanity check, we investigated whether there are signals that are detected only by haplarithmisis

447   (**Supplementary Table 6**) or logR analysis (**Supplementary Table 7**). We did not include those

448   in the analysis.

449   Copy number analysis was empowered by our quartet sample-set and haplarithmisis, as

450   we used all of the samples within a quartet to call a reliable genomic change. For each of the

451   detected DNA copy number alterations in one sample (e.g. placenta), we checked whether the

452   same alteration was present in other samples of the same quartet. This approach allowed us to

453   determine the parental origin of inherited CNVs, to detect *de novo* CNVs (i.e. in placental

454   samples or in both placental and umbilical cord samples), to find CNVs exclusive to any of the

455   samples (e.g. present solely in the placental genome) and to reveal genotyping artifacts.

456

457   **Other statistical and computational analysis.** Since both predictor (with or without IVF

458   treatment) and outcome (with or without *de novo* CNVs) are dichotomous, chi-squared test is the

459   most suitable test[44]. To determine appropriate sample size, we performed a power analysis using

460   chi-squared test with confidence level of at least 95% and feeding the conventional effect size

461   (w) values of 0.1, 0.3 and 0.5, which correspond to small, medium and large effect sizes,

462   respectively[45]. With a medium effect size at least 40 samples per cohort is required to reach to

463   the power of 0.8[24]. Given the sample size of about 50 per cohort (50 naturally conceived quartets

464   and 48 IVF quartets with appropriate birthweight for gestational age), this study has >80% power

465   (alpha = 0.05; middle effect size of 0.3; case to control ratio ca 1) to detect a significant

466   difference (<0.05).

467   We also compared large CNVs with mosaic or non-mosaic nature separately. The rate of

468   mosaic large CNVs found in naturally conceived versus IVF-conceived AGA groups were not

469   significantly different ($P = 0.157$ Pearson's chi square test). Similarly, the rate of non-mosaic

470   CNVs found in naturally conceived versus IVF-conceived AGA groups were not significantly

471   different ($P = 0.977$ Pearson's chi square test).

472   We applied Hotelling's T-square multivariate statistical test on the numerical

473   demographic features of AGA neonates of both IVF and natural cohorts, including birth weight,

474   birth length, placenta weight, parity and gestational age. This analysis showed no significant

475   difference in these cohorts (P = 0.22, Hotelling's two sample T2-test).

476   For statistical analyses we applied R (http://www.r-project.org/). For visualization

477   purposes we applied siCHILD[3] and R (http://www.r-project.org/).

478

479   **RNA sequencing and comparison with placental transcriptome.** We performed RNA

480   sequencing as described previously[46] (**Supplementary Table 4**). Briefly, 12 placental samples

481   were collected from singleton pregnancies at term (range, 38–42 weeks of gestational age).

482   Placental samples from different families were used for RNA sequencing and CNV analysis.

483   Full-thickness placental blocks of around 2–3 cm were taken from three different regions of

484   placenta. Total RNA was extracted with the mirVana miRNA Isolation Kit (Ambion, Life

485   Technologies), and RNA samples were pooled. DNA contamination was removed with the

486   DNA-free Kit (Invitrogen, Life Technologies). Agilent 2100 Bioanalyzer and RNA 6000 Nano

487   Kit (Agilent Technologies) were used to assess the integrity and concentration of total RNA

488   samples.

489   For RNA-Seq library preparation, 45 ng of total RNA was amplified with the Ovation

490   RNA-Seq System V2 Kit (NuGen). The SOLiD 5500 System DNA fragment library (Life

491   Technologies) was used for library preparation. Pooled samples were analyzed on three

492   FlowChip lanes, and approximately 40 million mappable paired-end reads were obtained per

493   sample. DESeq (version 3.0.2) was used to calculate reads per kilobase per million (RPKM). All

494   genes from placenta were grouped into four classes according to their RPKM values (0–1, 1–10,

495   10–100, and >100). Genes with moderate to high expression were defined by >10 mapped reads

496   per RPKM and treated as placenta expressed genes.

497

498   Protein-coding genes from *de novo* CNV regions were extracted from the UCSC browser

499 (https://genome.ucsc.edu/). Human genome build GRC37/hg19 was used as a reference. Genes

500 found at *de novo* CNV regions were studied for gene expression in placental tissue from 12

501 individuals. Enrichment analyses for Gene Ontology terms were performed for genes present at

502 *de novo* CNVs, by using g:Profiler (version Ensembl 90, Ensembl Genomes 37, rev 1741, build

503 date 2017-10-19)[47] with default settings. Briefly, this tool performs gene set enrichment analyses

504 for Gene Ontology, Human Phenotype Ontology and Kyoto Encyclopedia of Gene and Genomes

505 gene sets. Enrichment analyses are conducted by hypergeometric test and resulting enrichment

506 P-values are further adjusted for multiple testing by g:SCS (Set Counts and Sizes) method,

507 developed to address the non-independent structure of tested GO terms[47]. Separate enrichment

508 analyses were conducted to different groups of genes, based on cohort, CNV type and expression

509 in placenta (all genes and placenta expressed genes). Results were filtered hierarchically, and

510 only the most significant ontology per parent group was shown. *De novo* CNVs with gains and

511 losses were analyzed separately.

512

513 **Validation by ddPCR.** We applied ddPCR to validate low-level mosaic aberrations (**Fig. 2c** and

514 **Supplementary Table 3**). We used a proof-of-principle assay to evaluate the sensitivity of

515 TaqMan-based ddPCR. A DNA sample from a trisomy 21 (copy number, CN = 3) cell line was

516 mixed with a DNA sample from a normal diploid cell line (CN = 2) at different ratios, creating

517 admixture series of DNA samples with 100%, 75%, 50%, 25%, 10–15% and 0% abnormal

518 alleles (**Extended Data Fig. 6**). The cell lines of EBV-immortalized lymphocytes had been

519 established in-house using blood samples from patients of Centre for Human Genetics, UZ

520   Leuven, Leuven, Belgium, and were authenticated by conventional karyotyping. Mycoplasma

521   PCR was performed to ensure that the cell lines are not contaminated[48]. To evaluate the level of

522   Chr 21 mosaic gain, the FAM-fluorescent TaqMan copy number assay probe for the *RUNX1*

523   gene locus (dHsaCP1000308) was used as a target probe. The HEX-fluorescent TaqMan copy

524   number reference assay *AP3B1* probe (dHsaCP2500348) was used as an internal reference

525   control (Bio-Rad, Hercules, CA).

526   The ddPCR system was operated according to the manufacturer's instructions. Briefly,

527   before droplet generation, restriction enzyme HaeIII (New England Biolabs Inc.) was added

528   directly to the ddPCR reaction mixture and incubated for 20 min at room temperature.

529   Subsequently, droplets were generated in a DG8 disposable droplet generator cartridge using a

530   QX100 Droplet Generator (Bio-Rad). PCR amplification was performed by using the following

531   conditions: 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s and 59 °C for 1 min and,

532   finally, 98 °C for 10 min. Droplets were counted by using the QX100 Droplet Reader. Data were

533   analyzed by using QuantaSoft Software (Bio-Rad).

534   Mosaic ratio was calculated based on the total positive signal counts of *RUNX1*

535   normalized against reference *AP3B1* positive signal counts. All analyzed samples were run in

536   four-replicate reactions. Mosaic placental DNA samples were processed in the same manner by

537   performing initial DNA digestion, followed by droplet generation and PCR amplification. All

538   DNA samples were quantified by using the TaqMan copy number assay target probes for the

539   specific gene in the region of interest. The *AP3B1* probe was used as an internal reference

540   control (**Supplementary Table 3**).

541

542   **Code availability.** Custom code is available upon request.

543

Data availability. All SNP array data generated in this study were deposited in the

NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession

number GEO: GSE93353.

547

Methods-only References

550   40.   Sildver, K., Veerus, P., Lang, K. Birth weight percentiles and factors associated with
551   birth weight: a registry-based study in Estonia. *(Sünnikaalukõverad Eestis ja sünnikaalu*
552   *mõjutavad tegurid: registripõhine uuring) Eesti Arst* **94**, 465-470 (2015).
553   41.   Sankilampi, U., Hannila, M.L., Saari, A., Gissler, M. & Dunkel, L. New population-
554   based references for birth weight, length, and head circumference in singletons and twins from
555   23 to 43 gestation weeks. *Ann Med* **45**, 446-454 (2013).
556   42.   Conlin, L.K.*, et al.* Mechanisms of mosaicism, chimerism and uniparental disomy
557   identified by single nucleotide polymorphism array analysis. *Hum Mol Genet* **19**, 1263-1275
558   (2010).
559   43.   Nilsen, G.*, et al.* Copynumber: Efficient algorithms for single- and multi-track copy
560   number segmentation. *BMC Genomics* **13**, 591 (2012).
561   44.   Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G. & Newman, T.B. *Designing*
562   *Clinical Research*, (Lippincott Williams & Wilkins, Philadelphia, PA 19103 USA, 2015).
563   45.   Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, (Lawrence Erlbaum
564   Associates, United States of America, 1988).
565   46.   Metsalu, T.*, et al.* Using RNA sequencing for identifying gene imprinting and random
566   monoallelic expression in human placenta. *Epigenetics* **9**, 1397-1409 (2014).
567   47.   Reimand, J.*, et al.* g:Profiler-a web server for functional interpretation of gene lists (2016
568   update). *Nucleic acids research* **44**, W83-89 (2016).
569   48.   Spaepen, M., Angulo, A.F., Marynen, P. & Cassiman, J.J. Detection of bacterial and
570   mycoplasma contamination in cell cultures by polymerase chain reaction. *FEMS Microbiol Lett*
571   **78**, 89-94 (1992).