HGG Advances

ARTICLE

Whole-genome sequences reveal zygotic composition in chimeric twins

Christopher J. Yoon,[1,2,3,4] Chang Hyun Nam,[2] Taewoo Kim,[2] Jeong Seok Lee,[2,5] Ryul Kim,[2,5] Kijong Yi,[2,5] June-Young Koh,[2,5] Jiye Kim,[2] Hyein Won,[2] Ji Won Oh,[6] Obi L. Griffith,[3,4] Malachi Griffith,[3,4] Joohon Sung,[7] Tae Yeul Kim,[8] Duck Cho,[8,*] Ji Seon Choi,[9,*] and Young Seok Ju[1,2,10,*]

Summary

While most dizygotic twins have a dichorionic placenta, rare cases of dizygotic twins with a monochorionic placenta have been reported. The monochorionic placenta in dizygotic twins allows *in utero* exchange of embryonic cells, resulting in chimerism in the twins. In practice, this chimerism is incidentally identified in mixed ABO blood types or in the presence of cells with a discordant sex chromosome. Here, we applied whole-genome sequencing to one triplet and one twin family to precisely understand their zygotic compositions, using millions of genomic variants as barcodes of zygotic origins. Peripheral blood showed asymmetrical contributions from two sister zygotes, where one of the zygotes was the major clone in both twins. Single-cell RNA sequencing of peripheral blood tissues further showed differential contributions from the two sister zygotes across blood cell types. In contrast, buccal tissues were pure in genetic composition, suggesting that *in utero* cellular exchanges were confined to the blood tissues. Our study illustrates the cellular history of twinning during human development, which is critical for managing the health of chimeric individuals in the era of genomic medicine.

Introduction

Sexually reproducing organisms begin as a single diploid zygote (2n) fertilized from a haploid sperm (n) and a haploid oocyte (n). The DNA from the zygote is then replicated as the embryo continues to divide and produce all the cells of an organism. In most individuals, every cell in an organism can be traced back to a single cell, the zygote. Chimeras, in which an individual is composed of cells with multiple genomic constitutions traced to more than one distinct diploid zygote, are an exception to the norm (Figure 1A). In cattle, chimerism is frequently observed when multiple dizygotic bovine fetuses are conceived.[1,2] Blood anastomosis in the monochorionic dizygotic (MCDZ) placenta allows the transfusion of cells and hormones between the two fetuses, often resulting in the generation of freemartin, an infertile female cattle. In humans, chimerism is rare, as most human dizygotic twins have a dichorionic placenta that prohibits exchanges between the two twins.[3]

A few human MCDZ chimeric individuals have been reported.[4–8] Sex chromosomes[9–11] or short tandem repeat markers[8–10] have been utilized to evaluate the clonal composition of twins showing chimerism. However, these approaches lack the sensitivity and precision required to thoroughly assess the clonal composition of twins. Utilizing genome-wide variants can overcome these limitations and more comprehensively reveal the cellular history of chimerism. In the present study, we thoroughly investigated the inheritance patterns of millions of polymorphisms using whole-genome sequencing (WGS). We applied statistical evaluation to accurately decompose zygotic compositions at a single-cell and gamete resolution.[12] Furthermore, single-cell sequencing was used to trace the developmental outcomes of the chimeric cells in each twin.

Subjects and methods

Study participants

Two families (Family A and Family B) with known chimerism identified by ABO genotyping were recruited for the study.[5] Peripheral blood samples were obtained for all individuals. Buccal swab samples were also obtained from the chimeric twins in each family. In Family A, a healthy girl (Child-0) was conceived in the first pregnancy and triplets (Child-1, Child-2, and Child-3) were conceived in the second pregnancy. Family A's second pregnancy was conceived using assisted reproductive technology involving an ovulation induction with clomiphene citrate treatment. The triplets in Family A were 5 years old at the time of WGS and 6 years old at the time of peripheral blood mononuclear cell (PBMC) and buccal epithelial laser capture microdissection (LCM) isolation studies. The twins in Family B

[1]Research Center for Natural Sciences, Korea Advanced Institute of Science and Technology, Daejeon, Korea; [2]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea; [3]Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; [4]McDonnell Genome Institute, St. Louis, MO, USA; [5]Inocras, Inc., San Diego, CA, USA; [6]Department of Anatomy, Yonsei University College of Medicine, Seoul, Republic of Korea; [7]Department of Epidemiology, School of Public Health, Seoul National University, Seoul, Korea; [8]Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; [9]Department of Laboratory Medicine, International St. Mary's Hospital, Catholic Kwandong University College of Medicine, Incheon, Korea
[10]Lead contact
*Correspondence: duck.cho@skku.edu (D.C.), sunnychoi@ish.ac.kr (J.S.C.), ysju@kaist.ac.kr (Y.S.J.)
https://doi.org/10.1016/j.xhgg.2024.100301.

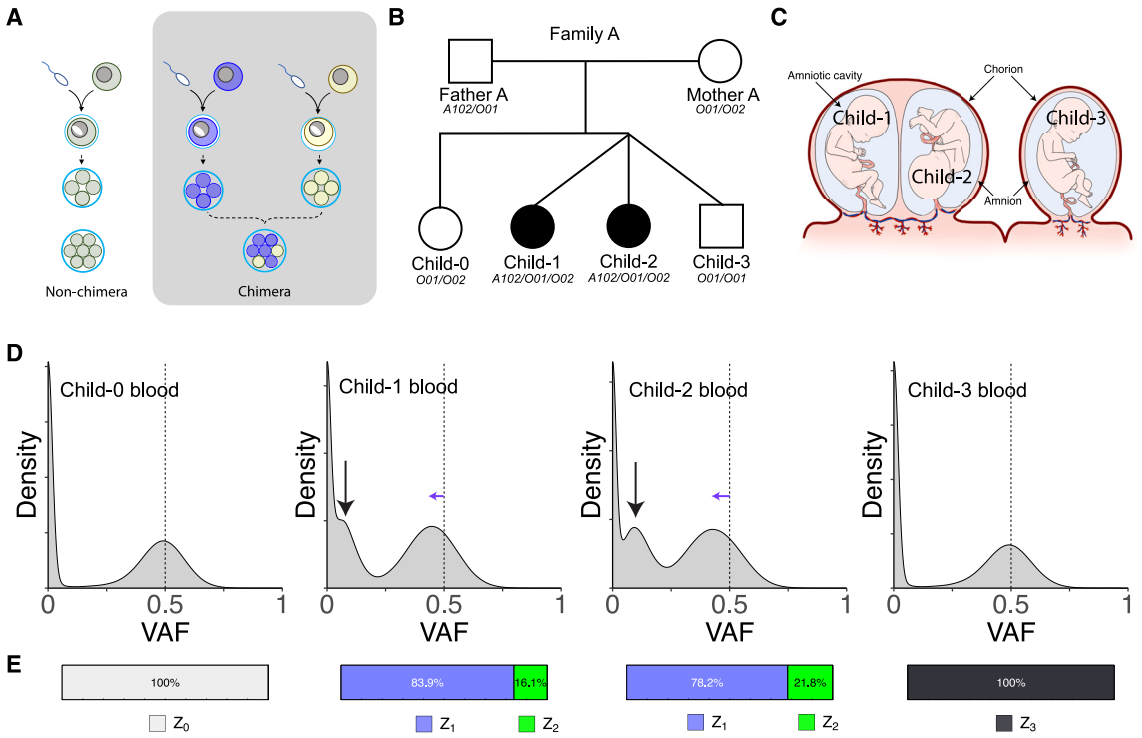

Figure 1. Quantification of hematopoietic chimerism in MCDZ twins

(A) Schematic of non-chimera and chimera. Non-chimeric individuals are traced to one common zygote, while chimeric individuals are traced back to two zygotes.

(B) A pedigree of Family A with chimera. Filled circles in Child-1 and Child-2 indicate the chimerism identified. ABO blood genotypes are shown on the bottom.

(C) The triplets' fetal membranes (chorion and amnion) in Family A. Child-1 and Child-2 showed monochorionic diamniotic configuration. Child-3 had a separate chorion and amnion.

(D) VAF density of SNPs in the children. The dotted lines indicate VAF = 0.5 is expected for a heterozygous variant. Black vertical arrows indicate an extra peak shown in the chimeric individuals. The blue horizontal arrows indicate the peak's left shift due to the merging heterozygous SNP peak (VAF = 0.5) and an additional peak from chimeric SNPs.

(E) The bar plots indicate the zygotic contribution estimated from the SNP read counts. MCDZ, monochorionic dizygotic; SNP, single-nucleotide polymorphism; VAF, variant allele frequency.

were conceived with *in vitro* fertilization followed by the transfer of two blastocyst stage embryos. Child-21 and Child-22 in Family B were 1 year old at the time of WGS.

Ethics approval and consent to participate

The study protocol was approved by the Institutional Review Boards of the Korea Advanced Institute of Science and Technology (IRB KH2019-174), International St. Mary's Hospital (IRB IS19TIME0070), and the Samsung Medical Center (IRB SMC-2019-01-049-005).

WGS data generation

DNA was extracted from the peripheral blood samples of all family members. Buccal swabs were obtained from the chimeric individuals. DNA libraries were generated according to Illumina's Truseq PCR free library protocol. Paired-end sequencing reads of WGS were obtained using NovaSeq (blood: 30×, buccal: 60×). Raw sequence files were aligned to the human reference genome GRCh38 using the BWA mem.[13]

Investigation of inherited heterozygous loci

Germline single-nucleotide polymorphisms (SNPs) and small insertions or deletions were used to estimate the level of chimerism in the children. Briefly, for an informative locus where one of the parents was a homozygous reference (ref/ref) and the other was a heterozygous variant (ref/var), there were two possible genotypes in the daughter embryo (homozygous reference and heterozygous variant). In genetically homogeneous offspring, the variant allele frequencies (VAFs) of the locus in sequencing is either 0% (ref/ref) or 50% (ref/var). In contrast, for chimeric offspring, homozygous reference and heterozygous variant cells co-existed in a single individual, leading to a dispersion of VAFs between 0% and 50%. We used TrioMix, a maximum likelihood estimation framework, to quantify chimerism in the twins using the inheritance pattern of common SNPs in GRCh38.[12]

Meiotic recombination

Meiotic recombination was identified by tracing the parental haplotype in the non-chimeric siblings (Child-0 and Child-3) and the monochorionic twin sisters (Child-1 and Child-2). The circular binary segmentation method[14] was used to estimate the recombination sites. All recombination sites were manually reviewed, and parental haplotypes were inferred from the recombination patterns using the most parsimonious recombination breakpoints between the siblings.[15]

Non-inherited variant discovery

Non-inherited variants (*de novo* mutations and postzygotic mutations) were identified using Varscan2[16] and in-house scripts from our previous reports.[17] Non-inherited variants were compared between Child-0, Child-1, Child-2, and Child-3. All *de novo* mutations shared between the children were manually reviewed and compared against meiosis recombination patterns to infer the parental origin of shared *de novo* mutations.

LCM and low-input library preparation

Cells obtained from the buccal swab were placed on membrane-covered microscope slides (Carl Zeiss Membrane Slides). Cells on the microscope slides were fixed using a PAXgene fixation kit, followed by hematoxylin and eosin staining and 70% alcohol treatment. Slides were mounted to an LCM machine (Carl Zeiss, PALM MicroBeam). Single buccal epithelial cells were dissected and collected on an adhesive cap (Carl Zeiss Adhesive Caps) with approximately 13–60 cells per cap. Each adhesive cap at a density of approximately 13–60 cells was independently processed for DNA extraction using a QIAamp DNA Micro Kit (Qiagen). This approach allowed enough DNA input material to prepare sequencing libraries and detect variants unique to each zygote. The NEBNext Ultra II FS DNA library kit was used to prepare low-input DNA libraries. Paired-end WGS reads were obtained using a NovaSeq machine as described above.

10× single-cell transcriptome sequencing

PBMCs were isolated from the whole blood using Ficoll density gradient centrifugation. PBMCs of the chimeric children were then prepared using a 10× genomics 3′ gene expression kit for single-cell RNA sequencing (scRNA-seq). 10× single-cell libraries were sequenced using the Illumina NovaSeq. Demultiplexed FASTQ files were processed using CellRanger (v.6.0.1). The output matrix was used as the input for Seurat (v.4) for clustering and cell-type annotation. Cell types were mapped to a two-dimensional uniform manifold approximation and projection space of multimodal reference atlas of circulating human immune cells.[18]

Phasing variants to each zygote and deconvolution of scRNA-seq

The zygote of origin of each cell was determined using a custom script utilizing high-confidence SNP locations specific to one of the two contributing zygotes in the chimera. High-confidence SNP loci were obtained by comparing the SNP VAFs in the buccal swab whole-genome sequences with each zygote's recombination patterns. Variants were phased to one of the parental homologous chromosomes using variant information from the additional siblings (Child-0 and Child-3). We could accurately assign their genotypes because the additional siblings were non-chimeric individuals. Observation of the variant read counts follows a binomial probability (*Pr*) distribution where k is an alternative read count, n is a total read depth at a locus, and α is the probability of sampling an alternative read, which is identical to the zygote fraction for homozygous variants.

$$Pr(k,n,\alpha) = \binom{n}{k}\alpha^k(1-\alpha)^{1-k}$$

Cumulative distribution function $F(k,n,\alpha)$ provides a confidence interval (CI) for observing the read counts for a given α.

$$F(k,n,\alpha) = \sum_{x=0}^{k} Pr(x,n,\alpha)$$

For a heterozygous variant, α is 0.5. Thus, a variant was considered a heterozygous variant with a 95% confidence under binomial probability if it satisfied the following condition:

$$0.025 < F(k,n,0.5) < 0.975.$$

We required VAF = 0 for a homozygous reference genotype and the observed variant read counts to satisfy the above equation. For Child-0 and Child-3, we compared the genotypes of the two siblings and the chromosomal phasing information to assign each variant to its respective homologous chromosome in the parents.

For the two zygotes to have a homozygous variant that differs in genotype (one zygote with a homozygous alternative genotype and the other zygote with a homozygous reference genotype), the two zygotes must receive the opposite allele from both parents, which is only possible in the four-chromosome region. Let us assume that Z_1 and Z_2 fractions are α_1 and 1-α_1, respectively, in Child-1's bulk buccal WGS and 1-α_2 and α_2 in Child-2's bulk buccal WGS. A variant locus j from the bulk buccal WGS of Child-1 was observed with a total read depth n_{j1} and an alternative read count k_{j1} and of Child-2 with a total read depth n_{j2} and a read count k_{j2}. To identify the homozygous variant j specific to Z_1, but not Z_2, with a 95% confidence interval, the read counts from the two bulk buccal WGS must satisfy the following conditions:

$$0.025 < F(k_{j1},n_{j1},\alpha_1) < 0.975,$$

$$F(k_{j1},n_{j1},1-\alpha_1) < 0.025 \vee F(k_{j1},n_{j1},1-\alpha_1) > 0.975,$$

$$F(k_{j2},n_{j2},\alpha_{j2}) < 0.025 \vee F(k_{j2},n_{j2},\alpha_{j2}) > 0.975, \text{and}$$

$$0.025 < F(k_{j2},n_{j2},1-\alpha_2) < 0.975.$$

Using these statistical assignments of unique variants, we obtained 44,448 homozygous variants unique to Z_1 and 44,193 homozygous variants unique to Z_2. Homozygous variants specific to Z_2, but not Z_1, could be similarly calculated:

$$0.025 < F(k_{j2},n_{j2},\alpha_2) < 0.975,$$

$$F(k_{j2},n_{j2},1-\alpha_2) < 0.025 \vee F(k_{j2},n_{j2},1-\alpha_2) > 0.975,$$

$$F(k_{j1},n_{j1},\alpha_{j1}) < 0.025 \vee F(k_{j1},n_{j1},\alpha_{j1}) > 0.975, \text{and}$$

$$0.025 < F(k_{j1},n_{j1},1-\alpha_1) < 0.975$$

For heterozygous variants specific to each zygote, we substitute α and 1-α with $\frac{\alpha}{2}$ and $\frac{1-\alpha}{2}$ for the alternative allele frequency. We obtained 462,339 heterozygous variants specific to Z_1 and 456,409 heterozygous variants specific to Z_2.

A VCF file containing the zygote-specific heterozygous and homozygous SNPs was used as an input for Demuxlet[19] to identify singlet cell barcodes with their zygotes of origin identity. Cell barcodes containing SNPs from both zygotes were labeled as doublets or multiplets and removed from further analysis.

Quantification and statistical analysis

Maximum likelihood estimation was used to estimate the degree of chimerism using TrioMix by quantifying the deviation from the expected Mendelian inheritance patterns of the SNPs.[12] Linear regression was performed between the expected ratio of Z_1 and Z_2 and the bulk sequencing VAFs of the mitochondrial heteroplasmic variants. A chi-squared test was used to estimate the chimera's two-, three, and four-chromosome regions. A two-proportion Z-test was used to estimate the differential distribution of zygotic origins in various cell types using scRNA-seq. Poisson's probabilities were used to assess the likelihood of observing shared *de novo* mutations in genetic sibling zygotes. All statistical calculations and visualizations were conducted using the R (v.4.0) programming language.

Results

Chimerism in the blood of MCDZ twins

We explored the genome sequences of Family A (Figure 1B).[5] Here, the first pregnancy gave birth to a healthy girl (Child-0). Then, the second pregnancy gave birth to triplets (Child-1, Child-2, and Child-3), who were conceived using assisted reproduction technology involving an ovulation induction with clomiphene citrate treatment. All triplet individuals were healthy, but two girls (Child-1 and Child-2) showed mixed ABO blood types (in the mixed-field agglutination of the blood groups). They had substantially dissimilar appearances, suggesting a non-identical genetic constitution. The triplets were noted to have a dichorionic triamniotic placenta (Figure 1C), where Child-1 and Child-2 shared a single chorion (i.e., a monochorionic diamniotic configuration between Child-1 and Child-2).

To systematically understand chimerism, we quantified the level of chimerism in the peripheral blood using the WGS of the family members. Here, we investigated the variant allele frequency (VAF) of inherited SNPs, in which one of the parents had a heterozygous genotype and the other parent had a homozygous reference genotype. Under Mendelian inheritance,[12] an offspring genome will have SNPs that are either heterozygous genotypes (VAF = ~0.5) or homozygous reference genotypes (VAF = ~0) with a 50%:50% chance. Suppose that the genomes of two different zygotes are admixed; intermediate VAFs (between $0 < \text{VAF} < 0.5$ according to the mixture level) will be observed in the genomic regions where the two zygotes inherited different parental chromosomes.

The peripheral blood of Child-0 and Child-3 showed a typical VAF pattern of the non-chimeric genome, or single zygote, with two possible genotypes (VAF = ~0, homozygous reference genotype; VAF = ~0.5, heterozygous genotype; Figure 1D). In contrast, the peripheral blood of Child-1 and Child-2 showed additional peaks near VAF = ~0.1 in Child-1 and Child-2 and a left shifted peak near VAF = ~0.4, indicating chimerism in their blood, implying the presence of admixed cells of dual zygotic origins in the blood of both twins (Figure 1D).

One of the possible sources of the chimerism is the male sibling (Child-3) in the triplet pregnancy. However, this hypothesis was rejected, since Child-3 specific variants and chromosome Y were not observed in the blood DNA of Child-1 and Child-2 (Figure S1A). Instead, a comparison of the SNPs in the blood tissues of Child-1 and Child-2 suggested that the chimerism of these twins was associated with a mutual exchange of cells from two zygotic lineages, independent of Child-3. Indeed, the genomic sequences of peripheral blood tissues of Child-1 and Child-2 were successfully decomposed by two sets of zygotes (Z_1 and Z_2; Figure 1E). A maximum likelihood estimation approach was used to quantify the genomic compositions of the two sibling zygotes[12] (Z_1:Z_2 = 78.2%:21.8% and 83.9%:16.1% for Child-1 and Child-2, respectively, Figure 1E). We investigated another MCDZ twin family (Family B; Figures S2A and S2B) showing a similar pattern of zygotic contributions (Z_4:Z_5 = 91.2%:8.8% in Child-21 and 90.0%:10.0% in Child-22; Figure S2C). In both the MCDZ twins, only one of the sister zygotes was commonly dominant in the blood. This indicated that immigrant cells became a predominant population in the blood tissues of both twins.

Decomposition of haplotype-resolved zygote-specific genome sequence in the chimera

Genomic sequences of the parents and multiple zygotes (Z_0, Z_1, Z_2, and Z_3, where Z_0 and Z_3 are zygotes generating Child-0 and Child-3, respectively) in the family allowed us to phase the haplotype genome sequences in the parents and their meiotic recombination breakpoints in each zygote[15] (Figure 2A). The genomic configurations of Child-1 and Child-2 can be categorized into three groups (Figure S3A): regions with two chromosomes (Z_1 and Z_2 receiving identical chromosome segments from both parents), three chromosomes (both zygotes receiving identical chromosome segments from one of the parents but receiving opposite homologous chromosomes from the other parent), or four chromosomes (both zygotes receiving opposite chromosome segments from both parents), according to the meiotic recombination and chromosomal segregation stochastically operative in the gametogenesis. Overall, the entire autosomal genome of Child-1 and Child-2 contained 26.6% two-chromosome regions, 46.2% three-chromosome regions, and 27.2% four-chromosome regions (Figure S3B), which was close to the random expectation (25%:50%:25%; chi-squared test $p = 0.9971$). Since the two-chromosome regions represent the identical sequence between the two zygotes, the remaining 73.4% (three- and four-chromosome regions) of the autosomal genome were in a true chimeric state.

Several genes of interest were located in the chimeric regions. For example, the *HLA* genes were located in a three-chromosome region (chromosome 6), where two sets of maternal chromosomes contributed to chimerism (Figure 2B). The *ABO* gene was located in a

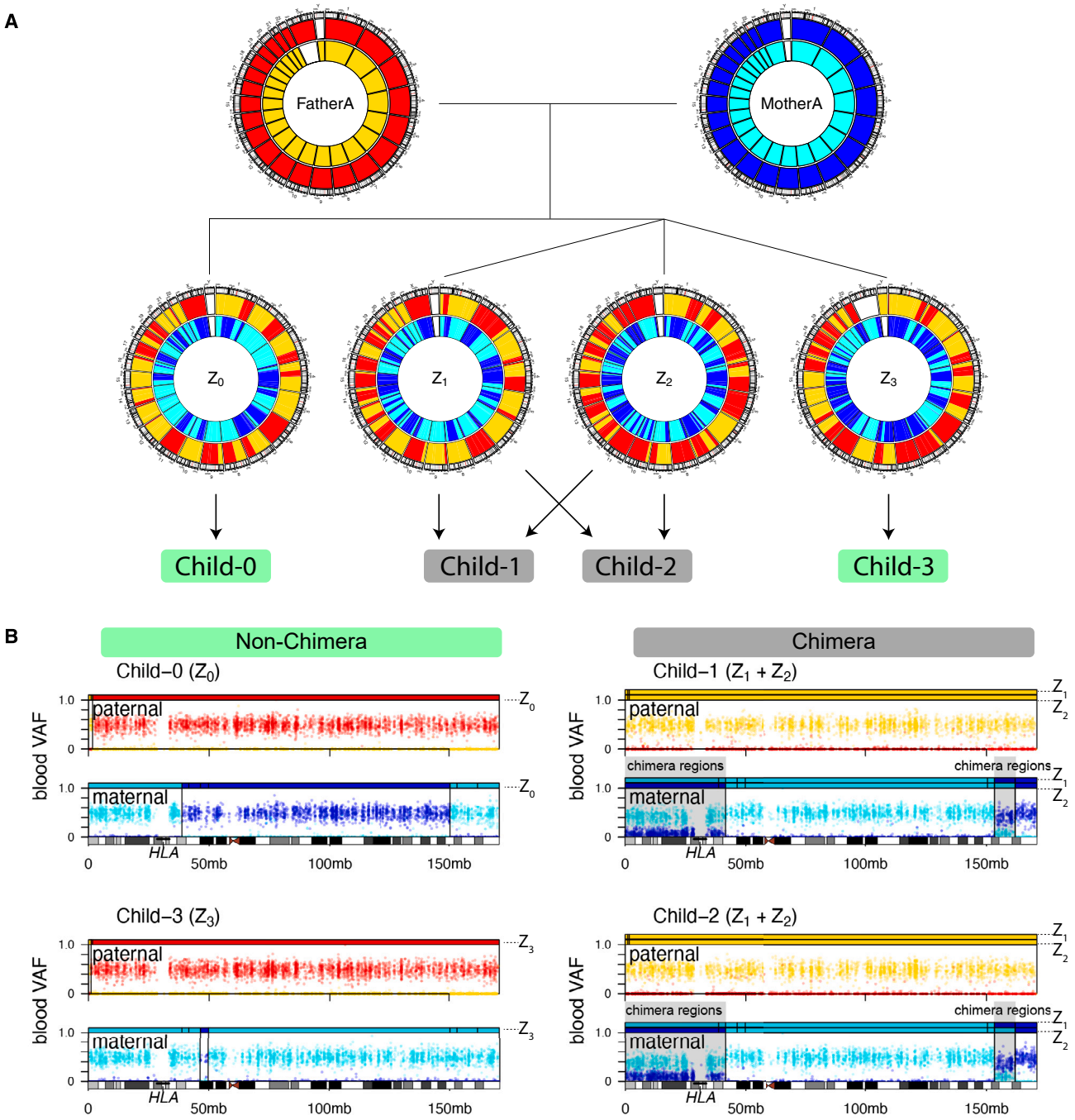

Figure 2. Haplotype-resolved diploid genomes of blood chimerism

(A) A Circos diagram of the parents and the four zygotes in Family A. Each haploid of parental chromosomes is assigned specific colors, and meiotic recombinations are seen as a color switch within the chromosomes of the zygotes. Contributing zygote(s) for each child are indicated with arrows.

(B) VAFs of each child are shown with their respective parental haploid of origin as their color (same color as A). The parental haploid genomes are also shown as rectangles over each plot. Two bars are drawn for chimeric children to indicate two originating zygotes. The regions where the zygotes inherited the same haplotype block are shown in identical colors with VAF = ~0.5. The regions where the zygotes inherited different haplotype blocks are highlighted with a gray background. VAF, variant allele frequency.

three-chromosome region (chromosome 9), where zygote Z_1 and Z_2 inherited different chromosomal segments from the father (Z_1: *O01* allele; Z_2: *A102* allele) but the same chromosome segment from the mother (*O02* allele for both Z_1 and Z_2), fully explaining the ambiguous blood type observed in Child-1 and Child-2 (Figure S1B).

Contribution of each zygote in various blood cell types

Inherited SNPs unique to Z_1 and Z_2 can be used as molecular markers to differentiate the zygotic origins of each single cell in Child-1 and Child-2. We decomposed the zygotic origin of PBMCs in the chimeric twins. To this end, PBMCs of Child-1 and Child-2 were profiled by

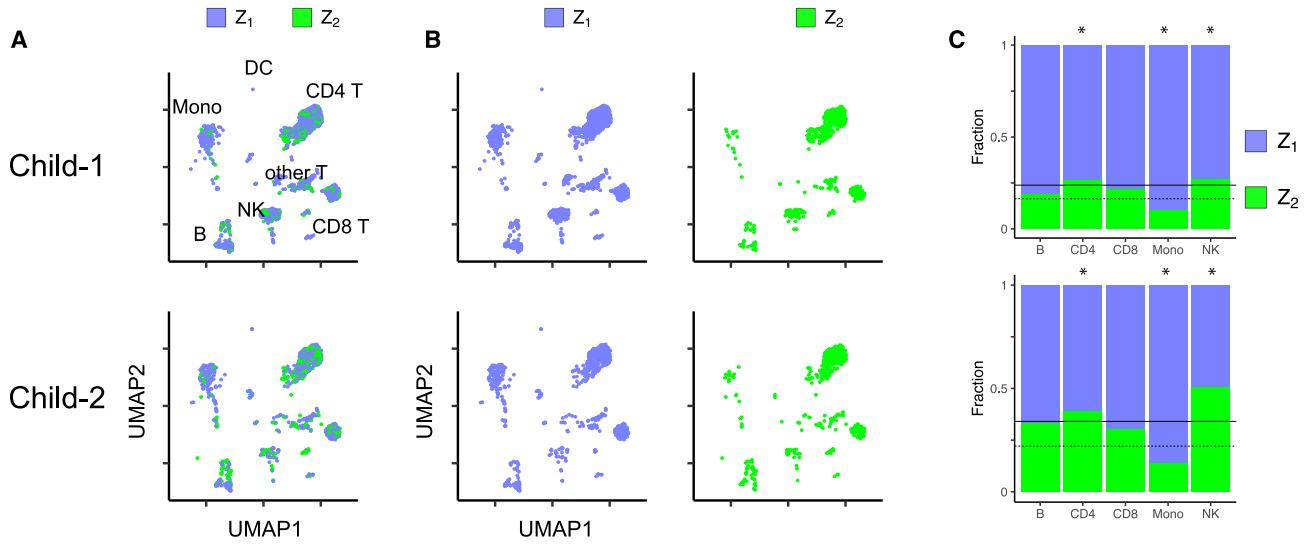

Figure 3. Single-cell-resolved chimerism of PBMCs
(A) UMAP projection of the PBMCs of Child-1 and Child-2 in Family A.
(B) Each contributing zygote is drawn separately for Child-1 and Child-2.
(C) Cell-type-specific zygotic fractions in the PBMCs of Child-1 and Child-2. The asterisks indicate statistically significant differences compared to overall cell counts (two-proportion Z-test, $*p < 0.05$). PBMC, peripheral blood mononuclear cell; UMAP, uniform manifold approximation and projection.

single-cell transcriptome sequencing and then were clustered into various cell types, including $CD4^+$ T lymphocytes, $CD8^+$ T lymphocytes, B lymphocytes, natural killer (NK) cells, and monocytes (Figure 3A). The zygotic origins of these cells were further assigned using their genotype information from the single-cell sequencing data (Figure 3B).[19] Although the dominance of the zygote (Z_1) was consistent in all cell types, the relative fraction between Z_1 and Z_2 differed for each cell type (Figure 3C; $p < 0.05$, two-proportion Z-test). Notably, the proportions of Z_2 were higher in $CD4^+$ T cells and NK cells but lower in monocytes in both the twin individuals, suggesting a genetic fitness of the two zygotes resulting in a relative differentiation potential to specific cell types.

Tracing the developmental history of chimerism in non-hematopoietic tissues

To explore potential chimerism in non-blood solid tissues, we sequenced the whole-genome of the buccal tissues of the twin individuals (Child-1 and Child-2). As buccal tissues are usually contaminated by blood cells to a substantial level,[20] we carefully isolated the buccal epithelial cells using the LCM technique[21] (Figure 4A). In contrast to the blood tissues, we found no supporting evidence of chimerism in LCM-isolated pure buccal epithelial cells (Figure 4B). Among LCM-isolated buccal epithelial cells from Child-1 ($n = 325$ cells), we did not observe any Z_2-originating epithelial cells (Figure S4). Conversely, among LCM-isolated buccal epithelial cells from Child-2 ($n = 67$ cells), no Z_1-originating epithelial cells were observed (Figure S4). Our data indicated that cellular exchanges are likely confined to hematopoietic stem cells. Our results suggest that the dominant zygotes are discordant across the tissues in Child-2 (i.e., Z_1 for blood cells and Z_2 for buccal epithelial cells and potentially other solid tissues as well).

Potential impact of ovulation induction in oogenesis

Individuals typically have approximately 70 (50–100) *de novo* mutations generated during parental gametogenesis.[23,24] On average, approximately 20% of *de novo* mutations are of maternal origin.[23,24] Usually, a sibling pair rarely shares *de novo* mutations because the chance of selecting two gametes sharing a close cellular lineage (which will share *de novo* mutations) within a large cellular pool of gametes is extremely low.[22] Indeed, sequencing data from the two external cohorts showed that most sibling pairs shared 0 or 1 *de novo* mutation (Figure 4C).

Unexpectedly, we observed an unusually high number of shared *de novo* mutations between Z_2 (one of the zygotes contributing to Child-1 and Child-2) and Z_3 (the zygote contributing to Child-3) ($n = 7$; $p = 4.37 \times 10^{-9}$, exact Poisson test $\lambda = 0.22$, 95% CI, 2.8–14.42; Figure 4C; Table S1). We speculated that all these shared *de novo* mutations were of maternal origins, as they were all located in the genomic regions of the identical maternal haplotypes in Z_2 and Z_3 (Figure 2A). Further, the two informative shared *de novo* mutations (chr3:195,327,945 A>T and chrX:10,147,343 C>T) were directly phased to the nearby maternal germline polymorphisms, confirming their maternal inheritance. In a thought experiment in which the same primary oocyte iteratively generated mature egg cells, a random pair of these egg cells would share seven *de novo* mutations (70 *de novo* mutations × 20% of maternal origins × 50% shared haplotypes). Therefore, our observations imply

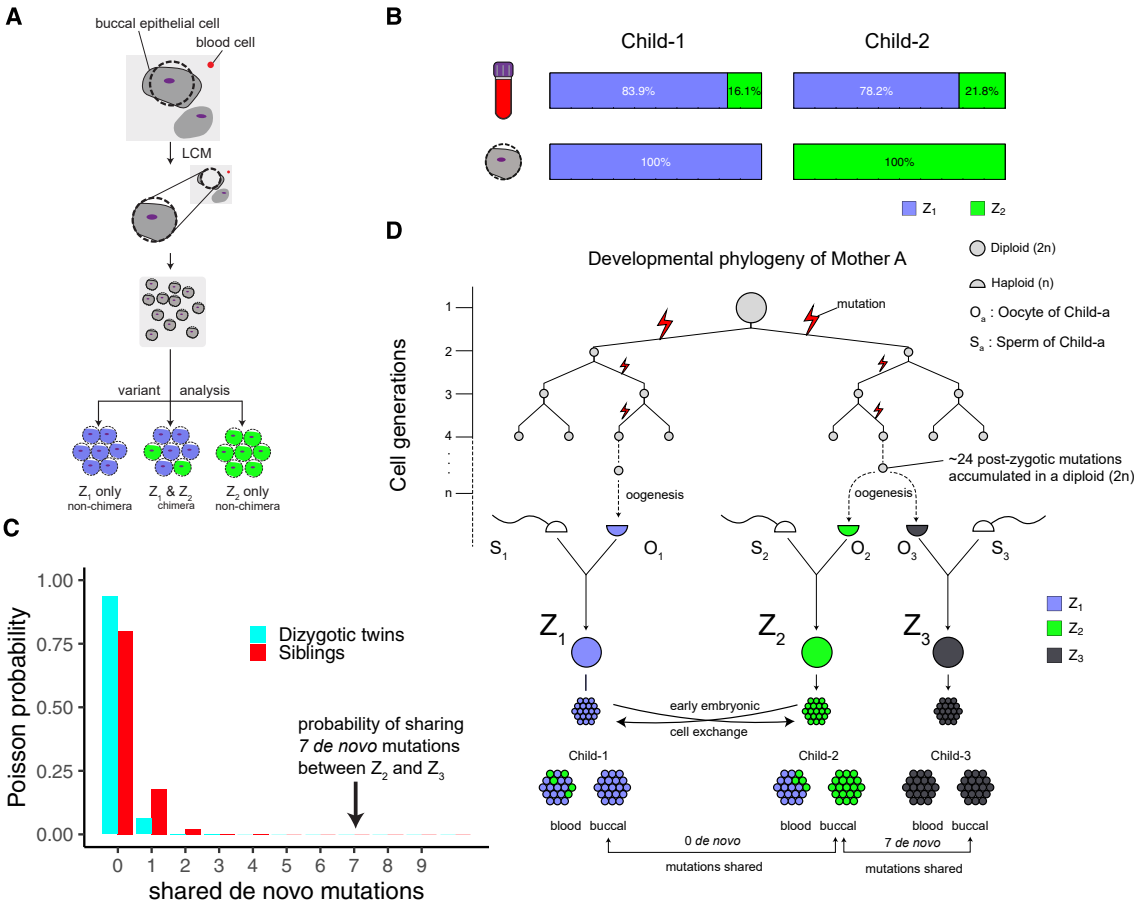

Figure 4. Tracing the chimerism of MCDZ chimera in non-hematopoietic tissues

(A) Single buccal epithelial cells are isolated using laser capture microdissection (LCM). Isolated buccal epithelial cells are pooled to generate sequence libraries. Variant analysis of each sequencing library was used to determine the contribution of the zygotes in the buccal epithelial cells.

(B) The contributions of Z_1 and Z_2 were estimated in Child-1 and Child-2's peripheral blood (top) and LCM-isolated buccal epithelial cells (bottom). The zygotic contributions for peripheral blood (top) for Child-1 and Child-2 are identical to Figure 1E, shown for comparison with those in buccal epithelial cells.

(C) The height of each bar represents the Poisson probability of shared *de novo* mutation counts based on reported data from dizygotic twins and siblings.[22] The number of shared *de novo* mutations (7) between Z_2 and Z_3 is shown with an arrow.

(D) The lineage history of the triplets is traced to gametogenesis. The half circles indicate haploid genomes of sperm or oocytes. The full circles indicate diploid genomes in the fertilized eggs and embryonic cells. MCDZ, monochorionic dizygotic.

that the oocytes contributing to Z_2 and Z_3 diverged very recently from the maternal germline, most likely from the primary oocyte (Figure 4D). Given the shared maternal haplotypes between Z_2 and Z_3 (56.9%), we speculated that the most recent common ancestor cell of maternal lines of the zygotes harbored 24 post-zygotic mutations (approximately a quarter of which are inherited as *de novo* mutations in the offspring) in the diploid genome. The mutation burden was even higher than the expected number of mutations in a diploid primary oocyte (n = 10–20), implying that two oocytes, Z_2 and Z_3, branched very recently in the mother (Figure 4D). Even though it is an observation from one family, we cannot rule out the possibility that ovulation induction treatment may have stimulated oogenesis and ovulation from a particular gamete lineage, contributing to two of the three fertilized oocytes in the triplet pregnancy in Family A.

Discussion

We used genome sequencing technologies to understand the zygotic composition and further infer chimerism's developmental history in human MCDZ twins. We used statistical quantification methods to understand the chimerism at the cellular level using WGS and scRNA-seq. Our study strongly suggests that the chimerism observed in the twins is most likely restricted to the hematopoietic system, indicating the origin from the anastomosis through the monochorionic placenta and subsequent reciprocal engraftment of hematopoietic stem cells of both zygotic origins into the bone marrow of both twin individuals. Due to the exposure of hematopoietic cells from both zygotic origins during central tolerance development in the thymus, cells of both zygotic origins are recognized as "self" and continue to

generate hematopoietic cells in both twins years after their birth.

While the twins were healthy during our study, the long-term consequences of chimerism on their health remain uncertain. The repertoire of self in these twins will be much greater than that of non-chimeric individuals, posing potential challenges for their immune systems. A long-term follow-up would be necessary to monitor the immunological consequences of chimerism. Such areas would include how the ratio of the two zygotes changes over time, perhaps in response to infection, malignancy, another inflammatory process, or even their healthy state. In addition, factors that determine one zygote to become the dominant zygote in the hematopoietic system remain unclear. These uncertainties underscore the need for continued vigilance and study in this area.

Our study also identified an unusually high number of shared *de novo* mutations from the two zygotes in one of the families. This high number of shared *de novo* mutations phased to the maternal origin suggests extremely close oocyte lineages. This is likely a process due to the ovulation induction that was used for fertility treatment rather than the chimerism itself. No study has yet investigated the role of ovulation induction on shared zygotic lineages. Comparing shared *de novo* mutations between a cohort of dizygotic twins conceived naturally and those conceived with ovulation induction would be needed to elucidate the causal relationship. As assisted reproductive technologies are being utilized more frequently to aid in human reproduction, there is a possibly increased probability of shared *de novo* mutations among the twins' zygotes.

While our study has provided a deeper understanding of chimerism in MCDZ twins, it is important to note that several limitations exist. Due to the rarity of these occurrences, our study was limited to two families. Therefore, more MCDZ twins must be studied to confirm our findings, particularly in areas such as skewed zygotic blood composition with a dominant zygote. Further research is essential to fully comprehend the complexities of chimerism in MCDZ twins. In addition, our study was limited to studying the two most accessible tissues, blood and buccal epithelial cells of mesoderm and ectoderm origin, respectively. Our investigation should be interpreted carefully, as it was limited to only two tissues (buccal epithelial and blood cells) from twin pairs. If cells of other tissues, particularly of endoderm origin, become available in the future through tissue biopsy, then confirming zygotic origins would be needed to validate our findings.

Finally, our study provides the framework for studying chimerism with high-throughput sequencing technologies. Chimera individuals are usually identified incidentally, as in our case, which was discovered via abnormal ABO blood typing. The true incidence of chimerism in the general population remains unknown. As more human genome sequences become routinely available, more chimeric individuals will be identified, and understanding their chimeric status will be critical for managing their health in the era of personalized and genomic medicine.

Data and code availability

The accession number for both families' bulk blood and buccal WGS data is European Nucleotide Archive: EGAS00001005997. The accession number for the PBMC scRNA-seq data is at the European Nucleotide Archive: EGAS50000000097. The accession number for the LCM-isolated buccal epithelial cell low-input WGS data is the European Nucleotide Archive: EGAS50000000098. All original code has been deposited at Zenodo: https://doi.org/10.5281/zenodo.10151688. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2024.100301.

Acknowledgments

This work was supported by the National Institutes of Health MSTP grant (T32 GM07200 to C.J.Y.) and the National Institutes of Health NRSA Predoctoral Fellowship (F30 HD106744 to C.J.Y.). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government, Ministry of Science and ICT (no. NRF-2022R1A5A102641311 to Y.S.J., Leading Researcher Program NRF-2020R1A3B2078973 to Y.S.J., and NRF-2020M3A9E4039670 to Y.S.J.).

Author contributions

Conceptualization, C.J.Y. and Y.S.J.; methodology, C.J.Y.; investigation, C.J.Y., J.S.L., K.Y., J.-Y.K., R.K., C.H.N., J.K., T.K., H.W., J.W.O., J.S.C., T.Y.K., D.C., and Y.S.J.; formal analysis, C.J.Y.; writing – original draft, C.J.Y.; writing – review & editing, C.J.Y. and Y.S.J.; funding acquisition, C.J.Y. and Y.S.J.; resources, J.S.C., T.Y.K., D.C., and J.S.; supervision, O.L.G., M.G., and Y.S.J.

Declaration of interests

Y.S.J. is the co-founder and chief genomics officer of Inocras, Inc. J.S.L. is the co-founder and chief innovation officer of Inocras, Inc.

Received: December 21, 2023
Accepted: May 14, 2024

References

1. Owen, R.D. (1945). Immunogenetic consequences of vascular anastomoses between bovine twins. Science *102*, 400–401. https://doi.org/10.1126/science.102.2651.400.
2. Padula, A.M. (2005). The freemartin syndrome: an update. Anim. Reprod. Sci. *87*, 93–109. https://doi.org/10.1016/j.anireprosci.2004.09.008.

3. Madan, K. (2020). Natural human chimeras: A review. Eur. J. Med. Genet. *63*, 103971. https://doi.org/10.1016/j.ejmg.2020.103971.
4. Souter, V.L., Kapur, R.P., Nyholt, D.R., Skogerboe, K., Myerson, D., Ton, C.C., Opheim, K.E., Easterling, T.R., Shields, L.E., Montgomery, G.W., and Glass, I.A. (2003). A Report of Dizygous Monochorionic Twins. N. Engl. J. Med. *349*, 154–158. https://doi.org/10.1056/nejmoa030050.
5. Chung, Y.N., Chun, S., Phan, M.-T.T., Nam, M.-H., Choi, B.M., Cho, D., and Choi, J.S. (2018). The first case of congenital blood chimerism in two of the triplets in Korea. J. Clin. Lab. Anal. *32*, e22580. https://doi.org/10.1002/jcla.22580.
6. Chen, K., Chmait, R.H., Vanderbilt, D., Wu, S., and Randolph, L. (2013). Chimerism in monochorionic dizygotic twins: case study and review. Am. J. Med. Genet. *161A*, 1817–1824. https://doi.org/10.1002/ajmg.a.35957.
7. Peters, H.E., König, T.E., Verhoeven, M.O., Schats, R., Mijatovic, V., Ket, J.C.F., and Lambalk, C.B. (2017). Unusual Twinning Resulting in Chimerism: A Systematic Review on Monochorionic Dizygotic Twins. Twin Res. Hum. Genet. *20*, 161–168. https://doi.org/10.1017/thg.2017.4.
8. Walker, S.P., Meagher, S., and White, S.M. (2007). Confined blood chimerism in monochorionic dizygous (MCDZ) twins. Prenat. Diagn. *27*, 369–372. https://doi.org/10.1002/pd.1670.
9. Fumoto, S., Hosoi, K., Ohnishi, H., Hoshina, H., Yan, K., Saji, H., and Oka, A. (2014). Chimerism of buccal membrane cells in a monochorionic dizygotic twin. Pediatrics *133*, e1097–e1100. https://doi.org/10.1542/peds.2013-1938.
10. Rodriguez-Buritica, D., Rojnueangnit, K., Messiaen, L.M., Mikhail, F.M., and Robin, N.H. (2015). Sex-discordant monochorionic twins with blood and tissue chimerism. Am. J. Med. Genet. *167A*, 872–877. https://doi.org/10.1002/ajmg.a.37022.
11. Kanda, T., Ogawa, M., and Sato, K. (2013). Confined blood chimerism in monochorionic dizygotic twins conceived spontaneously. AJP Rep. *3*, 33–36. https://doi.org/10.1055/s-0032-1331377.
12. Yoon, C.J., Kim, S.Y., Nam, C.H., Lee, J., Park, J.W., Mun, J., Park, S., Lee, S., Yi, B., Min, K.I., et al. (2022). Estimation of intrafamilial DNA contamination in family trio genome sequencing using deviation from Mendelian inheritance. Genome Res. *32*, 2134–2144. https://doi.org/10.1101/gr.276794.122.
13. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.
14. Seshan, V.E., and Olshen, A. (2016). DNAcopy: DNA copy number data analysis. R package version 1.
15. Roach, J.C., Glusman, G., Hubley, R., Montsaroff, S.Z., Holloway, A.K., Mauldin, D.E., Srivastava, D., Garg, V., Pollard, K.S., Galas, D.J., et al. (2011). Chromosomal haplotypes by genetic phasing of human families. Am. J. Hum. Genet. *89*, 382–397. https://doi.org/10.1016/j.ajhg.2011.07.023.
16. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. *22*, 568–576. https://doi.org/10.1101/gr.129684.111.
17. Lee, J.S., An, Y., Yoon, C.J., Kim, J.Y., Kim, K.H., Freeman, A.F., Yim, J.-J., Shin, E.-C., Holland, S.M., Lee, E.Y., and Ju, Y.S. (2020). Germline gain-of-function mutation of STAT1 rescued by somatic mosaicism in immune dysregulation-polyendocrinopathy-enteropathy-X-linked-like disorder. J. Allergy Clin. Immunol. *145*, 1017–1021. https://doi.org/10.1016/j.jaci.2019.11.028.
18. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
19. Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol. *36*, 89–94. https://doi.org/10.1038/nbt.4042.
20. Theda, C., Hwang, S.H., Czajko, A., Loke, Y.J., Leong, P., and Craig, J.M. (2018). Quantitation of the cellular content of saliva and buccal swab samples. Sci. Rep. *8*, 6944. https://doi.org/10.1038/s41598-018-25311-0.
21. Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. (1996). Laser capture microdissection. Science *274*, 998–1001. https://doi.org/10.1126/science.274.5289.998.
22. Goldmann, J.M., Seplyarskiy, V.B., Wong, W.S.W., Vilboux, T., Neerincx, P.B., Bodian, D.L., Solomon, B.D., Veltman, J.A., Deeken, J.F., Gilissen, C., and Niederhuber, J.E. (2018). Germline *de novo* mutation clusters arise during oocyte aging in genomic regions with high double-strand-break incidence. Nat. Genet. *50*, 487–492. https://doi.org/10.1038/s41588-018-0071-6.
23. Rahbari, R., Wuster, A., Lindsay, S.J., Hardwick, R.J., Alexandrov, L.B., Turki, S.A., Dominiczak, A., Morris, A., Porteous, D., Smith, B., et al. (2016). Timing, rates and spectra of human germline mutation. Nat. Genet. *48*, 126–133. https://doi.org/10.1038/ng.3469.
24. Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M.T., Hjorleifsson, K.E., Eggertsson, H.P., Gudjonsson, S.A., et al. (2017). Parental influence on human germline *de novo* mutations in 1,548 trios from Iceland. Nature *549*, 519–522. https://doi.org/10.1038/nature24018.