In vitro fertilization does not increase the incidence of *de novo* copy number alterations in fetal and placental lineages

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], Olga Žilina[7], Katre Teearu[7], Heidi Marjonen[9], Hanna Kahila[10], Jeroen Meekels[1], Viveca Söderström-Anttila[11], Anne-Maria Suikkari[11], Aila Tiitinen[10], Reedik Mägi[8], Sulev Kõks[12,13], Nina Kaminen-Ahola[9], Ants Kurg[7], Thierry Voet[3,14,17*], Joris Robert Vermeesch[3,17*] & Andres Salumets[5,8,10,15,17*]

Affiliations

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

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

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

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

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

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

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

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

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

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

[11] Väestöliitto Fertility Clinic, Helsinki, Finland

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

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

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

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

[16] These authors contributed equally

[17] Co-senior authors

* Correspondence should be addressed to: M.Z.E. (masoud.zamaniesteki@mumc.nl); Th.V. (thierry.voet@kuleuven.be); J.R.V. (joris.vermeesch@kuleuven.be) and A.S. (andres.salumets@ccht.ee)

Although chromosomal instability (CIN) is a common phenomenon in cleavage-stage embryogenesis following *in vitro* fertilization (IVF)[1-3], its rate in naturally conceived human embryos is unknown. CIN leads to mosaic embryos that contain a combination of genetically normal and abnormal cells and is significantly higher in *in vitro* produced preimplantation embryos as compared to in vivo-conceived preimplantation embryos[4]. Even though embryos with CIN-derived complex aneuploidies may arrest between the cleavage- and blastocyst-stage of embryogenesis[5,6], a high number of embryos containing abnormal cells can pass this strong selection barrier[7,8]. However, prevalence and extent of CIN during prenatal development and at birth, following IVF treatment, is not well understood. Here we profiled the genomic landscape of fetal and placental tissues postpartum from IVF and naturally conceived children to investigate the prevalence and persistence of large genetic aberrations that have likely arisen from IVF-related CIN. We demonstrate that CIN is not preserved at later stages of prenatal development, and that *de novo* numerical aberrations or large structural DNA imbalances occur at similar rates in IVF and naturally conceived live-born neonates. Our findings affirm that human IVF treatment has no detrimental effect on the chromosomal constitution of fetal and placental lineages.

CIN is a common event in human cleavage-stage embryogenesis that results in a high frequency of chromosomal mosaicism in IVF embryos[1-3] and likely provides an explanation for the low success rate per embryo transfer of IVF[9]. Strikingly, most IVF embryos are mixtures of euploid and aneuploid cells and only a low fraction of IVF embryos contain a normal karyotype in all blastomeres[1,10]. Between the cleavage- and blastocyst-stage of embryogenesis, CIN may cause

complex aneuploidies and the arrest of many embryos[5,6]. Even after this substantial selection, genetically aberrant cells can be preserved at high frequency in the developing blastocysts, such that about 40% of blastocysts have been reported to be uniformly euploid[7,8]. Despite the high prevalence of human IVF embryos with aberrant chromosomal configurations, its clinical implications are poorly understood. Recently, we established *Bos taurus* as a model organism to study CIN at cleavage stage embryogenesis[11] and showed that CIN in *in vitro*-produced embryos is significantly higher in prevalence when compared to *in vivo*-conceived embryos[4], rising the alarming concern about the suboptimal *in vitro* environment on genetic integrity of embryos. However, the fate of aberrant cells after cleavage-stage embryogenesis is not clear and it remains impossible to investigate naturally conceived human embryos *in vivo*.

Cell lineages containing genetic aberrations in mosaic cleavage-stage embryos have been speculated to survive in the trophectoderm, having no or little influence on the composition of the inner cell mass and its subsequent fetal development[12]. This possibility is in accordance with the confined placental mosaicism found in approximately 1–2% of ongoing natural pregnancies[13] and the low prevalence of genetic mosaicism in postnatal and adult DNA samples[14]. Furthermore, euploid blastomeres may outgrow blastomeres with chromosomal aberrations[15], allowing normal embryonic development and birth. This view of preferential selection of normal cells is supported by recent studies reporting the birth of healthy babies after intrauterine transfer of genetically mosaic IVF embryos[16,17]. In addition, a recent report described lineage-specific fate in chimeric mouse embryos. This report demonstrated that different mechanisms act on aneuploid cells in fetal or placenta lineages, such that apoptotic depletion of aneuploid cells occurs in fetal lineage, whereas aneuploid cells in placenta persist but slow down their divisions[18].

To investigate the persistence of IVF-related CIN, we profiled *de novo* DNA copy number alterations in DNA samples extracted from cord blood cells and placental tissue, which are derived from embryonic and extra-embryonic cell lineages, respectively, of both IVF and naturally conceived neonates using haplarithmisis[3] and molecular karyotyping. Importantly, the transferred embryos of IVF newborns were not screened nor counter selected for chromosomal aneuploidy status before intrauterine transfer. Following live birth, we scrutinized genome-wide single nucleotide polymorphism (SNP) profiles in DNA from the father, mother, placenta and neonate umbilical cord blood of 111 families (quartets; **Fig. 1a**), of which 49 and 62 quartets were from IVF and natural pregnancies, respectively. In the IVF cohort, the infertile patients underwent fresh or frozen IVF (n = 32) or intracytoplasmic sperm injection (ICSI, n = 17) embryo transfer, with about 4 out of 5 embryos transferred on days 2/3 (n=38, 77.5%) and the remaining on days 5/6 at blastocyst stage (n=11, 22.5%). These quartets were further subdivided based on birthweight at delivery, with 48 IVF and 50 natural conception families categorized as 'appropriate for gestational age' (AGA), and one IVF and 12 natural conception families as 'small for gestational age' (SGA) (**Supplementary Tables 1** and **2**). DNA copy number variants (CNVs) were classified by whether they were found exclusively in a single sample or a few to all samples within a quartet. This approach empowered our analysis and allowed us to differentiate inherited CNVs (present in either or both parents, umbilical cord blood and placenta), *de novo* CNVs (present in placenta and umbilical cord blood only) and placental *de novo* CNVs (present in placenta only). CNVs were detected by the combined use of logR values, B Allele Frequency (BAF) and parent-of-origin haplotypes (haplarithms). The latter analysis allowed us to account 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

Extended Data Fig. 3). This result indicates that the mosaic Chr 2 trisomy arose as CIN in the early cleavage-stage embryo, but segregated into the placental lineage only (**Fig. 3** and **Table 1**), because it was present across different parts of the placenta with the same degree of mosaicism. The presence of the low-degree mosaic Chr 2 trisomy reinforces an earlier observation in mouse embryos, which suggested a progressive reduction of aneuploid cell lineage in placenta[18] because only 10% of the cells carried the anomaly. However, the segmental trisomies in family P172 were present in only one of the placental locations, indicating that these genomic alterations either emerged later in gestation through local clonal expansion of aberrant cells in the placenta or appeared early in development, contributed to the trophectoderm and were largely overgrown by normal cells in placental tissue. The presence of multiple *de novo* DNA copy number variants has recently been shown to lead to genomic disorders[19]. The finding that multiple CNVs are present with a different degree of mosaicism, albeit in placenta genome only, is indicative that they occurred in a cumulative fashion[20].

To further investigate CNV heterogeneity across the placenta and reaffirm the (non-) mosaic nature of the detected *de novo* CNVs, we further analysed 20 DNA samples derived from spatially different locations of all but two placentas (F351 and F1021) with *de novo* CNVs (**Table 1**). As expected, *de novo* non-mosaic CNVs found in both placenta and cord blood were consistently present in all the biopsies, whereas *de novo* mosaic CNVs were not present in the other biopsies (**Extended Data Fig. 4**). Importantly, the full Chr 2 mosaic trisomy was found in all the biopsies (**Extended Data Fig. 3**), indicating its early embryonic origin (see above).

We identified two clinically significant non-mosaic CNVs in the SGA cohort, with the genetic aberrations found both in umbilical cord and placental DNA. P153 carried a *de novo* 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

high expression according to the whole RNA sequencing of placental tissue (**Fig. 2d** and **Supplementary Table 4**), and genes from regions with *de novo* gains (18.0%, 77/428) and losses (19.1%, 9/47, $P = 0.84$ odds ratio test) demonstrated similar proportion of moderate-to-high expression genes in placental tissue. Moreover, enrichment analyses of genes located in *de novo* CNV regions did not reveal any molecular pathway related to placental biology. All this indicates that these new CNVs are likely sporadic and do not play any role in placental biology.

In this study, our main outcome measure was the extent and prevalence of *de novo* large CNVs (>100 kb) in live-born IVF-conceived neonates when compared to naturally conceived ones. Although we have not studied the embryonic loss with or without aneuploidies in IVF- or naturally conceived pregnancies, we speculate that when all cells of day-3 (cleavage-stage) embryos are investigated for the presence of (segmental) aneuploidies, on average 80% of the embryos present with at least one aneuploid cell (range 70%-90%)[1,3,10]. The IVF success rate per embryo transfer of day-3 embryos is about 35% in Helsinki and Tartu IVF centres. Assuming that all euploid day-3 embryos survive, they would constitute at most 20% of the total number of embryos transferred resulting in pregnancy/delivery, but indicating that at least 15% of all embryos did contain some aneuploid cells. This would amount to 60% and 40% of the live births resulting from respectively euploid and mosaic euploid/aneuploid cleavage-stage embryos. Conversely, about 20% (15/80%) of mosaic embryos survive and result in healthy live births, whereas 80% (65/80%) are selected against. Since it is unlikely that all euploid embryos make it to term, the latter is a lower estimate.

Mosaicism rates in human IVF embryos are still debated and vary across studies primarily because of the definition of mosaicism used, the technology applied, and the variance 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].

However, in this study, the IVF procedure, including culture conditions, were similar in the participating IVF clinics.

Taken together, our findings suggest that the high level of chromosomal mosaicism in IVF cleavage-stage embryos is not preserved at later stages of prenatal development. Although we found a low-level mosaic trisomy in one IVF placental genome, it was not present in the fetal lineage of the same pregnancy, suggesting a proliferative defect of trisomic cells. The sporadic *de novo* genomic aberrations found in approximately 10.8% of pregnancies were scattered across the genome and represented random and rare events of embryonic CIN, without any observed functional consequence on placental biology or fetal health. Thus, our findings confirm that the IVF procedure has no detrimental effect on large genomic rearrangements in fetal and placental lineages of live-born neonates, which helps to remove a major health concern for IVF neonates.

ACKNOWLEDGMENTS

We gratefully thank all families that participated in this study in Estonia and Finland. This research was funded by institutional research grant IUT34-16 from the Estonian Ministry of Education and Research to A.S; grant EU48695 from Enterprise Estonia to A.S; grant EU692065 from Horizon 2020 innovation programme (WIDENLIFE) to A.K; grant EU324509 from European Union's FP7 Marie Curie Industry-Academia Partnerships and Pathways programme to A.S; Helsinki University Hospital Fund and Faculty of Medicine, University of Helsinki to N.K-A; EVA (Erfelijkheid Voortplanting & Aanleg) specialty program fund of Maastricht University Medical Centre (MUMC+) to M.Z.E; Estonian Research Council grant IUT20-60, IUT24-6 and European Union through the European Regional Development Fund Project No. 2014-2020.4.01.15-0012 GENTRANSMED and 2014-2020.4.01.16-0125 to R.M; and KU

Leuven funding (C1/018) and FWO grant G.0392.14N to J.R.V and Th.V. We thank B. de Greef, A. van Montfoort and N. Davarzani for statistical consultations.

AUTHOR CONTRIBUTIONS

M.Z.E., Tr.V., A.K., Th.V., J.R.V. and A.S. conceived the study and designed the experiments. 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., 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. and M.Z.E. performed ddPCR assays. M.K. carried out RNA sequencing analysis. M.N., K.T., O.Z. and R.M. performed PennCNV and QuantiSNP analyses. M.Z.E. drafted the initial version of the manuscript. M.Z.E., Tr.V., A.K., Th.V., J.R.V. and A.S. wrote and edited the manuscript. M.Z.E., Th.V., J.R.V. and A.S. jointly supervised this study. All the authors read and approved the manuscript for submission.

COMPETING FINANCIAL INTERESTS

M.Z.E., J.R.V., and Th.V. are co-inventors on a patent applications ZL913096-PCT/EP2014/068315-WO/2015/028576 ‘‘Haplotyping and copy-number typing using polymorphic variant allelic frequencies’’.

References

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

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

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

Figure Note

Figure legends

Figure 1 | Haplarithmisis reveals genetic mosaicism with parent-of-origin information. (**a**) Schematic representation of a quartet. (**b**) Haplarithmisis can determine the maternal DNA contribution to the bulk placental DNA sample (see **Methods**). (**c**) A maternally inherited CNV (red arrows) and placental *de novo* CNV with paternal origin found on Chr 5 of P070 (purple arrow). B allele fraction (BAF) of a SNP is the signal intensity of allele B over signal intensities of both expected alleles for that SNP ($BAF = (\frac{B}{A+B})$), and logR (relative copy number) of that SNP is the base 2 logarithm of the summed normalized signal intensities of both alleles over expected signal intensity values ($logR = \frac{R_{Sample}}{R_{expexted}}$). (**d**) Overview of *de novo* CNVs found in this study (see also **Source data**). G and L denote gains and losses, respectively, as shown in **Table 1**.

Figure 2 | Mosaic *de novo* CNVs and overlap with placental transcriptome. Haplarithms revealed *de novo* mosaic (**a**) partial trisomies in P172 placenta of paternal origin (see also **Extended Data Fig. 2**) and (**b**) full-chromosome trisomy in P106 placenta of maternal origin (see also **Extended Data Fig. 3**). These aberrations are indicated by purple arrows. (**c**) Validation of detected mosaic *de novo* CNVs by ddPCR (each circle and error bar indicate mean and standard deviation, respectively, of four independent measurements). (**d**) Placental expression of genes detected in *de novo* CNV regions, expressed as reads per kilobase per million mapped reads (RPKM).

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.

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

374

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

(**Table 1** and **Extended Data Figs. 2**, **3** and **4**). All samples were collected by the same medical personnel. Maturity and health of term placenta were confirmed by histological examination. A total of 9 mL of peripheral venous blood from parents and 4 mL of umbilical cord blood were collected into ethylenediaminetetraacetic acid-containing tubes (BD Vacutainer, Becton, Dickinson and Co.). In Estonia, DNA was isolated from blood and placental tissue by using the PureLink Genomic DNA Kit (Invitrogen, Life Technologies). DNA samples were eluted with 200 µL of PureLink Genomic Elution Buffer (10 mM Tris-HCl, pH 9.0, 0.1 mM EDTA) and stored at -20 °C until analysis. Finnish parental DNA from blood samples was extracted by NucleoSpin Blood XL Kit (Macherey-Nagel) and DNA from umbilical cord blood by NucleoSpin Tissue Kit (Macherey-Nagel). NanoDrop 2000 (Thermo Fisher Scientific Inc.) was used to assess DNA concentration and purity.

SNP genotyping. For SNP genotyping, the Infinium PsychArray-24 v1.1 (Illumina; GEO: GPL22819) and PsychArray-24 v1.3 BeadChip SNP arrays (Illumina; GEO: GPL26750) were used, which contains approximately 590,000 SNP markers with a median marker spacing of 5.5 kb. Genotype calls, SNP B-allele frequency values and logR values of all samples were computed by the Illumina GenomeStudio software.

Haplarithmisis. We applied haplarithmisis[3] on each quartet DNA sample. Discrete SNP calls derived from umbilical cord blood were used as seeds to phase the parental genotypes. Parental haplarithms were used to infer DNA copy number state, parent-of-origin and level of mosaicism of *de novo* copy number alterations in placental/newborn's DNA. Level of mosaicism was calculated as described[42].

Haplarithmisis can be used to estimate contributions of maternal and fetal genomes in placental DNA samples. We carried out a simulation analysis by blending maternal and fetal alleles with different proportions (from $1\%_{Mother}$: $99\%_{Child}$ to $99\%_{Mother}$: $1\%_{Child}$) and applied haplarithmisis (**Extended Data Fig. 5**). This process allowed us to determine accurately the maternal and fetal contributions to placental tissue (**Supplementary Table 5** and **Extended Data Fig. 1**).

Copy number profiling. We applied both PennCNV[36] (with 0.2% false positive rate and 86% sensitivity for constitutional CNVs) and QuantiSNP[37] (with 0.1% false positive rate for constitutional CNVs) algorithms as well as the copy number module of siCHILD[3] on the raw logR values with minor modifications. Raw logR-values were smoothed by using a moving average window of five consecutive SNP probes, wave-corrected for GC% bias by a Lowess fit and normalized to a trimmed mean of normal diploid chromosomes. Normalized logR-values were segmented by piecewise-constant fitting[43] (gamma = 14). A copy number alteration was considered reliable if more than six consecutive SNP probes supported the copy number change, its length was more than 33 kb (six times 5.5 kb median distance between two probes) and it was supported by raw or segmented parent-of-origin specific haplotypes, coined haplarithms. As a sanity check, we investigated whether there are signals that are detected only by haplarithmisis (**Supplementary Table 6**) or logR analysis (**Supplementary Table 7**). We did not include those in the analysis.

Copy number analysis was empowered by our quartet sample-set and haplarithmisis, as we used all of the samples within a quartet to call a reliable genomic change. For each of the detected DNA copy number alterations in one sample (e.g. placenta), we checked whether the

same alteration was present in other samples of the same quartet. This approach allowed us to determine the parental origin of inherited CNVs, to detect *de novo* CNVs (i.e. in placental samples or in both placental and umbilical cord samples), to find CNVs exclusive to any of the samples (e.g. present solely in the placental genome) and to reveal genotyping artifacts.

Other statistical and computational analysis. Since both predictor (with or without IVF treatment) and outcome (with or without *de novo* CNVs) are dichotomous, chi-squared test is the most suitable test[44]. To determine appropriate sample size, we performed a power analysis using chi-squared test with confidence level of at least 95% and feeding the conventional effect size (w) values of 0.1, 0.3 and 0.5, which correspond to small, medium and large effect sizes, respectively[45]. With a medium effect size at least 40 samples per cohort is required to reach to the power of 0.8[24]. Given the sample size of about 50 per cohort (50 naturally conceived quartets and 48 IVF quartets with appropriate birthweight for gestational age), this study has >80% power (alpha = 0.05; middle effect size of 0.3; case to control ratio ca 1) to detect a significant difference (<0.05).

We also compared large CNVs with mosaic or non-mosaic nature separately. The rate of mosaic large CNVs found in naturally conceived versus IVF-conceived AGA groups were not significantly different ($P = 0.157$ Pearson's chi square test). Similarly, the rate of non-mosaic CNVs found in naturally conceived versus IVF-conceived AGA groups were not significantly different ($P = 0.977$ Pearson's chi square test).

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

Protein-coding genes from *de novo* CNV regions were extracted from the UCSC browser (https://genome.ucsc.edu/). Human genome build GRC37/hg19 was used as a reference. Genes found at *de novo* CNV regions were studied for gene expression in placental tissue from 12 individuals. Enrichment analyses for Gene Ontology terms were performed for genes present at *de novo* CNVs, by using g:Profiler (version Ensembl 90, Ensembl Genomes 37, rev 1741, build date 2017-10-19)[47] with default settings. Briefly, this tool performs gene set enrichment analyses for Gene Ontology, Human Phenotype Ontology and Kyoto Encyclopedia of Gene and Genomes gene sets. Enrichment analyses are conducted by hypergeometric test and resulting enrichment P-values are further adjusted for multiple testing by g:SCS (Set Counts and Sizes) method, developed to address the non-independent structure of tested GO terms[47]. Separate enrichment analyses were conducted to different groups of genes, based on cohort, CNV type and expression in placenta (all genes and placenta expressed genes). Results were filtered hierarchically, and only the most significant ontology per parent group was shown. *De novo* CNVs with gains and losses were analyzed separately.

Validation by ddPCR. We applied ddPCR to validate low-level mosaic aberrations (**Fig. 2c** and **Supplementary Table 3**). We used a proof-of-principle assay to evaluate the sensitivity of TaqMan-based ddPCR. A DNA sample from a trisomy 21 (copy number, CN = 3) cell line was mixed with a DNA sample from a normal diploid cell line (CN = 2) at different ratios, creating admixture series of DNA samples with 100%, 75%, 50%, 25%, 10–15% and 0% abnormal alleles (**Extended Data Fig. 6**). The cell lines of EBV-immortalized lymphocytes had been established in-house using blood samples from patients of Centre for Human Genetics, UZ

Leuven, Leuven, Belgium, and were authenticated by conventional karyotyping. Mycoplasma PCR was performed to ensure that the cell lines are not contaminated[48]. To evaluate the level of Chr 21 mosaic gain, the FAM-fluorescent TaqMan copy number assay probe for the *RUNX1* gene locus (dHsaCP1000308) was used as a target probe. The HEX-fluorescent TaqMan copy number reference assay *AP3B1* probe (dHsaCP2500348) was used as an internal reference control (Bio-Rad, Hercules, CA).

The ddPCR system was operated according to the manufacturer's instructions. Briefly, before droplet generation, restriction enzyme HaeIII (New England Biolabs Inc.) was added directly to the ddPCR reaction mixture and incubated for 20 min at room temperature. Subsequently, droplets were generated in a DG8 disposable droplet generator cartridge using a QX100 Droplet Generator (Bio-Rad). PCR amplification was performed by using the following conditions: 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s and 59 °C for 1 min and, finally, 98 °C for 10 min. Droplets were counted by using the QX100 Droplet Reader. Data were analyzed by using QuantaSoft Software (Bio-Rad).

Mosaic ratio was calculated based on the total positive signal counts of *RUNX1* normalized against reference *AP3B1* positive signal counts. All analyzed samples were run in four-replicate reactions. Mosaic placental DNA samples were processed in the same manner by performing initial DNA digestion, followed by droplet generation and PCR amplification. All DNA samples were quantified by using the TaqMan copy number assay target probes for the specific gene in the region of interest. The *AP3B1* probe was used as an internal reference control (**Supplementary Table 3**).

Code availability. Custom code is available upon request.

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.

Methods-only References

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