



OPEN ACCESS

Edited by:

Kazuhiko Nakabayashi,
National Center for Child Health and
Development (NCCHD), Japan

Reviewed by:

Xu Wang,
Auburn University,
United States
Miho Ishida,
University College London,
United Kingdom

***Correspondence:**

Ronaldo da Silva Francisco Junior
ronaldoj@lncc.br
Enrique Medina-Acosta
quique@uenf.br

†Present address:

Ronaldo da Silva Francisco Junior,
Laboratório de Bioinformática,
Laboratório Nacional de Computação
Científica, Petrópolis Brazil
Victor Ramos,
Laboratory of Molecular Immunology,
The Rockefeller University, New York,
NY, United States

†These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Genetics

Received: 11 July 2019

Accepted: 24 October 2019

Published: 26 November 2019

Citation:

da Silva Francisco Junior R,
dos Santos Ferreira C,
Santos e Silva JC,
Terra Machado D, Côrtes Martins Y,
Ramos V, Simões Carnivalli G,
Garcia AB and Medina-Acosta E
(2019) Pervasive Inter-Individual
Variation in Allele-Specific
Expression in Monozygotic Twins.
Front. Genet. 10:1178.
doi: 10.3389/fgene.2019.01178

Pervasive Inter-Individual Variation in Allele-Specific Expression in Monozygotic Twins

Ronaldo da Silva Francisco Junior^{1††*}, Cristina dos Santos Ferreira^{2†},
Juan Carlo Santos e Silva², Douglas Terra Machado², Yasmin Côrtes Martins¹,
Victor Ramos^{3†}, Gustavo Simões Carnivalli⁴, Ana Beatriz Garcia²
and Enrique Medina-Acosta^{2*}

¹ Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Petrópolis, Brazil, ² Laboratório de Biotecnologia, Núcleo de Diagnóstico e Investigação Molecular, Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, Brazil, ³ Department of Genetics, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil, ⁴ Department of Computational Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Despite being developed from one zygote, heterokaryotypic monozygotic (MZ) co-twins exhibit discordant karyotypes. Epigenomic studies in biological samples from heterokaryotypic MZ co-twins are of the most significant value for assessing the effects on gene- and allele-specific expression of an extranumerary chromosomal copy or structural chromosomal disparities in otherwise nearly identical germline genetic contributions. Here, we use RNA-Seq data from existing repositories to establish within-pair correlations for the breadth and magnitude of allele-specific expression (ASE) in heterokaryotypic MZ co-twins discordant for trisomy 21 and maternal 21q inheritance, as well as homokaryotypic co-twins. We show that there is a genome-wide disparity at ASE sites between the heterokaryotypic MZ co-twins. Although most of the disparity corresponds to changes in the magnitude of biallelic imbalance, ASE sites switching from either strictly monoallelic to biallelic imbalance or the reverse occur in few genes that are known or predicted to be imprinted, subject to X-chromosome inactivation or A-to-I(G) RNA edited. We also uncovered comparable ASE differences between homokaryotypic MZ twins. The extent of ASE discordance in MZ twins (2.7%) was about 10-fold lower than the expected between pairs of unrelated, non-twin males or females. The results indicate that the observed within-pair dissimilarities in breadth and magnitude of ASE sites in the heterokaryotypic MZ co-twins could not solely be attributable to the aneuploidy and the missing allelic heritability at 21q.

Keywords: allele-specific expression, allele imbalance, Down syndrome, genomic imprinting, heterokaryotypic monozygotic co-twins, mitochondrial heteroplasmy, random monoallelic expression, trisomy 21

INTRODUCTION

Monozygotic (MZ) twinning entails the partitioning of progenitor cells derived from one zygote collapsing into two sets that form two separate fetuses (co-twins) of nearly identical genotypes. MZ co-twins develop through monochorionic or dichorionic placentation as a result of when the sets of progenitor cells are split. The exact mechanisms that trigger MZ twinning are vague but genetic

(Liu et al., 2018), epigenetic, and environmental factors have been implicated (Knopman et al., 2014).

A considerable body of experimental evidence demonstrates that most MZ co-twins are not identical but discordant for (epi)genetic traits (Bennett et al., 2008; Baranzini et al., 2010; Furukawa et al., 2013; Souren et al., 2016) and congenital diseases (Chaiyasap et al., 2014; Huang et al., 2019). In stark contrast to homokaryotypic MZ co-twins, the heterokaryotypic MZ co-twins differ for constitutive chromosomal anomalies (Scott and Ferguson-Smith, 1973; Nieuwint et al., 1999). Typically, a pair of heterokaryotypic MZ co-twins exhibits discordant karyotypes for autosomal or gonosomal aneuploidies (i.e., trisomy 21, trisomy 13, XO or XXY) arising most likely post-zygotically and leading to mosaicism at various degrees (Gilbert et al., 2002; Tachon et al., 2014). Heterokaryotypic MZ co-twins may be discordant for structural chromosomal rearrangements (Leung et al., 2009; Essaoui et al., 2013), including genome-wide copy number variation (CNV) that is also commonplace in homokaryotypic MZ twins (Abdellaoui et al., 2015; Huang et al., 2019). Other likely causes for genotypic discordance in MZ monozygotic co-twins include alterations in gene expression (Buil et al., 2015), parent-of-origin effects associated to abnormal non-random (skewed) X-chromosome inactivation (XCI) (Orstavik et al., 1995), and genomic imprinting (Weksberg et al., 2002; Begemann et al., 2018). There are 43 well-documented cases of heterokaryotypic MZ co-twins in humans (**Table S1**). Most of the reported cases are spontaneous pregnancies, rather than associated with assisted reproductive technology.

Epigenomic studies in heterokaryotypic MZ co-twins are of the most significant value for assessing the effects on gene- and allele-specific expression of an extranumerary chromosomal copy or structural chromosomal disparities in otherwise nearly identical genomes.

Oligo microarray (Yan et al., 2002; Lo et al., 2003; Morley et al., 2004) and genome-wide transcriptome shotgun sequencing (RNA-Seq) studies in multiple biological samples have unveiled that many genes are subjected to the differential transcriptional expression of one allele of a pair of alleles (Dixon et al., 2015; Pirinen et al., 2015; Weissbein et al., 2016). Allele-specific expression (ASE) refers to the departure from the Mendelian 1:1 allelic expression ratio assumption. Typically, the patterns of allele expression include symmetrically (strictly) biallelic, asymmetrically biallelic (biallelic imbalance or allelic bias), and strictly monoallelic (Dixon et al., 2015; Pirinen et al., 2015; Weissbein et al., 2016).

RNA-Seq analysis allows determining the breadth and magnitude of ASE sites simultaneously. At a given experimental condition, each cell type should exhibit an array of ASE sites, an ASE signature, or transcriptome fingerprint, which is expected to be remarkably particular to the individual biological sample. The ASE signatures may be altered by environmental, health, and disease conditions (Moyerbrailean et al., 2016; Weissbein et al., 2016). In essence, the same source of cells from MZ co-twins should exhibit identical ASE signatures. However, studies based on transcriptome sequence analysis disclosed widespread discordance in ASE sites in biological samples from apparently healthy homokaryotypic MZ twins (Cheung et al., 2008;

Buil et al., 2015). Therefore, at the RNA level, the occurrence of ASE discordance constitutes a form of a cryptic, unexplained/missing heritability in individuals who share, in principle, “identical” genomes. On the other hand, genome-wide ASE discordance implies that the mechanisms for reliable transfer or flow of genetic information from DNA to RNA within humans are loose, with profound implication(s) for human health and disease (Chakravarti, 2011).

The causes of ASE discordance are associated with (epi)genetic factors, gene-gene, and gene-environment interactions (**Figure 1, Dataset S1**). For genes that are not subjected to either epigenetic regulatory mechanisms such as genomic imprinting (Baran et al., 2015), and XCI (Tukiainen et al., 2017), ASE mostly relates to the expression effects associated to quantitative trait loci (eQTLs), which can be ascribed to sequence variants of both alleles (cis effect), whereas the extent of the ASE effect relies on trans genetic variants and environmental factors interacting with the cis genetic variants (Buil et al., 2015).

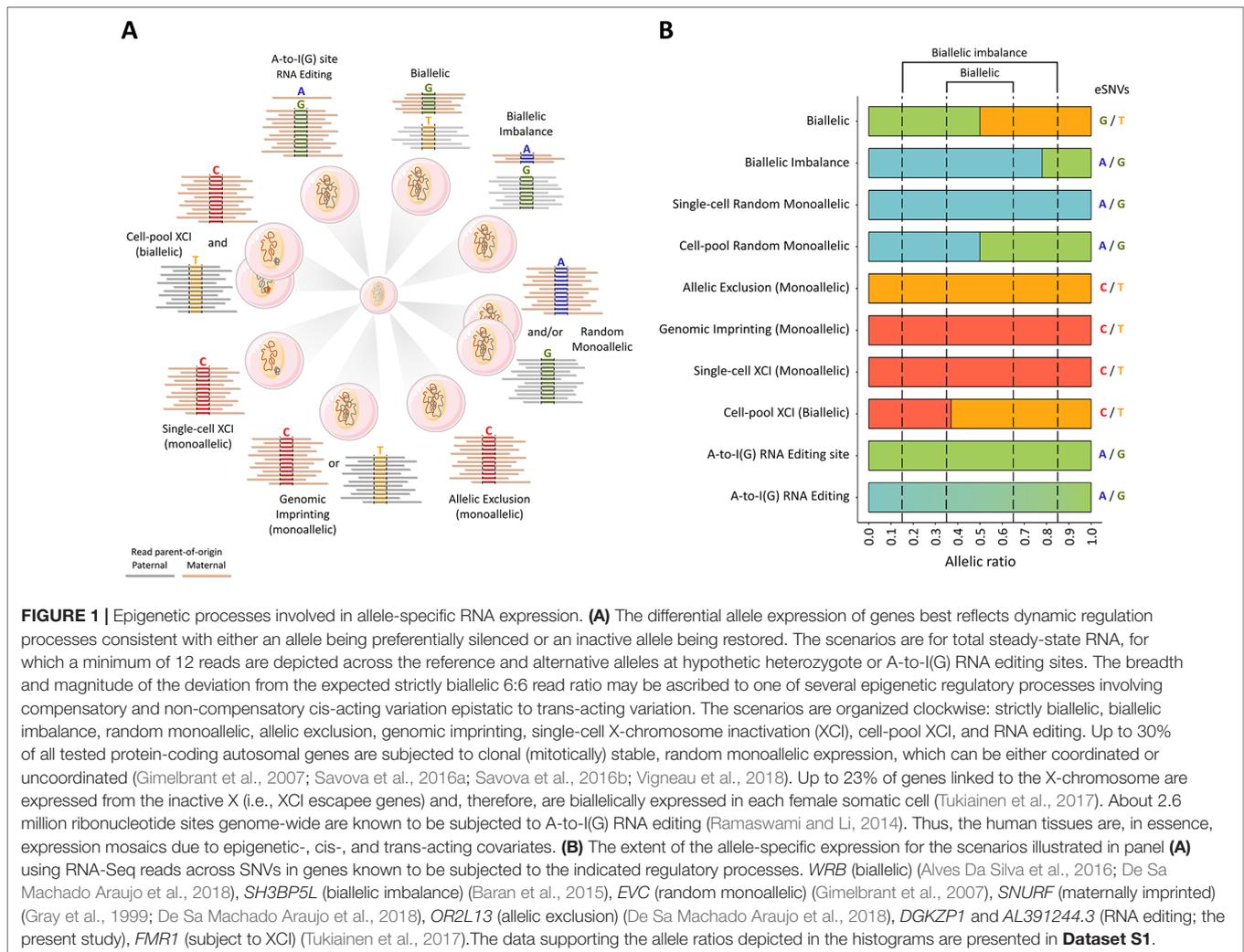
Furthermore, over 2.6 million ribonucleotide sites are known to be post-transcriptionally subjected to allele-specific editing at varying extents in several human tissues, thus contributing, at a much higher degree, to the phenotypic expression of likely mutational sites in the form of differential epitranscriptomes (Li et al., 2009; Ramaswami and Li, 2014; Zhou et al., 2018). Among the genetic factors, there are also differences in meiotic recombination and chromosomal aberrations (Weissbein et al., 2016).

Here, we carried out a comparative computation analysis of RNA-Seq data from heterokaryotypic MZ co-twins discordant for trisomy 21 and homokaryotypic MZ co-twins. We cross-referenced the ASE sites with public data repositories to exemplify the sources and consequences of within-pair disparities to annotate ASE effects in genes that are subjected to the (epi)genetic processes of genomic imprinting, XCI, and RNA editing. We identified considerable ASE disparity between either heterokaryotypic or homokaryotypic co-twins.

MATERIALS AND METHODS

Bioprojects

We used primary (unprocessed) RNA sequence filed data from the Sequence Read Archive (SRA) public experiments in 10 twin pairs, being one pair of heterokaryotypic co-twins and nine pairs of homokaryotypic co-twins. The biological samples included: primary fetal fibroblasts (GEO BioProject PRJNA239814) from the study by Letourneau and collaborators (2014), induced pluripotent stem cells (iPSC) from the study by (Hibaoui et al., 2014) (GEO BioProject PRJNA227902), and cultured B-cells (Epstein-Barr virus transformed lymphoblastoid cell lines from peripheral adult blood B-lymphocyte; GEO BioProject PRJNA170210) (**Dataset S2**). We selected the transcriptome study by Letourneau and collaborators (2014) on a pair of MZ co-twins who were karyotypically discordant for trisomy 21 (T21) of maternal origin (Dahoun et al., 2008), and therefore are heterokaryotypic twins (i.e., co-twins that differ concerning constitutive chromosomal anomalies).



Comparative transcriptomics in these heterokaryotypic twins lead to the proposal of the so-called domains of genome-wide gene expression dysregulation in Down syndrome (Letourneau et al., 2014). The case is emblematic because, in addition to the discordant maternal T21 aneuploidy, primary fetal fibroblasts from the MZ twins exhibited missing allelic heritability at 21qter as a result of recombination event(s) (Dahoun et al., 2008). A diagram of the discordant maternal 21q inheritance in the pair of co-twins heterokaryotypic for trisomy 21 is represented in **Figure S1**. To estimate the extent and magnitude of ASE discordance in unrelated, non-twin individuals, we included the RNA-Seq run experiments from BioProject PRJNA316578 (**Dataset S2**), which comprises whole blood samples from two males and two females, mean age 34-year-old, healthy controls.

Identification, Quantification, and Sorting Out Allele-Specific Expression Sites in Transcriptome Data

We implemented PipASE, an in-house computational pipeline to identify, quantify, and sort out ASE sites in the transcriptome

data (**Figure S2**). PipASE scans genome-wide for expressed single nucleotide variants (eSNVs) in high quality aligned reads. We recognize that RNA-Seq read counts and, therefore, expressed allele rates, maybe artifactually made discordant between co-twins as a result from sequencing chemistry and forward/reverse strand biases in the error rate of the high-throughput sequencing technology (Heap et al., 2010; Pickrell et al., 2012; Liu et al., 2014; Soderlund et al., 2014; Hu et al., 2015; Wood et al., 2015; Raghupathy et al., 2018; Richard Albert et al., 2018). Therefore, primary sources of technical artifacts such as systematic errors in sequencing and mapping sequence reads to a haploid reference genome were curbed by including in the PipASE the following specific algorithms that reduce or control the mapping bias: i) relaxing the number of mismatches admitted per string, yet excluding reads with spurious mismatches at the last bases of reads aligning just to one DNA strand; ii) excluding reads aligning around insertions, deletions, and simple tandem repeats; iii) excluding reads mapping to paralogous genomic regions (i.e., segmental duplications); iv) requiring ≥ 12 high-quality read depth to call a candidate informative site, and v) prioritizing the ranking of ASE sites by multiple consistent expression patterns.

Raw reads were trimmed with Trimmomatic (Bolger et al., 2014), and aligned to the hg38 reference genome using the Spliced Transcripts Alignment to a Reference (STAR, v3.5a) software (Dobin et al., 2013). We required uniquely and high-quality mapped reads (MAPQ \geq 30) by filtering them using the sequence alignment/map tools (SAMtools) (Li et al., 2009). We processed the RNA-Seq data according to the best practice guidance using the ASEReadCounter tool from the open-source Genome Analysis Toolkit (GATK, v3.8), instrumented for variant discovery in high-throughput sequencing data (McKenna et al., 2010; Depristo et al., 2011; Van Der Auwera et al., 2013). Annotated single nucleotide polymorphisms (SNP) and private SNVs were identified using HaplotypeCaller from GATK at each hypothetical heterozygous position according to HapMap (International HapMap, 2003) and database of SNP (Sherry et al., 2001). The annotation of ASE variant site positions to the hg38 reference genome was performed using the R/Bioconductor biomart package (Durinck et al., 2005; Durinck et al., 2009). SNP population data (MAF, ancestral allele) were integrated using rsnp package version 0.3.0 (Chamberlain et al., 2018). For the assessment of ASE, the read counts from the replicas were amalgamated, and Q1 values across each informative eSNV site were calculated for all biosamples on a per twin basis. For ASE sites that occurred only once in each set of biosamples, the ASE value was given by the informative run. Thus, ASE sites are supported by at least one informative run. For example, BioProject PRJNA239814, which refers to fetal fibroblasts biosamples collected from the MZ twin pair discordant for trisomy 21, comprises 12 RNA-Seq run experiments, being six per twin. The project includes four biosamples for each twin, and two of which are replicas. For that project, the distribution of informative ASE sites is 51.7, 18.2, 18.5, and 11.6% sites supported by at least 1, 2, 3, and 4 biosamples, respectively. ASE across imputed heterozygous SNP sites was calculated as the difference of RNA-Seq read counts between the two alleles, using the equation $ASE = |0.5 - \text{Ref_allele_read count} / (\text{Ref_allele_read count} + \text{Alt_allele_read count})|$. The allelic expression imbalance value per site (ranging between 0 and 0.5) is, therefore, a measure of departure from the expected Mendelian 1:1 allelic expression ratio (Babak et al., 2015; Baran et al., 2015). We annotated the ASE data by calculating the expected null reference/alternative ratios and binomial test P-values (Wang and Clark, 2014) using the *binom.test* R code function (R Core Team, 2019), and according to their gene structure sequence context (exon, intron, 5' UTR, 3' UTR, and intergenic) using the GRCh38.92 Ensembl release 96 in gtf format and the *GenomicFeatures* annotation package in R code (Lawrence et al., 2013). I-square statistical test was used to assess the degree of heterogeneity in the ASE profiles of genes supported by multiple eSNVs. The test is based on the chi-square and degree of freedom values, and it was used to measure the inconsistency of ASE profiles in each gene. We ranked genes according to the following criteria: homogeneity (I-square <30%), moderate heterogeneity (between 30 and 50%), substantial heterogeneity (between 50 and 75%), and considerable heterogeneity (> 75%). The negative I-square values were considered as 0% (Wang and Clark, 2014; Von

Hippel, 2015). A flowchart for the PipASE used for scanning and sorting out genome-wide, allele-specific differences between MZ co-twins is shown in **Figure S2**.

Cross-Referencing With Public Data Repositories

For every ASE site observed in each RNA-Seq sample, we extracted functional information by computational cross-referencing with public databases regarding pathogenic expression-altering or loss-of-function risk variant alleles (Adzhubei et al., 2010; Landrum et al., 2016; Vaser et al., 2016), genomic imprinted genes (Jirtle and Murphy, 2012; Wei et al., 2014; Baran et al., 2015; Pirinen et al., 2015), A-to-I(G) RNA editing sites (Ramaswami and Li, 2014), germline ASE discordant sites in MZ twins (Cheung et al., 2008), and XCI escapee and non-escapee genes (Carrel and Willard, 2005; Cotton et al., 2013; Balaton et al., 2015; Cotton et al., 2015; Tukiainen et al., 2017; Garieri et al., 2018; Shvetsova et al., 2019; Wainer Katsir and Linial, 2019). Allelic expression profiles were validated computationally by data integration with the ASE profiles observed in multiple human tissues from the Genotype-Tissue Expression (GTEx) project (The GTEx Project, 2015), using the *Data Integrator* tool available at the UCSC Genome Browser, that contains track hubs for the second source GTEx data (release V6, October 2015), mainly as previously reported (De Sa Machado Araujo et al., 2018).

Canonical A-to-I(G) Ribonucleic Acid Editing

ASE sites were queried in the RADAR database, which comprises a list of about 2.6 million rigorously annotated database of A-to-I(G) RNA editing sites. For cross-referencing of the ASE sites, we merged RADAR data version 1 (available online from the RADAR browser) and version 2, which is based on the GTEx RNA-Seq dataset from 30 tissues (hg19; version 6p), and reports RNA editing levels for sites with \geq 20 reads (Tan et al., 2017), kindly provided as a flat database by Dr. Jin Billy Li at Stanford University (Ramaswami and Li, 2014). The hg19 coordinates were lifted over to hg38 using “hg19ToHg38.over.chain” file and R scripts based on *AnnotationHub* (Morgan, 2017) and *rtracklayer* libraries (Lawrence et al., 2009). We limited the analysis to base positions corresponding to canonical A-to-I(G) variants, excluding all SNVs that map within segmental duplications or simple repeats in the hg38 reference genome, using the ShortMatch tool with query strings of 50 bases in length containing the variant at position 26th. The filter-selection step above followed published quality guidelines (Lin et al., 2012b; Ramaswami et al., 2012; Piskol et al., 2013). For every ASE site matching a RADAR reference editing site location, we calculated the A-to-I(G) RNA editing levels as the ratio of G-containing reads divided by the sum of A- and G-containing reads in RNA-Seq experiments of each pair of co-twins. The strength of the co-association between the levels of RNA editing at ASE sites within twin-pairs was measured using linear models in R.

RESULTS

Transcriptome-Wide, Allele-Specific Differences Observed in Monozygotic Co-Twins Discordant for Both Trisomy 21 and Recombination

Recombination and sequence variation are major evolutionary sources of diversity in the human genome. We, therefore, wished first to evaluate how these two forces impacted on ASE in “identical” co-twins. Between the MZ co-twins discordant for T21, we identified 1,227 (3.8%) ASE sites whose allelic patterns were discordant (i.e., monoallelic *versus* biallelic) in fibroblasts and 3,295 (6%) such sites in iPSC (Figure 2A, Dataset S3). We estimated the magnitude of expression change between conditions for the variants called (Figure 2B). The bulk of the ASE sites exhibited a LogASE value close to zero, which means that the majority of the ASE sites were not altered in trisomy 21 condition. Importantly, 19 eSNVs were significantly altered in fibroblasts of the trisomy 21 (T1DS) affected twin, being 16 sites with LogASE ≥ 0.8 and three sites with LogASE ≤ -0.8. Noteworthy, 11 implicated genes mapped to the 21q region discordant for maternal inheritance due to a recombination event. Among those genes, *CASP6*, *FAM86GP*, and *PDXDC1/PKD1P6* were expressed monoallelically, whereas the *IL17RA* gene was expressed biallelically in fibroblasts from the T1DS twin. In iPSC, we observed 260 eSNVs with LogASE ≥ 0.8 and 214 ≤ -0.8

annotated in 274 genes (Figure 2B). Of the 19 ASE sites with ASE values ≤ -0.8 or ≥ 0.8 in fibroblasts, 14 were also called in iPSC. However, only 10 sites were altered in both cell types with values of ASE ≥ 0.8 (Dataset S3), and are located within the 21q region spanning the recombination event. The overall distribution of genes by the numbers of ASE sites observed in fibroblasts and iPSC is shown in Figure 2C.

The discrepancies in ASE between the MZ co-twins discordant for T21 observed in both fibroblasts (Figure 3A), and iPSC (Figure S3A) were widespread in the genome (average of 20 ASE sites per Mb). We validated the heterokaryotypic status of the MZ twins discordant for T21 by comparing the within-pair global allele-ratios and plotted them as expression karyotypes (e-karyotypes) (Figure 3B and Figure S3B).

Allele-Specific Expression Disparity Observed in Homokaryotypic Twin-Pairs

To begin to sort out the likely causes of the widespread ASE discordance found in co-twins, we examined the breadth and magnitude of discordant ASE sites in nine pairs of co-twins not discordant for aneuploidy and recombination. Surprisingly, the breadth and magnitude of ASE concordance and discordance in the control twin pairs were comparable to those observed in the heterokaryotypic twin pair, with an average 1,074 discordant sites (2.7%) per twin pair (Figure S4). The discordant ASE sites were also distributed genome-wide (Figure S5). Despite their

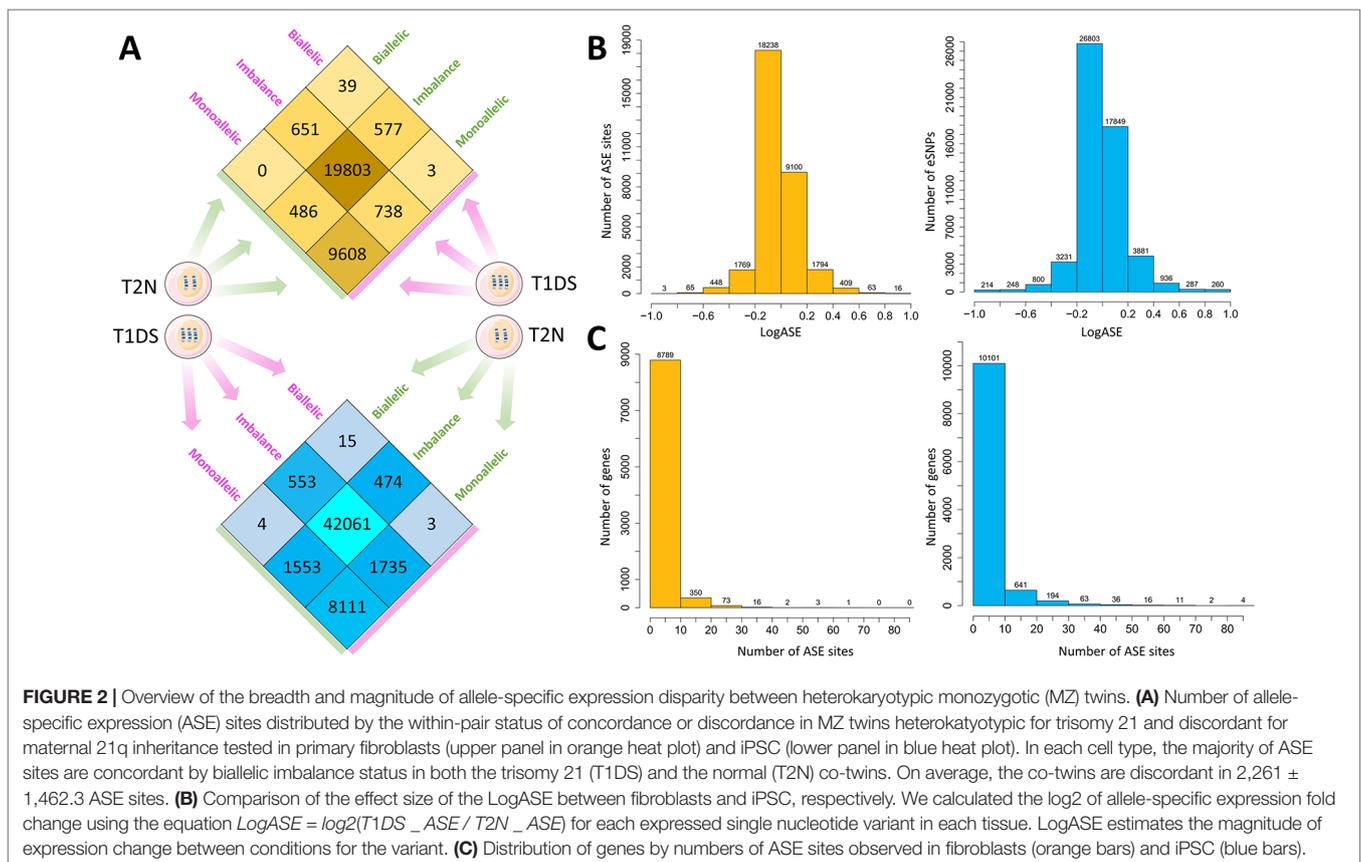
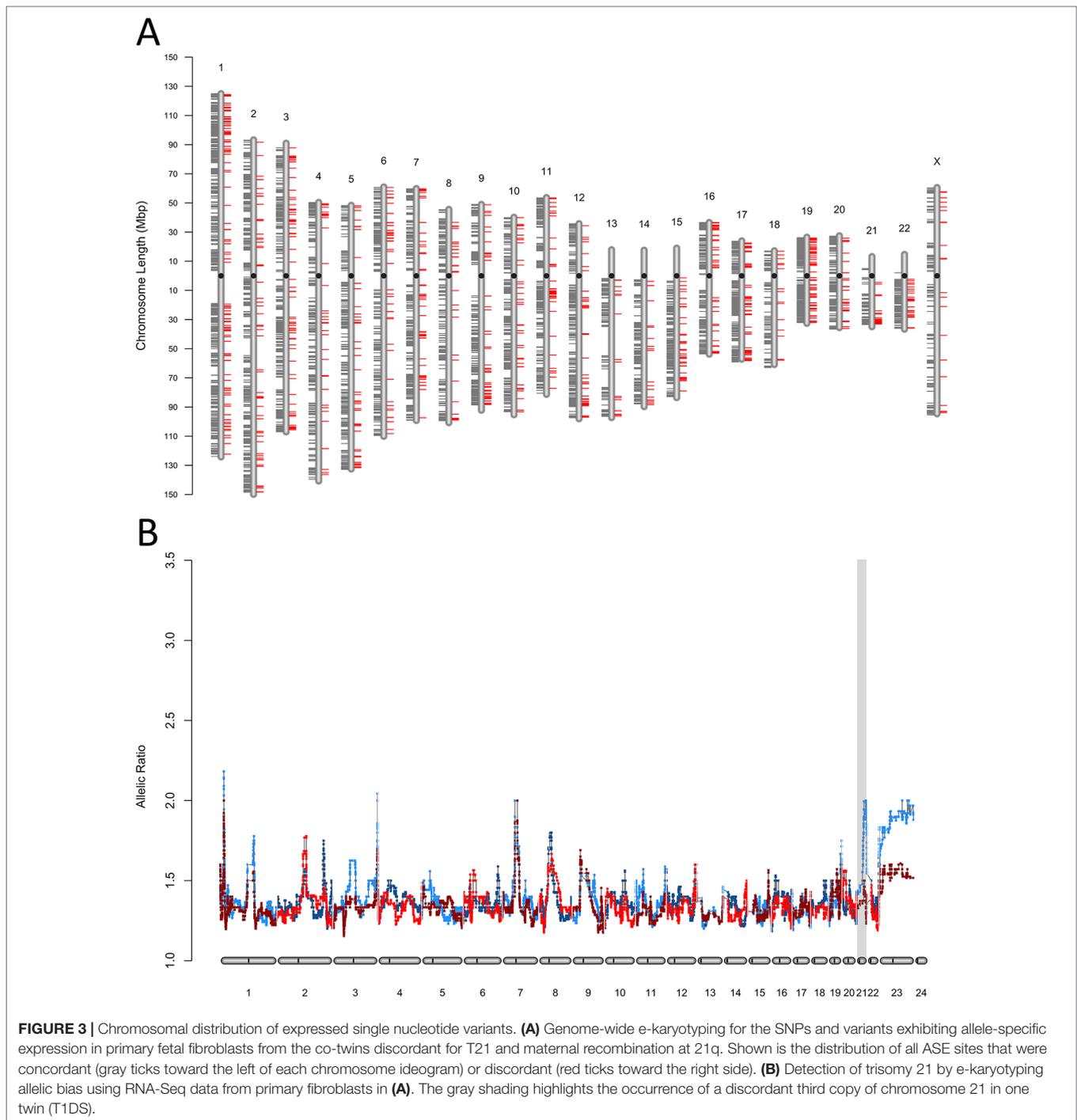


FIGURE 2 | Overview of the breadth and magnitude of allele-specific expression disparity between heterokaryotypic monozygotic (MZ) twins. **(A)** Number of allele-specific expression (ASE) sites distributed by the within-pair status of concordance or discordance in MZ twins heterokaryotypic for trisomy 21 and discordant for maternal 21q inheritance tested in primary fibroblasts (upper panel in orange heat plot) and iPSC (lower panel in blue heat plot). In each cell type, the majority of ASE sites are concordant by biallelic imbalance status in both the trisomy 21 (T1DS) and the normal (T2N) co-twins. On average, the co-twins are discordant in $2,261 \pm 1,462.3$ ASE sites. **(B)** Comparison of the effect size of the LogASE between fibroblasts and iPSC, respectively. We calculated the log2 of allele-specific expression fold change using the equation $LogASE = \log_2(T1DS_ASE / T2N_ASE)$ for each expressed single nucleotide variant in each tissue. LogASE estimates the magnitude of expression change between conditions for the variant. **(C)** Distribution of genes by numbers of ASE sites observed in fibroblasts (orange bars) and iPSC (blue bars).



diverse parental origins, there were, on average, 19,488 ASE sites common within the nine pairs of homokaryotypic MZ twin pairs; 90 (0.46%) sites were discordant in the entire set of twin pairs. Nevertheless, there were, on average, 571 ASE sites discordant in a given twin pair, but concordant in another. The recurrent sites in all nine pairs best reflect identity by state. We note, however, that monozygotic twin developed with a shared circulation, and therefore, the ASE profiles assessed in cultured transformed B-cells isolated at an early age will tend to be similar.

Unfortunately, we could not trace the chorion type of the nine homokaryotypic twin-pairs, which were sampled at the age ranging 19 to 65 (**Dataset S2**).

For the entire set of MZ twin pairs, the average distribution of eSNVs per gene was the following: 34.3% ($n = 3,162$) of genes were called by one eSNV; 57.8% ($n = 5,333$) were supported by 2 to 10 eSNVs; 7.9% ($n = 729$) were called by 11 to 200 eSNVs; 0.02% ($n = 2.4$) were called by 201 to 500 eSNVs, and 0.01% ($n = 1.1$) exhibited >500 eSNVs (**Dataset S4A** and **S4B**). We carried

a statistical test for heterogeneity to query for intervention effects (variation in effect estimates beyond chance) across a given genomic region. For the entire set of biosamples, we found, on average, that 43.6% ($n = 2,619$) of genes supported by multiple eSNVs exhibited considerable homogeneity across the eSNV profiles; 3.8% ($n = 225$) had moderate heterogeneity; 6.1% ($n = 366$) had substantial heterogeneity; and 46.5% ($n = 2,795$) had considerable heterogeneity (**Dataset S4C and S4D**). We note that genes exhibiting considerable heterogeneity are large (on average 130 Kbp, i.e., *CD226*) and are supported on average by 7.2 (range 2 to 318) eSNVs. Conversely, the most homogeneous profiles are in genes with an average size of 12 Kbp (i.e., *JRK*), which are supported on average by 3.7 eSNVs (range 2 to 38 sites). Moreover, comparing genes supported by the same number of eSNV (i.e., 30 sites), we note that the eSNVs are distributed differently, toward the 3' UTR in genes ranked as homogeneous (i.e., *LGALS8* and *PLEC*) and spread along the gene body in those ranked as heterogeneous (i.e., *CD226* and *GLEC17A*).

We also validated the homokaryotypic status of the nine MZ control twin pair by comparing the within-pair global allel-ratios and plotted them as e-karyotypes (**Figure S6**). Jointly, the e-karyotyping analyses demonstrate that there is pervasive missing allelic heritability between the transcriptome of MZ co-twins and that the bulk of the ASE site within-pair disparities in the heterokaryotypic co-twins cannot be solely attributed to the differential occurrence of aneuploidy and the missing allelic heritability at 21q.

Allele-Specific Expression Disparity Observed in Unrelated, Non-Twin Males and Females

Unrelated, non-twin males and females exhibited comparable extents of ASE discordance genome-wide: 24.8% (6,546/26,371 eSNV sites) in males (**Dataset S5A**) and 25.57% (5,992/23,431 eSNV sites) in females (**Dataset S5B**). Therefore, the extent of ASE discordance in unrelated, non-twin males and females is about 10-fold higher than the observed between pairs of MZ twin-pairs (2.7%). In the unrelated male and female set, 47.4 and 45.4% of genes supported by ≥ 2 eSNVs, respectively, exhibited considerably heterogeneous ASE profiles, whereas 43.4 and 45.5% of genes were ranked as considerably homogeneous (**Dataset 5C**). Similar to the finding in MZ twins, genes exhibiting considerable heterogeneity are large (on average 83 Kbp, i.e., *GAK* in males and 221 Kbp in females, i.e., *SAMD3*) and are supported on average by 8.2 (range 2 to 104) eSNVs in males and 7.7 (range 2 to 106) eSNVs in females. Conversely, the most homogeneous profiles are in genes with an average size of 18 Kbp (i.e., *UBE2I* in males and *EEF1D* in females), which are supported on average by 3.7 (range 2 to 39) eSNVs in males and 3.5 (2 to 36) eSNVs in females. Again, comparing genes supported by the same number of eSNVs (i.e., 25 sites), we note that the eSNVs are distributed differently, toward the 3' UTR in genes ranked as homogeneous (i.e., *HCP5* and *PRRC2B* in males and *AC004151.1* and *NOTCH1* in females) or spread along the gene body in those ranked as heterogeneous (i.e., *FCGBP* and *GAK* in males and *SAMD3* and *SYNE3* in females).

Assessment of the Underlying Causes of the Observed Pervasive Missing Allelic Inheritability

The underlying causes of the observed pervasive missing allelic inheritability can include i) genome-wide DNA sequence variations within pairs of MZ co-twins, as supported by recent findings in MZ twins discordant for autism spectrum disorder (ASD) using whole-genome sequencing (Huang et al., 2019) and ii) differential expression of alleles. Given that none of the ten MZ twin pairs referred here has genomic sequences available in public repositories, we first cross-referenced the observed ASE sites with data about the distribution of eSNVs reported between the MZ co-twins discordant for ASD (Huang et al., 2019). On average, the MZ co-twins discordant for ASD exhibited 54 eSNVs disparities annotated in exons, 3,912 in introns, 13 in 5' UTR, and 74 in 3' UTR for 2,786 genes (**Table S2**). Remarkably, between either the MZ co-twins heterokaryotypic for T21 or the homokaryotypic MZ co-twins, we identified, on average, 10,111 ASE discordant sites in annotated exons, 8,037 in introns, 2,066 in 5' UTR, and 18,032 in 3' UTR for 8,495 genes. Thus, a mean 120-fold increase in discordant ASE sites per annotation category. This fold difference cannot be attributed solely to the average distribution rate of discordant eSNVs of 1.1×10^{-4} per exonic site reported across human genes between the genomes of MZ co-twins (Huang et al., 2019). Furthermore, there is a 20-fold deficit in ASE sites annotated in intergenic regions as compared with the number of eSNV sites discordant by whole genome sequencing, which supports the view that the biased distribution of ASE sites discordances within genes may be biologically relevant. We also validated some of the ASE discordant sites by cross-referencing with the sets ASE sites in MZ twins from the study by Cheung et al. (2008) (**Dataset S3**).

Next, we cross-referenced the ASE sites with data about genes known or predicted to be expressed from one allele at a time through genomic imprinting, XCI, and A-to-I(G) RNA editing. Overall, we identified discordant ASE sites in either 205 known or candidate imprinted genes (**Dataset S3**), 12 X-linked genes (**Table S3**), and 3,955 sites likely subjected to A-to-I(G) RNA editing (**Dataset S3**).

Allele-Specific Expression Switching in Imprinted Genes

We note that, on average, 4,574 ASE sites were monoallelic concordant within co-twins. Annotation of those sites revealed that 8,867 genes exhibited multiple monoallelically eSNVs with no biallelically expressed sites (**Datasets S3, S6–S14**). Among those genes, we annotated five known imprinted genes (*DR1*, *BRD2*, *VAR2*, *MEG3*, and *H19*), each one ranked with ≥ 8 eSNVs. Cross-reference of those genes with secondary data from the GTEx project validated their monoallelic expression in multiple tissues (**Dataset S15**), and therefore their imprinted status (Jirtle and Murphy, 2012; Wei et al., 2014; Baran et al., 2015; Pirinen et al., 2015). Unfortunately, the GTEx project does not include samples of embryonic fibroblasts, iPSC, or B-cells. In contrast, most other genes ranked with ≥ 8 monoallelic eSNVs were expressed biallelically in multiple tissues in the GTEx

database (**Dataset S15**). We speculate that the monoallelic concordance at multiple sites observed in the co-twins reflects extended homozygosis, rather than parent-of-origin effects. We note five gene exceptions. First, the *AC091729.3* gene, which exhibited an average of 12 monoallelically eSNVs in two of the ten MZ co-twin pairs, is also expressed monoallelically in an isoform-specific fashion with four SNPs in 35 tissues in the GTEx samples (**Dataset S15**). Second, the *SPINK5* gene with 11 monoallelically eSNVs, expressed monoallelically exclusively in the bladder in the GTEx data. While the ASE profile across the *AC091729.3* gene was homogeneous, the *SPINK5* gene ranked moderately heterogeneous. We view these two genes as potential leads and suggest that the *AC091729.3* gene is subjected to isoform-specific genomic imprinting, whereas the *SPINK5* gene is imprinted in a tissue-specific manner. Third, the known imprinted genes *SNURF*, *SNHG14*, and *ZNF264*, which are expressed monoallelically in multiple tissues in the GTEx samples, exhibited ≥ 10 biallelically imbalance eSNVs in iPSC (**Dataset S15**). Interestingly, various biallelically imbalance eSNVs in these three genes are listed in the RADAR database and are likely subjected to A-to-I(G) RNA-editing: 5 out of 16 eSNVs (*SNURF*), 19/35 (*SNHG14*), 4/14 (*ZNF264*) (**Dataset S3**). We, therefore, suggest that the epitranscriptome modification of these gene products by RNA editing alters their expected imprinting phenotype (at least *in vitro*) in iPSCs.

Estimated Impact of Canonical A-to-I(G) Ribonucleic Acid-Editing on Allele-Specific Expression Disparity

The number of ASE sites that positionally correspond to canonical A-to-I(G) sites was, on average, $2,012 \pm 786$ per twin pair (**Datasets S3, S6–S14**), and all the sites cover 419 ± 116 genes. The vast majority of sites exhibited a concordant biallelic imbalance profile (pink and light blue dots in **Figure 4** and **Figure S7**). Thus, within the co-twins, there was an overall concordance in the biallelic imbalance state. The number of sites that exhibited discordant allelic profiles, being biallelic in one twin and monoallelic in the other, was however minimal, albeit more abundant in the homokaryotypic than in the heterokaryotypic co-twins (green dots in **Figure 4** and **Figure S7**). This observation indicated that in the cells analyzed, few eSNVs were 100% edited (i.e., expressed strictly monoallelically) in the complete set of 10 pairs of twins. Between the heterokaryotypic co-twins, there were only seven such discordant sites, being monoallelically expressed in T1DS and biallelically imbalance in the normal co-twin (**Figure 4**). The seven discordant sites occur in seven protein-coding genes, including *CD46* (an immune type I receptor) and *ING5* (a tumor suppressor).

In the heterokaryotypic twins, 117 expressed genes exhibited high proportions of ASE sites (≥ 4 sites per gene) coincident with RNA editing sites. Notably, for the *CYP20A1* and *ZNF621* genes, 74 and 77% of all ASE sites are canonical A-to-I(G) sites validated in the RADAR database. The extent of allele imbalance ranged from 6 to 98%. However, about 5% of all sites exhibited discordant RNA editing levels higher than 25% between co-twins,

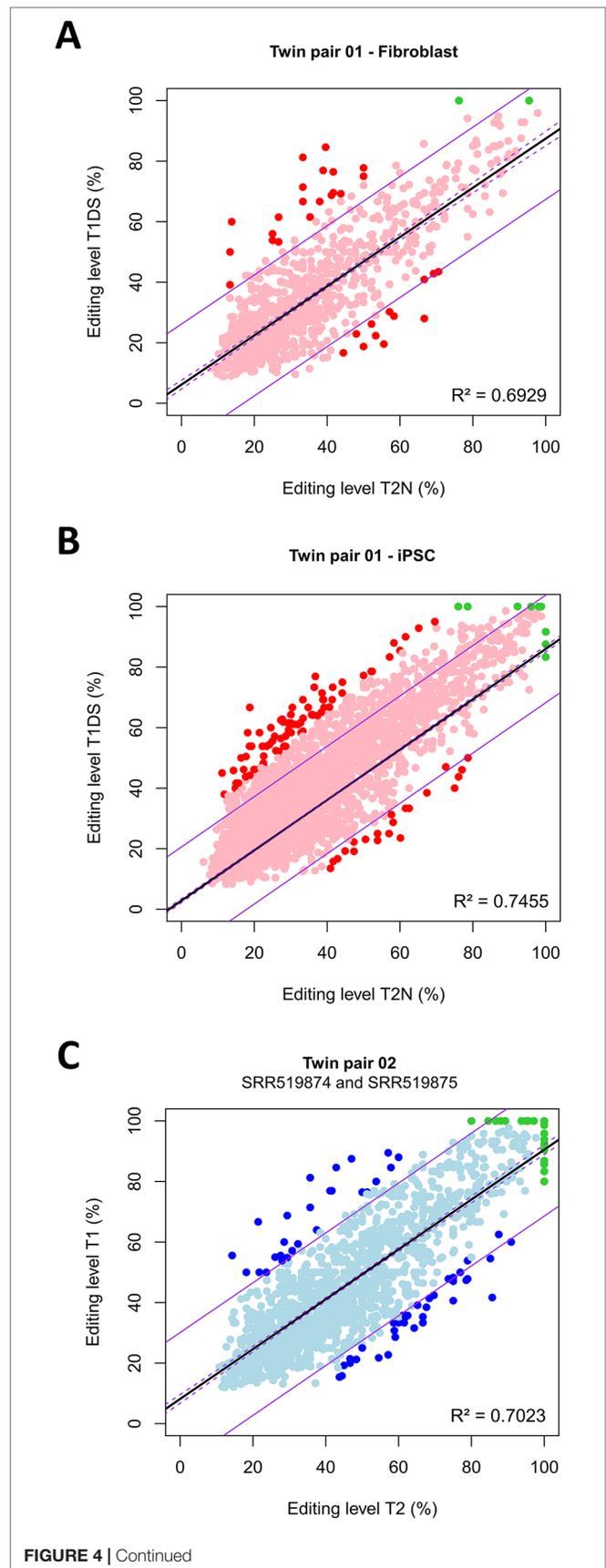


FIGURE 4 | Continued

FIGURE 4 | Within twin-pair disparities in allele expression proportions at expressed single nucleotide variants (eSNVs) that are coincident with canonical A-to-I(G) RNA editing sites. Shown is the distribution of eSNVs that positionally match canonical RNA editing sites between heterokaryotypic co-twins, assayed either in fetal fibroblasts (A), fetal fibroblast-derived iPSC (B), or between homokaryotypic co-twins tested in culture-B-cells (C). Each dot corresponds to an eSNV. The vast majority of sites exhibited a concordant biallelic imbalance profile (pink and light blue dots). Red dots represent eSNVs that were discordant between co-twins in that they showed allelic proportions differences higher than 25%, regardless of the discordance or concordance in the karyotype. Green dots represent eSNVs that exhibited discordant allelic profiles, being biallelic in one twin and monoallelic in the other. The linear models (solid black lines), the confidence interval of the models (broken purple lines), and the predictions (solid purple lines) were constructed using R. Model equations: (A) $Y = 6.15718 + 0.81302X$; (B) $Y = 2.832681 + 0.831631X$; (C) $Y = 8.15439 + 0.82362X$. For all pairs, $P < 2.2e-16$.

regardless of whether hetero- or homokaryotypic conditions (red and blue dots in **Figure 4**, **Figure S5** and **Dataset S3**). For example, the gene *CYP20A1* presented the highest percentage of allele imbalance between the heterokaryotypic co-twins in fibroblasts (T1DS = 81.25%/T2N = 33.33%) whereas in iPSC the gene *GCFC2/MRPL19* exhibited the highest discrepancy (T1DS = 66.66%/T2N = 18.75%).

Since A-to-I(G) RNA-editing of mRNAs can create stop codons (protein-truncating effect variants) or result in non-synonymous mutations, it was important to annotate the sites with discordant allelic proportions. In the heterokaryotypic co-twins, none of the annotated sites created stop codons, but nine sites are predicted to cause non-synonymous mutations (**Dataset S16**). The cyclin-dependent kinase 13 gene *CDK13*, presented two non-synonymous editing sites that change lysine (Lys; chr7_39950928) and glutamine (Gln; chr7_39950949) to arginine. Within the 5,372 eSNVs annotated as canonical RNA editing sites in the transcriptomes of the homokaryotypic twin pairs, none creates stop codons, and 21 sites correspond to non-synonymous mutations (**Dataset S16**).

MZ Co-Twins Are Discordant in the Allele-Specific Expression of X-Chromosome Inactivation Non-Escapee Genes

We scanned for discordant ASE sites in X-linked genes within the seven female twin pairs, and integrated data for the intersected sites about the XCI classification status from public repositories (Carrel and Willard, 2005; Cotton et al., 2013; Balaton et al., 2015; Cotton et al., 2015; Tukiainen et al., 2017; Garieri et al., 2018; Shvetsova et al., 2019; Wainer Katsir and Linial, 2019). The analysis was restricted to non-escapee genes because, in pooled cells, those genes must exhibit biallelic expression profiles. For this specific analysis, we only accepted gene products that displayed at least two discordant eSNVs (i.e., monoallelic in one twin versus biallelic in the sister twin). For the heterokaryotypic discordant co-twins, there was ASE disparity in the *UBLAA* gene products in fibroblasts, whereas, in iPSC, there was ASE disparity in the *FANCB* and *FTX* gene products (**Table S3**). Three control twin pairs expressed genes with at least two eSNVs: *TAB3*, *WDR44*, and *XIAP* genes in twin pair 05; *IDS*, *MAP7D3*, *RLIM*,

RPL10, *SLC9A7*, *TBC1D25*, *TLR7*, *XIAP*, and *ZNF275* genes in twin pair 08; and the *ZNF275* gene in twin pair 09 (**Table S3**). None of the ASE disparities above are A-to-I(G) RNA editing sites in the RADAR database.

The Overall Impact of Allele-Specific Expression of Pathogenic Variants

We annotated 32 eSNVs associated with 131 human pathologies in the transcriptomes of the heterokaryotypic twins (**Dataset S17A**). Most pathogenic eSNVs are linked to autosomal recessive phenotypes and were coexpressed with the wild type allele, likely outbalancing the predicted deleterious effects. Four pathogenic alleles (rs1799990*G > A, rs1800562*G > A, rs200855215*A > G, and rs4784677*A > G) were expressed monoallelically, and are associated with Jakob-Creutzfeldt disease (OMIM #123400), hemochromatosis (OMIM #235200), Leber optic atrophy (OMIM #535000), and Bardet-Biedl syndrome 2 (OMIM #615981), respectively. One pathogenic allele (rs11583680*C > A), associated with autosomal dominant familial hypercholesterolemia (OMIM #603776), was also coexpressed with the wild type allele. In the homokaryotypic twin sets, 23 eSNVs, predicted to be pathogenic in the ClinVar database, predominantly coexpressed with the wild type alleles (**Dataset S17B**). For example, rs1799958*G > A, associated with deficiency of butyryl-coenzyme A dehydrogenase (OMIM #201470), was coexpressed with the wild-type allele in MZT3.

Evidence for Expressed Mitochondrial Microheteroplasmy

We also identified an ASE form of mitochondrial microheteroplasmy (Souren et al., 2016), albeit at lower limits, in all 10 MZ twin pairs, demonstrated by the presence of 237 eSNVs (median number of 25 eSNVs per dataset, **Table S4**). The observed limited number of mitochondrial eSNVs does not relate exclusively with the early embryonic age at sampling because the age of the control twin pairs ranged from 19- to 65-year-old (median age 26 years) (**Dataset S2**). Thus, for the set of donors investigated, we did not observe the accumulation of mitochondrial eSNVs with age (Smigrodzki and Khan, 2005).

Lastly, we queried ClinVar, PolyPhen, and SIFT public databases for evidence about the pathogenicity prediction for the mitochondrial eSNVs to assess the most functionally crucial mitochondrial point mutations. Fifteen eSNVs are predicted to be likely pathogenic in at least one database (**Table S4**). For example, rs28358569*A > G, monoallelically expressed in MZT9, is related to mitochondrial non-syndromic sensorineural hearing loss (OMIM #500008) and aminoglycoside-induced deafness (OMIM #580000); rs193302980*C > T and rs2853508*A > G are related to familial breast cancer (OMIM #114480).

Gene Ontology Analysis of Discordant Allele-Specific Expression Sites

In fibroblasts, the *CASP6* and *PDXDC1* genes, represented by ASE sites exhibiting biallelic to monoallelic switch ($\text{LogASE} \geq 0.8$) within the heterokaryotypic co-twins were related with

nitrogen compound and organic substance metabolic processes (**Dataset S18A**). On the other hand, *IL17RA*, the only gene with $\text{LogASE} \leq -0.8$ and mapping outside the well-characterized 21q recombination region, is enriched in immunological processes such as leukocyte migration, signal transduction, cytokine production, and cell activation (**Dataset S18B**). In iPSCs, most of the genes (72 of 100 genes) with discordant ASE sites (either bi-to-mono or mono-to-biallelic switches) are related to the regulation of biological process (**Datasets S9C and Datasets S9D**).

DISCUSSION

We aimed to compile variant sites with expression profiles that are dissimilar between MZ co-twins who are discordant or not for a specific condition. Our scanning strategy permitted the identification, quantification, and classification of differential allelic expression by way of ASE discordant sites (i.e., eSNV) occurring genome-wide between co-twins who are either discordant or not for T21. Remarkably, the breadth and magnitude of ASE discordant sites were high and comparable between either heterokaryotypic or homokaryotypic co-twins. On average, we identified about 1,342 ASE discordant sites in the 10 pairs of MZ co-twins.

The extent of the ASE sites in T21 discordant co-twins was comparable between the non-discordant co-twins, assayed in three cell types (fibroblasts, iPSC, and B-cells). Overall, the analyses indicate that ASE discordance between MZ co-twins stems from aneuploidy, recombination, genomic imprinting, and RNA editing. We interpret the widespread occurrence of ASE discordance between MZ co-twins as being the result of sister chromatid-specific alterations in transcription. The discordant ASE sites observed between co-twins best reflect a combined effect of genetic and epigenetic processes on differential allele expression.

We note that e-karyotyping unveils dynamic arrays of ASE sites that can be considered as signatures that exhibit remarkable singularity to the individual biological sample. For example, for the heterokaryotypic co-twins, the sets of eSNVs observed in fibroblasts or iPSC do not overlap entirely. Overall, 38.67% ($n = 24,103$) of ASE sites were called in both fibroblasts and iPSC samples, 804 (3.3%) of which exhibited discordant allelic-expression profiles in fibroblasts, but concordant in iPSC. Similarly, 1,318 sites (5.4%) were concordant in fibroblasts but discordant in iPSC. Moreover, 187 sites (0.7%) were discordant in both sample types. The relative lack of overlap among the experiments is likely explained by the differential expression of genes in these cell types. Therefore, e-karyotyping signatures might have forensic value and resolution power to discriminate clinically non-discordant co-twins. The e-karyotyping signatures may be specific to the level of each experimental condition for the same source of a biological sample. In principle, no sharing of e-karyotyping signatures is expected to occur within co-twins.

The observation of allelic bias is becoming commonplace in high-throughput transcriptome analyses (Deveale et al., 2012; Marinov et al., 2014; Metsalu et al., 2014; Wood et al., 2015; Weissbein et al., 2016). It is acceptable that the expression of

most genes can be altered among biological sample replicas and that the total cellular RNA is not constant. Allelic bias in RNA-Seq can be, in part, attributed to the differential impact of the *in vitro* culture conditions (Gimelbrant et al., 2007; Weissbein et al., 2014). Thus, part of the ASE discordance observed between fibroblasts and iPSC in co-twins in our analysis may be due to acquired chromosomal abnormalities during the iPSC derivation and their propagation in culture.

Because the onset of MZ twinning, XCI, and genomic imprinting may occur at about the same time of embryological development (Machin, 1996), twinning may affect the distribution of cells bearing the inactivated X-chromosome or abnormal epigenetic marks of imprinting, and therefore, the varying manifestation of allelic differences from these processes. Surprisingly, the effect primarily occurs in female co-twins rather than male co-twins, and, thus, it is likely due to the presence of more than one X-chromosome in females (Lubinsky and Hall, 1991; Matias et al., 2014). Furthermore, there are cases of MZ female co-twins discordant for skewed XCI and imprinted disorders (Orstavik et al., 1995) and non-imprinted diseases (Bennett et al., 2008). Interestingly, 1,050 ASE sites map to 205 known and candidate imprinted genes. ASE discordance, yet at a considerably lower extent than the described here, has been reported between a pair of MZ “identical” co-twins clinically discordant for multiple sclerosis (Souren et al., 2016). Importantly, altered allelic expression of two imprinted genes (*ZNF331* and *GNAS*) and five non-imprinted genes (*ABLIM1*, *UBE2I*, *KIAA1267*, *CD6*, and *ATHL1*) were detected between the multiple sclerosis discordant co-twins.

Fourteen X-linked genes subjected to XCI (the non-escapee genes *UBL4A*, *FANCB*, *FTX*, *TAB3*, *WDR44*, *XIAP*, *IDS*, *MAP7D3*, *RLIM*, *RPL10*, *SLC9A7*, *TBC1D25*, *TLR7*, and *ZNF275*) in four female twin pairs exhibited ASE disparities in which one co-twin presented a biallelic profile and the sister female showed a monoallelic pattern. The RNA-Seq experiments were from pooled cells rather than single-cells and, thus, a biallelic model is the anticipated expression profile for non-escapee genes. The observed ASE disparities cannot be attributed to differences in cell culture conditions that result in decreased percentage of XCI mosaicism, which are expected to affect the expression profiles of all the 457 genes that are subjected to XCI (Carrel and Willard, 2005; Cotton et al., 2013; Balaton et al., 2015; Cotton et al., 2015; Tukiainen et al., 2017; Garieri et al., 2018; Shvetsova et al., 2019; Wainer Katsir and Linial, 2019).

What mechanisms might explain the observed ASE discordance in MZ co-twins? Data from prior RNA-Seq studies in co-twins (Baranzini et al., 2010; Lin et al., 2012a; Brown et al., 2014; Hibaoui et al., 2014; Buil et al., 2015; Dixon et al., 2015; Ding et al., 2017; Santoni et al., 2017) indicate that the differential allele expression of autosomal genes best reflects dynamic regulation processes consistent with either an allele being preferentially silenced or an inactive allele being restored. The biallelic expression of genes is a regulatory mechanism that outbalances the harmful effects of pathogenic expression-altering or loss-of-function risk variant alleles (Adzhubei et al., 2010; Landrum et al., 2016; Vaser et al., 2016). In each human euploid somatic cell, autosomal genes are anticipated to be

symmetrically expressed from both the parental alleles in a cell type-specific manner throughout development. However, the biallelic RNA expression pattern is not a phenotypic hallmark of all genes since 10–30% of human autosomal genes assayed for polymorphic variant sites [i.e., expressed SNPs (eSNPs) or eSNVs] are dynamically subjected to the epigenetic phenomena of clonal (mitotically) stable, random monoallelic expression (Gimelbrant et al., 2007; Savova et al., 2016a; Savova et al., 2016b; Savova et al., 2017), or allelic bias (Dixon et al., 2015). Most enigmatic, genes that are biallelically expressed in a cell can be regulated in a neighboring cell to randomly switch their RNA expression from biallelic to monoallelic at a time (Chess, 2013; Eckersley-Maslin and Spector, 2014; Eckersley-Maslin et al., 2014). Also, distinct subsets of autosomal and X-linked genes are subjected to epigenetic silencing of one allele, in a parent-of-origin dependent manner by autosomal genomic imprinting (Baran et al., 2015) or in a random fashion by XCI (i.e., in females) (Tukiainen et al., 2017).

ASE discordance in X-linked genes that are subjected to XCI has been reported between MZ female co-twins in humans (Cheung et al., 2008; Antonarakis et al., 2018) and in mice (Wang et al., 2010). Subtle departure from equal allelic expression ratios is often genetically determined in cis (i.e., eQTLs) and trans, but part of the disparity can also be ascribed to the random sampling effect of X inactivation (Carrel and Willard, 2005; Cheung et al., 2008; Cotton et al., 2013; Balaton et al., 2015; Cotton et al., 2015; Tukiainen et al., 2017; Garieri et al., 2018; Shvetsova et al., 2019; Wainer Katsir and Linial, 2019). Moreover, allelic imbalance on the X-chromosome could also affect autosomal allelic expression. Notwithstanding, we note that the extent of ASE discordance in homokaryotypic male twin-pairs (on average, 3.2%, $n = 1,478$ discordant sites) is comparable genome-wide to that in female twin pairs (2.5%, $n = 958$ discordant sites).

There are genetic and functional consequences of the autosomal variant sites in genes that are expressed from a single allele in one cell at a time. Mainly, i) they bestow more extensive genetic diversity in humans (Savova et al., 2016a); ii) they often are gain-of-function rather than pathogenic expression-altering or loss-of-function risk variants (i.e., for neurodevelopmental disorders), and influence expression variance in cis; iii) the range of expression level of monoallelically expressed genes is higher than biallelically expressed genes (Savova et al., 2017); iv) ultimately, they increase cell-to-cell expression variability with a beneficial impact of avoiding genetic disease phenotypes (Savova et al., 2016a).

The extranumerary chromosome 21 in trisomic cells of Down syndrome patients is well known to result in genome-wide dysregulation of gene expression represented by chromosomal domains with genes whose expression levels are copy-dosage compensated, upregulated, or downregulated as compared with euploid cells (Letourneau et al., 2014). In the co-twins discordant for T21 and 21q recombination, the ASE discrepant profiles can be viewed ultimately as unexplained heritability or missing heritability due to the discordance in trisomy 21, recombination at 21q, altered genomic imprinting, random monoallelic expression, and RNA editing. Allelic imbalance (allelic-specific heterogeneity) in dosage-sensitive genes can

arise by a stochastic adaptive regulation in both euploid and aneuploid cells as a consequence of low-level mRNA abundance and increased transcriptional burst frequency, rather than burst size (Deng and Distèche, 2019; Larsson et al., 2019; Symmons et al., 2019). The extent of the biallelic imbalance across eSNVs most likely reflects a gene network expression effect operating in the form of eQTLs (Mott et al., 2014; Pettigrew et al., 2016). Thus, part of the unexplained heritability or missing heritability could be explained by differences in cell-specific gene interactions. Moreover, the ASE profile discrepancies between fibroblasts and iPSC could be due to non-imprinted parental origin effects in each cell type associated with the aneuploidy. For example, the rs93366794 eSNP exhibited a concordant biallelic profile in iPSC but discordant in fibroblasts, being monoallelic in the T21 twin and biallelic in the normal co-twin. Interestingly, in an RNA-Seq study of a healthy brain, the *WRD4* gene bearing the rs93366794 site was reported to be expressed monoallelically from the paternal allele, but a mono-to-biallelic switch occurred in the offspring with *versus* without ASD (Lin et al., 2018).

Although the analysis presented here allowed the identification of a pervasive disparity in ASE profiles between co-twins, the biological significance of the extent and breadth of the observed differences in allele expression must only be assessed by independent experiments. However, the following results are of worth noting: i) among the eSNVs, there were several alleles known to be associated with disease conditions or predicted to be pathogenic; ii) the canonical imprinted gene *SNURF*, which is expressed monoallelically in over 50 tissues in the GTEx dataset, was expressed biallelically in iPSCs; iii) in all the 10 twin sets, there was expressed mitochondrial microheteroplasmy; iv) among all the genes expressed in the 10 twin pairs, there were 55 ± 17 genes that exhibited elevated proportions (ranging from 50 to 100%) of ASE sites coincident with RNA editing sites.

The breadth and magnitude of ASE discordance disclose unprecedented epigenomic-wide inter-individual variation occurring in MZ co-twins. Prior ASE studies in MZ twins are restricted to a specific gene or gene sets and, therefore, do not uncover the apparent state of pervasive missing allelic heritability in MZ co-twins shown in the present study. Although independent validation through wet experimentation (i.e., allele-specific quantitative reverse transcription-PCR, RNA-fluorescence *in situ* hybridization, or allele-specific pyrosequencing) is required for the biologically relevant candidate ASE discordant sites (here regarded as potential leads), two critical implications emerge from the epigenomic-wide inter-individual variation observed in MZ co-twins: i) as in the case of inter-individual variation in DNA methylation (Maunakea et al., 2010; Bell et al., 2012; Young et al., 2017; Garg et al., 2018), ASE discordance may have to be looked at when assessing and calculating the impact of phenotypic variation in the differential susceptibility to specific human conditions and diseases (Skipper, 2008; Sun et al., 2013); ii) ASE discordance also might be considered instrumental for developing RNA biomarker signatures for forensic body fluid identification and kinship analysis (Blay et al., 2019).

The present systematic and integrative meta-analysis has three important limitations: sample size, the certainty of correct calling a positive eSNV site for a theoretical heterozygote position,

and comparisons made in three different cell types. First, the experimental setting must be viewed as a case-study regarding a heterokaryotypic MZ twin pair discordant for trisomic 21 and chromosome 21 distal recombination. Reported cases of heterokaryotypic MZ twin pairs are rare. In **Table S1**, we listed all relevant cases studied in the literature. Notwithstanding, there is only one RNA-Seq public (i.e., not controlled) study in a pair of heterokaryotypic MZ twins, namely, the selected discordant index case. We used the index case as a reference case to investigate whether the underlying discordance in karyotype and recombination affect the ASE profiles. We initially hypothesized that any likely discordance in ASE differences will be restricted to chromosome 21 and that the differences will be more significant across and beyond the recombination event. However, the initial analysis indicated the occurrence of genome-wide rather than chr21-restricted ASE discordance. To investigate whether the observed genome-wide ASE discordance was limited to the unique index case, we investigated nine pairs of homokaryotypic MZ twins. Surprisingly, we observed genome-wide discordance in ASE, similar in breadth and magnitude to that observed in the index case, albeit in different cell lines. Second, to decrease the chances of false-positive ASE calls, we called eSNV sites using base quality control Q30 and ≥ 12 read depth, which are selecting criteria that excel in stringency published reports of the kind (Q20 and ≥ 8 reads) (Cheung et al., 2008; Baran et al., 2015; Tan et al., 2017; Tukiainen et al., 2017). Because the probability of correct SNV calling increases at higher coverage levels for a theoretical heterozygote position, we provided results for three read coverages (12, 20, and 40 reads). Coverage of 40 reads provides a 99.9% probability of correct SNV call (**Datasets S3-S14**). Third, the observation of genome-wide ASE discordance in the same type of cell lines (nine MZ twin pairs) and different cell lines (index case) is a very reassuring remark. Cheung et al. (2008) used 100K Affymetrix SNP array on the same set of homokaryotypic MZ twin biosamples and identified 201 SNPs with significant evidence of differential allelic expression. Of those, we confirmed 137 eSNVs as discordant, 38 sites of which were common to the nine twin-pairs (**Datasets S4-S14**). Unfortunately, no public next-generation sequencing data are available for DNA-Seq and RNA-Seq matched biosamples from MZ twins. Thus, we cannot address at present the question of how much of the genetic variation does contribute to the total percentage of ASE in MZ twins.

CONCLUDING REMARKS

Our genome-wide scans for allelic expression discordance reveal an apparent state of pervasive missing allelic heritability in MZ co-twins. The extent and breadth of the ASE discordant sites are not exclusively associated with differences due to chromosomal aberrations and recombination, but also relate to the epigenome-wide differential allele expression phenomena of genomic imprinting and RNA editing. We conclude that most of ASE discordant sites observed within MZ pairs (either homokaryotypic or heterokaryotypic co-twins) cannot be attributed solely to the estimated within-pair incongruencies in DNA (Huang et al., 2019) or correspond to random

transcriptional allelic noise varying across experiments (Chakravarti, 2011). The epigenome-wide ASE discordance may have essential effects on physiology, phenotype, or inheritance, and implications for the Developmental Origins of Health and Disease (DOHaD) approach in co-twins (Yamada and Chong, 2017).

WEB RESOURCES

The URLs for public data used herein are as follows:

UCSC Genome Browser, <https://genome.ucsc.edu/>
 GTEx portal, <https://www.gtexportal.org/>
 NCBI SRA, <https://www.ncbi.nlm.nih.gov/sra/>
 NCBI GEO, <http://www.ncbi.nlm.nih.gov/geo/>
 dbGaP, <http://www.ncbi.nlm.nih.gov/gap>
 PhenoScanner, <http://www.phenoscanter.medschl.cam.ac.uk/phenoscanter>
 e-GRASP, <http://www.mypeg.info/egrasp>
 HaploReg, <http://compbio.mit.edu/HaploReg>
 PheGenI, <https://www.ncbi.nlm.nih.gov/gap/phegeni>
 Geneimprint, <http://www.geneimprint.com/>
 Metaimprint, <http://bioinfo.hrbmu.edu.cn/MetaImprint/>
 dbMAE, <https://mae.hms.harvard.edu/>
 OMIM, <http://www.omim.org>
 RADAR, <http://rnaedit.com/>
 Otago's Catalogue of Imprinted Genes, <http://igc.otago.ac.nz/>
 R software package, <http://www.R-project.org>
 GATK, <https://software.broadinstitute.org/gatk/>

DATA AVAILABILITY STATEMENT

RNA-Seq data experiments used in this manuscript are from publicly available BioProjects. The primary data can be accessed from the NCBI SRA repository (<https://www.ncbi.nlm.nih.gov/sra/>) under the following run entries: SRR1182244, SRR1182246, SRR1182249, SRR1182248, SRR1182252, SRR1182253, SRR1182245, SRR1182247, SRR1182250, SRR1182251, SRR1182254, SRR1182255, SRR1028343, SRR1028344, SRR1028345, SRR1028346, SRR1028347, SRR1028348, SRR1028349, SRR519874, SRR519875, SRR519876, SRR519877, SRR519878, SRR519879, SRR519880, SRR519881, SRR519882, SRR519883, SRR519884, SRR519885, SRR519886, SRR519887, SRR519888, SRR519889, SRR519890, SRR519891, SRR3390461, SRR3390473, SRR3389246, SRR3390437.

AUTHOR CONTRIBUTIONS

RJ, CF, JS, DM, EM-A: Conceived and designed experiments. RJ, CF, JS, YC, VR, GC, EM-A: Develop scripts and carried computational analyses. JS, DM, AG: performed SRA RNA-Seq analysis for control SNPs queries. RJ, CF, JS, DM, AG, EM-A: Performed cross-reference of ASE sites. DM, RJ, CF, JS, EM-A: Prepared figures. EM-A: Supervised the work and drafted the manuscript. All authors provided substantial

contributions to the interpretation of data, revised and approved the manuscript.

FUNDING

This study was supported by grants from the Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ (BR) (<http://www.faperj.br/>) [grant number E26/010.001036/2015 to EM-A] and from the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (BR) (<http://cnpq.br/>) [grant number 308780/2015-9 to EM-A]. JS and DM received undergraduate PIBIC/CNPq fellowships from the Universidade Estadual do Norte Fluminense Darcy Ribeiro UENF (BR) (<http://www.uenf.br/>). CF is a recipient of a graduate fellowship from the Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (<http://www.capes.gov.br/>). The agencies had no role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to the high-performance platform of the MultiUser Equipment (EMU) of the Center for Genomic Medicine at the Clinics Hospital of Ribeirão Preto (FMRP-USP) for the support, particularly to the Drs. Wilson Araújo da Silva Junior, Raul Torrieri, and Marcelo Gomes. Special thanks to Dr. Thiago Motta Venancio from the Universidade Estadual do Norte Fluminense for the support with access to his computer processing facility. The authors are grateful to the members of the Molecular Identification and Diagnostics Unit of the Laboratory of Biotechnology, Universidade Estadual do Norte Fluminense, for their insights and productive brainstorm lab meetings.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01178/full#supplementary-material>

DATASET S1 | RNA-Seq read counts and allele ratios supporting **Figure 1**.

DATASET S2 | BioProjects and RNA-Seq experiments used in this study.

DATASET S3 | ASE concordant and discordant sites identified in heterokaryotypic MZ twins assayed in fibroblasts and induced pluripotent stem cells (iPSC) derived from the co-twins discordant for trisomy 21 and maternal recombination at 21q.

DATASET S4 | Summary of the average distribution of eSNVs per gene in the entire set of biosamples and heterogeneity scores across genes. **(A)** Summary of informative of eSNVs, **(B)** Summary of eSNVs by ASE profile. Ranking of genes by ASE heterogeneity/homogeneity scores in either heterokaryotypic **(C)** and homokaryotypic **(D)** twins.

DATASET S5 | Concordant and discordant ASE sites identified in unrelated, non-twin males **(A)** and female **(B)** pairs assayed in peripheral mononuclear cells. Ranking of genes by ASE heterogeneity/homogeneity scores in unrelated, non-twin males and females **(C)**.

DATASETS S6–S14 | Concordant and discordant ASE sites identified in homokaryotypic MZ twins assayed in cultured B-cells derived from nine pairs of co-twins not discordant for a specific health condition.

DATASET S15 | Cross-reference of ASE sites from the ten monozygotic twin pairs with allele expression profiles from secondary data of the GTEx project in multiple tissues.

DATASET S16 | Summary of A-to-I(G) RNA-editing sites resulting in synonymous and non-synonymous substitutions in the ten twin pairs set.

DATASET S17 | Protein-coding genetic eSNVs associated with disease risk or disease pathogenesis in the ten twin pairs set.

DATASET S18 | Gene ontology enrichment annotations.

TABLE S1 | Review of cases of heterokaryotypic co-twins reported in the literature.

TABLE S2 | Distribution of ASE sites by gene structure sequence context. Cross-reference of ASE sites disparities observed in the present study with the annotation features of the eSNV disparities found in the whole-genome sequencing study by Huang et al. 2019.

TABLE S3 | ASE disparities in X-linked genes subject to XCI between MZ twins.

TABLE S4 | Evidence for allele-specific expression form of mitochondrial microheteroplasmy in MZ twin pairs.

FIGURE S1 | Representation of the discordant maternal 21q inheritance in the pair of co-twins heterokaryotypic for trisomy 21 reported by Dahoun et al. (2008). The MZ co-twins are discordant for trisomy 21 of maternal origin in twin 1 (T1DS) and maternal allelic disparity at 21qter likely carried by disomic twin 2 (T2N). The discordant inheritance of 21q is probably due to meiosis I subtelomeric recombination event likely occurring between the maternal chromosomes 21 within the 1.7Mb interval (hg38) delimited by the short tandem repeat marker D21S1445, where alleles were identical in both twins and the short tandem repeat marker D21S1611, where different alleles were inherited.

FIGURE S2 | Flowchart of analysis. The in-house computational pipeline, PipASE, used for scanning and sorting out genome-wide, allele-specific differences between MZ co-twins.

FIGURE S3 | Chromosomal distribution of eSNVs in iPSC. **(A)** Genome-wide e-karyotyping for the SNVs exhibiting allele-specific expression in iPSC derived from the co-twins discordant for T21 and maternal recombination at 21q. Shown is the distribution of all ASE sites that were concordant (gray ticks towards the left of each chromosome ideogram) or discordant (blue ticks towards the right side). **(B)** Assessment of chromosomal aberrations by e-karyotyping allelic bias using RNA-Seq data from iPSC in **(A)**.

FIGURE S4 | Overview of the breadth and magnitude of allele-specific expression disparity between nine MZ control twin pairs. The RNA-Seq SRA entries for the nine twin pairs are SRR519874, SRR519875, SRR519876, SRR519877, SRR519878, SRR519879, SRR519880, SRR519881, SRR519882, SRR519883, SRR519884, SRR519885, SRR519886, SRR519887, SRR519888, SRR519889, SRR519890, and SRR519891. For each SRA entry above, the panels are **(A)** numbers of ASE sites distributed by the within-pair status of concordance or discordance in control homokaryotypic MZ control twins tested in cultured B-cells. The majority of ASE sites are concordant for a biallelic imbalance status. On average, the co-twins are discordant in 1074 ± 252.03 ASE sites. **(B)** Comparison of the effect size of LogASE. We calculated the \log_2 of allele-specific expression fold change using the equation $\text{LogASE} = \log_2(T1_ASE / T2_ASE)$ for each eSNV in each tissue. LogASE estimates the magnitude of expression change between conditions for the variant. **(C)** Distribution of genes by numbers of ASE sites observed in cultured B-cells.

FIGURE S5 | Chromosomal distribution of eSNVs in the homokaryotypic twin pairs. Genome-wide e-karyotyping for the SNVs exhibiting allele-specific expression in cultured B-cells from nine control twin pairs. Shown in each panel, **(A)** through **(I)**, is the distribution of all ASE sites that were concordant (gray ticks towards the left

of each chromosome ideogram) or discordant (red ticks towards the right side). The RNA-Seq SRA entries for the nine twin pairs are SRR519874, SRR519875, SRR519876, SRR519877, SRR519878, SRR519879, SRR519880, SRR519881, SRR519882, SRR519883, SRR519884, SRR519885, SRR519886, SRR519887, SRR519888, SRR519889, SRR519890, and SRR519891, respectively.

FIGURE S6 | Assessment of chromosomal aberrations by e-karyotyping allelic bias using RNA-Seq data from control twin pairs. (A) through (I). The RNA-Seq SRA entries for the nine twin pairs used as controls are SRR519874, SRR519875, SRR519876, SRR519877, SRR519878, SRR519879, SRR519880, SRR519881, SRR519882, SRR519883, SRR519884, SRR519885, SRR519886, SRR519887, SRR519888, SRR519889, SRR519890, and SRR519891, respectively. For each SRA entry above, a plot is shown, which represents the distribution of allele ratios in cultured B-cells from nine pairs of co-twins who are not discordant for a specific health condition. None of the control twin pairs present detectable chromosomal aberrations.

FIGURE S7 | Within twin-pair disparities in allele expression proportions at eSNVs that are coincident with canonical A-to-I(G) RNA editing sites in

homokaryotypic twins. Shown is the distribution of eSNVs that positionally match canonical RNA editing sites between nine homokaryotypic co-twins, (A) through (I), assayed in culture-B-cells. Each dot corresponds to an eSNV. The vast majority of sites exhibited a concordant biallelic imbalance profile (pink and light blue dots). Red dots represent eSNPs that were discordant between co-twins in that they exhibited allelic proportions differences higher than 25%, regardless of the discordance or concordance in the karyotype. Green dots represent eSNVs that exhibited discordant allelic profiles, being biallelic in one twin and monoallelic in the other. The linear models (solid black lines), the confidence interval of the models (broken purple lines) and of the prediction (solid purple lines) were constructed using R. Model equations: (A) $Y = 8.15439 + 0.82362X$; (B) $Y = 7.20067 + 0.80502X$; (C) $Y = 6.0758 + 0.8080X$; (D) $Y = 12.74607 + 0.78267X$; (E) $Y = 14.78510 + 0.79019X$; (F) $Y = 5.64298 + 0.81255X$; (G) $Y = 7.97718 + 0.82152X$; (H) $Y = 8.36441 + 0.81231X$; (I) $Y = 6.60496 + 0.72912X$. For all pairs, $P < 2.2e-16$. The RNA-Seq SRA entries for the nine twin pairs used as controls are SRR519874, SRR519875, SRR519876, SRR519877, SRR519878, SRR519879, SRR519880, SRR519881, SRR519882, SRR519883, SRR519884, SRR519885, SRR519886, SRR519887, SRR519888, SRR519889, SRR519890, and SRR519891.

REFERENCES

- Abdellaoui, A., Ehli, E. A., Hottenga, J. J., Weber, Z., Mbarek, H., Willemsen, G., et al. (2015). CNV Concordance in 1,097 MZ Twin Pairs. *Twin Res. Hum. Genet.* 18, 1–12. doi: 10.1017/thg.2014.86
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Alves Da Silva, A. F., Machado, F. B., Pavarino, E. C., Biselli-Perico, J. M., Zampieri, B. L., Da Silva Francisco Junior, R., et al. (2016). Trisomy 21 Alters DNA Methylation in Parent-of-Origin-Dependent and -Independent Manners. *PLoS One* 11, e0154108. doi: 10.1371/journal.pone.0154108
- Antonarakis, M. G., Georgios, S., Xavier, B., Emilie, F., Pascale, R., Christelle, B., et al. (2018). Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-specific expression in human fibroblasts. doi: 10.1073/pnas.1806811115
- Babak, T., Deveale, B., Tsang, E. K., Zhou, Y., Li, X., Smith, K. S., et al. (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat. Genet.* 47, 544–549. doi: 10.1038/ng.3274
- Balaton, B. P., Cotton, A. M., and Brown, C. J. (2015). Derivation of consensus inactivation status for X-linked genes from genome-wide studies. *Biol. Sex Differ.* 6, 35. doi: 10.1186/s13293-015-0053-7
- Baran, Y., Subramaniam, M., Biton, A., Tukiainen, T., Tsang, E. K., Rivas, M. A., et al. (2015). The landscape of genomic imprinting across diverse adult human tissues. *Genome Res.* 25, 927–936. doi: 10.1101/gr.192278.115
- Baranzini, S. E., Mudge, J., Van Velkinburgh, J. C., Khankhanian, P., Khrebtukova, I., Miller, N. A., et al. (2010). Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 464, 1351–1356. doi: 10.1038/nature08990
- Begemann, M., Rezwani, F. I., Beygo, J., Docherty, L. E., Kolarova, J., Schroeder, C., et al. (2018). Maternal variants in NLRP and other maternal effect proteins are associated with multilocus imprinting disturbance in offspring. *J. Med. Genet.* 55, 497–504. doi: 10.1136/jmedgenet-2017-105190
- Bell, J. T., Tsai, P. C., Yang, T. P., Pidsley, R., Nisbet, J., Glass, D., et al. (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 8, e1002629. doi: 10.1371/journal.pgen.1002629
- Bennett, C. M., Boye, E., and Neufeld, E. J. (2008). Female monozygotic twins discordant for hemophilia A due to nonrandom X-chromosome inactivation. *Am. J. Hematol.* 83, 778–780. doi: 10.1002/ajh.21219
- Blay, N., Casas, E., Galvan-Femenia, I., Graffelman, J., De Cid, R., and Vavouri, T. (2019). Assessment of kinship detection using RNA-seq data. *Nucleic Acids Res.* gkz776. doi: 10.1093/nar/gkz776 10.1101/546937.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brown, A. A., Buil, A., Vinuela, A., Lappalainen, T., Zheng, H. F., Richards, J. B., et al. (2014). Genetic interactions affecting human gene expression identified by variance association mapping. *Elife* 3, e01381. doi: 10.7554/eLife.01381
- Buil, A., Brown, A. A., Lappalainen, T., Vinuela, A., Davies, M. N., Zheng, H. F., et al. (2015). Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* 47, 88–91. doi: 10.1038/ng.3162
- Carrel, L., and Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434, 400–404. doi: 10.1038/nature03479
- Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16, 195. doi: 10.1186/s13059-015-0762-6
- Chaiyasap, P., Kulawongnuchai, S., Srichomthong, C., Tongsimma, S., Suphapeetiporn, K., and Shotelersuk, V. (2014). Whole genome and exome sequencing of monozygotic twins with trisomy 21, discordant for a congenital heart defect and epilepsy. *PLoS One* 9, e100191. doi: 10.1371/journal.pone.0100191
- Chakravarti, A. (2011). Widespread promiscuous genetic information transfer from DNA to RNA. *Circ. Res.* 109, 1202–1203. doi: 10.1161/RES.0b013e31823c4992
- Chamberlain, S., Ushey, K., and Zhu, H. (2018). rsnp: Get 'SNP' ('Single-Nucleotide' 'Polymorphism') Data on the Web [Online]. Available at URL: <https://cran.r-project.org/web/packages/rsnp/>.
- Chess, A. (2013). Random and non-random monoallelic expression. *Neuropsychopharmacology* 38, 55–61. doi: 10.1038/npp.2012.85
- Cheung, V. G., Bruzel, A., Burdick, J. T., Morley, M., Devlin, J. L., and Spielman, R. S. (2008). Monozygotic twins reveal germline contribution to allelic expression differences. *Am. J. Hum. Genet.* 82, 1357–1360. doi: 10.1016/j.ajhg.2008.05.003
- Core Team, R. (2019). R: A language and environment for statistical computing [Online]. Available at URL: <https://www.R-project.org/>.
- Cotton, A. M., Ge, B., Light, N., Adoue, V., Pastinen, T., and Brown, C. J. (2013). Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol.* 14, R122. doi: 10.1186/gb-2013-14-11-r122
- Cotton, A. M., Price, E. M., Jones, M. J., Balaton, B. P., Kobor, M. S., and Brown, C. J. (2015). Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. *Hum. Mol. Genet.* 24, 1528–1539. doi: 10.1093/hmg/ddu564
- Dahoun, S., Gagos, S., Gagnebin, M., Gehrig, C., Burgi, C., Simon, F., et al. (2008). Monozygotic twins discordant for trisomy 21 and maternal 21q inheritance: a complex series of events. *Am. J. Med. Genet. A* 146A, 2086–2093. doi: 10.1002/ajmg.a.32431
- De Sa Machado Araujo, G., Da Silva Francisco Junior, R., Dos Santos Ferreira, C., Mozer Rodrigues, P. T., Terra Machado, D., Louvain De Souza, T., et al. (2018). Maternal 5(m)CpG Imprints at the PARD6G-AS1 and GCSAML Differentially Methylated Regions Are Decoupled From Parent-of-Origin Expression Effects in Multiple Human Tissues. *Front. Genet.* 9, 36. doi: 10.3389/fgene.2018.00036

- Deng, X., and Disteche, C. M. (2019). Rapid transcriptional bursts upregulate the X chromosome. *Nat. Struct. Mol. Biol.* 26, 851–853. doi: 10.1038/s41594-019-0314-y
- Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Deveale, B., Van Der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.* 8, e1002600. doi: 10.1371/journal.pgen.1002600
- Ding, N., Zhang, Z., Yang, W., Ren, L., Zhang, Y., Zhang, J., et al. (2017). Transcriptome analysis of monozygotic twin brothers with childhood primary myelofibrosis. *Genomics Proteomics Bioinf.* 15, 37–48. doi: 10.1016/j.gpb.2016.12.002
- Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature* 518, 331–336. doi: 10.1038/nature14222
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., et al. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21, 3439–3440. doi: 10.1093/bioinformatics/bti525
- Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191. doi: 10.1038/nprot.2009.97
- Eckersley-Maslin, M. A., and Spector, D. L. (2014). Random monoallelic expression: regulating gene expression one allele at a time. *Trends Genet.* 30, 237–244. doi: 10.1016/j.tig.2014.03.003
- Eckersley-Maslin, M. A., Thybert, D., Bergmann, J. H., Marioni, J. C., Flicek, P., and Spector, D. L. (2014). Random monoallelic gene expression increases upon embryonic stem cell differentiation. *Dev. Cell* 28, 351–365. doi: 10.1016/j.devcel.2014.01.017
- Essaoui, M., Nizon, M., Beaujard, M. P., Carrier, A., Tantau, J., De Blois, M. C., et al. (2013). Monozygotic twins discordant for 18q21.2qter deletion detected by array CGH in amniotic fluid. *Eur. J. Med. Genet.* 56, 502–505. doi: 10.1016/j.ejmg.2013.06.007
- Furukawa, H., Oka, S., Matsui, T., Hashimoto, A., Arinuma, Y., Komiyama, A., et al. (2013). Genome, epigenome and transcriptome analyses of a pair of monozygotic twins discordant for systemic lupus erythematosus. *Hum. Immunol.* 74, 170–175. doi: 10.1016/j.humimm.2012.11.007
- Garg, P., Joshi, R. S., Watson, C., and Sharp, A. J. (2018). A survey of inter-individual variation in DNA methylation identifies environmentally responsive co-regulated networks of epigenetic variation in the human genome. *PLoS Genet.* 14, e1007707. doi: 10.1371/journal.pgen.1007707
- Garieri, M., Stamoulis, G., Blanc, X., Falconnet, E., Ribaux, P., Borel, C., et al. (2018). Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-specific expression in human fibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* 115, 13015–13020. doi: 10.1073/pnas.1806811115
- Gilbert, B., Yardin, C., Briault, S., Belin, V., Lienhardt, A., Aubard, Y., et al. (2002). Prenatal diagnosis of female monozygotic twins discordant for Turner syndrome: implications for prenatal genetic counselling. *Prenat. Diagn.* 22, 697–702. doi: 10.1002/pd.383
- Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., and Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science* 318, 1136–1140. doi: 10.1126/science.1148910
- Gray, T. A., Saitoh, S., and Nicholls, R. D. (1999). An imprinted, mammalian bicistronic transcript encodes two independent proteins. *Proc. Natl. Acad. Sci. U.S.A.* 96, 5616–5621. doi: 10.1073/pnas.96.10.5616
- Heap, G. A., Yang, J. H., Downes, K., Healy, B. C., Hunt, K. A., Bockett, N., et al. (2010). Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum. Mol. Genet.* 19, 122–134. doi: 10.1093/hmg/ddp473
- Hibaoui, Y., Grad, I., Letourneau, A., Sailani, M. R., Dahoun, S., Santoni, F. A., et al. (2014). Modelling and rescuing neurodevelopmental defect of Down syndrome using induced pluripotent stem cells from monozygotic twins discordant for trisomy 21. *EMBO Mol. Med.* 6, 259–277. doi: 10.1002/emmm.201302848
- Hu, Y. J., Sun, W., Tzeng, J. Y., and Perou, C. M. (2015). Proper Use of Allele-Specific Expression Improves Statistical Power for cis-eQTL Mapping with RNA-Seq Data. *J. Am. Stat. Assoc.* 110, 962–974. doi: 10.1080/01621459.2015.1038449
- Huang, Y., Zhao, Y., Ren, Y., Yi, Y., Li, X., Gao, Z., et al. (2019). Identifying genomic variations in monozygotic twins discordant for autism spectrum disorder using whole-genome sequencing. *Mol. Ther. Nucl. Acids* 14, 204–211. doi: 10.1016/j.omtn.2018.11.015
- International Hapmap, C. (2003). The International HapMap Project. *Nature* 426, 789–796. doi: 10.1038/nature02168
- Jirtle, J., and Murphy, C. (2012). Geneimprint database [Online]. Available: <http://www.geneimprint.com> [Accessed September 2017].
- Knopman, J. M., Krey, L. C., Oh, C., Lee, J., Mccaffrey, C., and Noyes, N. (2014). What makes them split? Identifying risk factors that lead to monozygotic twins after in vitro fertilization. *Fertil. Steril.* 102, 82–89. doi: 10.1016/j.fertnstert.2014.03.039
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44, D862–D868. doi: 10.1093/nar/gkv1222
- Larsson, A. J. M., Coucoravas, C., Sandberg, R., and Reinius, B. (2019). X-chromosome upregulation is driven by increased burst frequency. *Nat. Struct. Mol. Biol.* 26, 963–969. doi: 10.1038/s41594-019-0306-y
- Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25, 1841–1842. doi: 10.1093/bioinformatics/btp328
- Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118. doi: 10.1371/journal.pcbi.1003118
- Letourneau, A., Santoni, F. A., Bonilla, X., Sailani, M. R., Gonzalez, D., Kind, J., et al. (2014). Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature* 508, 345–350. doi: 10.1038/nature13200
- Leung, W. C., Choi, H., Lau, W. L., Ng, L. K., Lau, E. T., Lo, F. M., et al. (2009). Monozygotic dichorionic twins heterokaryotypic for duplication chromosome 2q13-q23.3. *Fetal Diagn. Ther.* 25, 397–399. doi: 10.1159/000236153
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M., et al. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333, 53–58. doi: 10.1126/science.1207018
- Lin, M., Hrabovsky, A., Pedrosa, E., Wang, T., Zheng, D., and Lachman, H. M. (2012a). Allele-biased expression in differentiating human neurons: implications for neuropsychiatric disorders. *PLoS One* 7, e44017. doi: 10.1371/journal.pone.0044017
- Lin, W., Piskol, R., Tan, M. H., and Li, J. B. (2012b). Comment on widespread RNA and DNA sequence differences in the human transcriptome. *Science* 335, 1302. doi: 10.1126/science.1210624 author reply 1302.
- Lin, C. Y., Chang, K. W., Lin, C. Y., Wu, J. Y., Coon, H., Huang, P. H., et al. (2018). Allele-specific expression in a family quartet with autism reveals mono-to-biallelic switch and novel transcriptional processes of autism susceptibility genes. *Sci. Rep.* 8, 4277. doi: 10.1038/s41598-018-22753-4
- Liu, Z., Yang, J., Xu, H., Li, C., Wang, Z., Li, Y., et al. (2014). Comparing computational methods for identification of allele-specific expression based on next generation sequencing data. *Genet. Epidemiol.* 38, 591–598. doi: 10.1002/gepi.21846
- Liu, S., Hong, Y., Cui, K., Guan, J., Han, L., Chen, W., et al. (2018). Four-generation pedigree of monozygotic female twins reveals genetic factors in twinning process by whole-genome sequencing. *Twin Res. Hum. Genet.* 21, 361–368. doi: 10.1017/thg.2018.41
- Lo, H. S., Wang, Z., Hu, Y., Yang, H. H., Gere, S., Buetow, K. H., et al. (2003). Allelic variation in gene expression is common in the human genome. *Genome Res.* 13, 1855–1862. doi: 10.1101/gr.1006603
- Lubinsky, M. S., and Hall, J. G. (1991). Genomic imprinting, monozygous twinning, and X inactivation. *Lancet* 337, 1288. doi: 10.1016/0140-6736(91)92956-3
- Machin, G. A. (1996). Some causes of genotypic and phenotypic discordance in monozygotic twin pairs. *Am. J. Med. Genet.* 61, 216–228. doi: 10.1002/(SICI)1096-8628(19960122)61:3<216::AID-AJMG5>3.0.CO;2-S

- Marinov, G. K., Williams, B. A., Mccue, K., Schroth, G. P., Gertz, J., Myers, R. M., et al. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* 24, 496–510. doi: 10.1101/gr.161034.113
- Matias, A., Silva, S., Martins, Y., and Blickstein, I. (2014). Monozygotic twins: ten reasons to be different. *Diagnóstico Prenatal.* 25, 53–57. doi: 10.1016/j.diapre.2013.09.003
- Maunakea, A. K., Chepelev, I., and Zhao, K. (2010). Epigenome mapping in normal and disease States. *Circ. Res.* 107, 327–339. doi: 10.1161/CIRCRESAHA.110.222463
- Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Metsalu, T., Viltrop, T., Tiirats, A., Rajashekar, B., Reimann, E., Koks, S., et al. (2014). Using RNA sequencing for identifying gene imprinting and random monoallelic expression in human placenta. *Epigenetics* 9, 1397–1409. doi: 10.4161/15592294.2014.970052
- Morgan, M. (2017). AnnotationHub: Client to access AnnotationHub resources. R package version 2.14.5 [Online]. Available at URL: <https://bioconductor.org/packages/release/bioc/html/AnnotationHub.html>
- Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., et al. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature* 430, 743–747. doi: 10.1038/nature02797
- Mott, R., Yuan, W., Kaisaki, P., Gan, X., Cleak, J., Edwards, A., et al. (2014). The architecture of parent-of-origin effects in mice. *Cell* 156, 332–342. doi: 10.1016/j.cell.2013.11.043
- Moybrailean, G. A., Richards, A. L., Kurtz, D., Kalita, C. A., Davis, G. O., Harvey, C. T., et al. (2016). High-throughput allele-specific expression across 250 environmental conditions. *Genome Res.* 26, 1627–1638. doi: 10.1101/gr.209759.116
- Nieuwint, A., Van Zalen-Sprock, R., Hummel, P., Pals, G., Van Vugt, J., Van Der Harten, H., et al. (1999). 'Identical' twins with discordant karyotypes. *Prenat. Diagn.* 19, 72–76. doi: 10.1002/(SICI)1097-0223(199901)19:1<72::AID-PD465>3.0.CO;2-V
- Orstavik, R. E., Tommerup, N., Eiklid, K., and Orstavik, K. H. (1995). Non-random X chromosome inactivation in an affected twin in a monozygotic twin pair discordant for Wiedemann-Beckwith syndrome. *Am. J. Med. Genet.* 56, 210–214. doi: 10.1002/ajmg.1320560219
- Pettigrew, K. A., Frinton, E., Nudel, R., Chan, M. T. M., Thompson, P., Hayiou-Thomas, M. E., et al. (2016). Further evidence for a parent-of-origin effect at the NOP9 locus on language-related phenotypes. *J. Neurodev. Disord.* 8, 24. doi: 10.1186/s11689-016-9157-6
- Pickrell, J. K., Gilad, Y., and Pritchard, J. K. (2012). Comment on widespread RNA and DNA sequence differences in the human transcriptome. *Science* 335, 1302. doi: 10.1126/science.1210484 author reply 1302.
- Pirinen, M., Lappalainen, T., Zaitlen, N. A., Consortium, G. T., Dermitzakis, E. T., Donnelly, P., et al. (2015). Assessing allele-specific expression across multiple tissues from RNA-seq read data. *Bioinformatics* 31, 2497–2504. doi: 10.1093/bioinformatics/btv074
- Piskol, R., Peng, Z., Wang, J., and Li, J. B. (2013). Lack of evidence for existence of noncanonical RNA editing. *Nat. Biotechnol.* 31, 19–20. doi: 10.1038/nbt.2472
- Raghupathy, N., Choi, K., Vincent, M. J., Beane, G. L., Sheppard, K., Munger, S. C., et al. (2018). Hierarchical analysis of rRNA-seq reads improves the accuracy of allele-specific expression. *Bioinformatics* 34, 2177–2184. doi: 10.1093/bioinformatics/bty078
- Ramaswami, G., and Li, J. B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* 42, D109–D113. doi: 10.1093/nar/gkt996
- Ramaswami, G., Lin, W., Piskol, R., Tan, M. H., Davis, C., and Li, J. B. (2012). Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods* 9, 579–581. doi: 10.1038/nmeth.1982
- Richard Albert, J., Koike, T., Younesy, H., Thompson, R., Bogutz, A. B., Karimi, M. M., et al. (2018). Development and application of an integrated allele-specific pipeline for methylomic and epigenomic analysis (MEA). *BMC Genomics* 19, 463. doi: 10.1186/s12864-018-4835-2
- Santoni, F. A., Stamoulis, G., Garieri, M., Falconnet, E., Ribaux, P., Borel, C., et al. (2017). Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression. *Am. J. Hum. Genet.* 100, 444–453. doi: 10.1016/j.ajhg.2017.01.028
- Savova, V., Chun, S., Sohail, M., Mccole, R. B., Witwicki, R., Gai, L., et al. (2016a). Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. *Nat. Genet.* 48, 231–237. doi: 10.1038/ng.3493
- Savova, V., Patsenker, J., Vigneau, S., and Gimelbrant, A. A. (2016b). dbMAE: the database of autosomal monoallelic expression. *Nucleic Acids Res.* 44, D753–D756. doi: 10.1093/nar/gkv1106
- Savova, V., Vinogradova, S., Pruss, D., Gimelbrant, A. A., and Weiss, L. A. (2017). Risk alleles of genes with monoallelic expression are enriched in gain-of-function variants and depleted in loss-of-function variants for neurodevelopmental disorders. *Mol. Psychiatry* 22, 1785–1794. doi: 10.1038/mp.2017.13
- Scott, J. M., and Ferguson-Smith, M. A. (1973). Heterokaryotypic monozygotic twins and the acardiac monster. *J. Obstet. Gynaecol. Br. Commonw.* 80, 52–59. doi: 10.1111/j.1471-0528.1973.tb02131.x
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311. doi: 10.1093/nar/29.1.308
- Shvetsova, E., Sofronova, A., Monajemi, R., Gagalova, K., Draisma, H. H. M., White, S. J., et al. (2019). Skewed X-inactivation is common in the general female population. *Eur. J. Hum. Genet.* 27, 455–465. doi: 10.1038/s41431-018-0291-3
- Skipper, M. (2008). Gene expression - One allele or two?. *Nat. Rev. Genet.* 9, 4–5. doi: 10.1038/nrg2287
- Smigrodzki, R. M., and Khan, S. M. (2005). Mitochondrial microheteroplasmy and a theory of aging and age-related disease. *Rejuvenation Res.* 8, 172–198. doi: 10.1089/rej.2005.8.172
- Soderlund, C. A., Nelson, W. M., and Goff, S. A. (2014). Allele Workbench: transcriptome pipeline and interactive graphics for allele-specific expression. *PLoS One* 9, e115740. doi: 10.1371/journal.pone.0115740
- Souren, N. Y., Gerdes, L. A., Kumpfel, T., Lutsik, P., Klopstock, T., Hohlfeld, R., et al. (2016). Mitochondrial DNA Variation and Heteroplasmy in Monozygotic Twins Clinically Discordant for Multiple Sclerosis. *Hum. Mutat.* 37, 765–775. doi: 10.1002/humu.23003
- Sun, C., Burgner, D. R., Ponsonby, A. L., Saffery, R., Huang, R. C., Vuillemin, P. J., et al. (2013). Effects of early-life environment and epigenetics on cardiovascular disease risk in children: highlighting the role of twin studies. *Pediatr. Res.* 73, 523–530. doi: 10.1038/pr.2013.6
- Symmons, O., Chang, M., Mellis, I. A., Kalish, J. M., Park, J., Susztak, K., et al. (2019). Allele-specific RNA imaging shows that allelic imbalances can arise in tissues through transcriptional bursting. *PLoS Genet.* 15, e1007874. doi: 10.1371/journal.pgen.1007874
- Tachon, G., Lefort, G., Puechberty, J., Schneider, A., Jeandel, C., Boulot, P., et al. (2014). Discordant sex in monozygotic XXY/XX twins: a case report. *Hum. Reprod.* 29, 2814–2820. doi: 10.1093/humrep/deu275
- Tan, M. H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A. N., et al. (2017). Dynamic landscape and regulation of RNA editing in mammals. *Nature* 550, 249–254. doi: 10.1038/nature24041
- The GTEx Project, C. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660. doi: 10.1126/science.1262110
- Tukiainen, T., Villani, A. C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., et al. (2017). Landscape of X chromosome inactivation across human tissues. *Nature* 550, 244–248. doi: 10.1038/nature24265
- Van Der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinf.* 43, 11 11–33. doi: 10.1002/0471250953.bi1110s43
- Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., and Ng, P. C. (2016). SIFT missense predictions for genomes. *Nat. Protoc.* 11, 1–9. doi: 10.1038/nprot.2015.123
- Vigneau, S., Vinogradova, S., Savova, V., and Gimelbrant, A. (2018). High prevalence of clonal monoallelic expression. *Nat. Genet.* 50, 1198–1199. doi: 10.1038/s41588-018-0188-7
- Von Hippel, P. T. (2015). The heterogeneity statistic I(2) can be biased in small meta-analyses. *BMC Med. Res. Methodol.* 15, 35. doi: 10.1186/s12874-015-0024-z
- Wainer Katsir, K., and Linal, M. (2019). Human genes escaping X-inactivation revealed by single cell expression data. *BMC Genomics* 20, 201. doi: 10.1186/s12864-019-5507-6

- Wang, X., and Clark, A. G. (2014). Using next-generation RNA sequencing to identify imprinted genes. *Heredity (Edinb)* 113, 156–166. doi: 10.1038/hdy.2014.18
- Wang, X., Soloway, P. D., and Clark, A. G. (2010). Paternally biased X inactivation in mouse neonatal brain. *Genome Biol.* 11, R79. doi: 10.1186/gb-2010-11-7-r79
- Wei, Y., Su, J., Liu, H., Lv, J., Wang, F., Yan, H., et al. (2014). MetaImprint: an information repository of mammalian imprinted genes. *Development* 141, 2516–2523. doi: 10.1242/dev.105320
- Weissbein, U., Benvenisty, N., and Ben-David, U. (2014). Quality control: genome maintenance in pluripotent stem cells. *J. Cell Biol.* 204, 153–163. doi: 10.1083/jcb.201310135
- Weissbein, U., Schachter, M., Egli, D., and Benvenisty, N. (2016). Analysis of chromosomal aberrations and recombination by allelic bias in RNA-Seq. *Nat. Commun.* 7, 12144. doi: 10.1038/ncomms12144
- Weksberg, R., Shuman, C., Caluseriu, O., Smith, A. C., Fei, Y. L., Nishikawa, J., et al. (2002). Discordant KCNQ1OT1 imprinting in sets of monozygotic twins discordant for Beckwith-Wiedemann syndrome. *Hum. Mol. Genet.* 11, 1317–1325. doi: 10.1093/hmg/11.11.1317
- Wood, D. L., Nones, K., Steptoe, A., Christ, A., Harliwong, I., Newell, F., et al. (2015). Recommendations for accurate resolution of gene and isoform allele-specific expression in RNA-Seq data. *PLoS One* 10, e0126911. doi: 10.1371/journal.pone.0126911
- Yamada, L., and Chong, S. (2017). Epigenetic studies in developmental origins of health and disease: pitfalls and key considerations for study design and interpretation. *J. Dev. Orig. Health Dis.* 8, 30–43. doi: 10.1017/S2040174416000507
- Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., and Kinzler, K. W. (2002). Allelic variation in human gene expression. *Science* 297, 1143. doi: 10.1126/science.1072545
- Young, P. E., Kum Jew, S., Buckland, M. E., Pamphlett, R., and Suter, C. M. (2017). Epigenetic differences between monozygotic twins discordant for amyotrophic lateral sclerosis (ALS) provide clues to disease pathogenesis. *PLoS One* 12, e0182638. doi: 10.1371/journal.pone.0182638
- Zhou, Z. Y., Hu, Y., Li, A., Li, Y. J., Zhao, H., Wang, S. Q., et al. (2018). Genome wide analyses uncover allele-specific RNA editing in human and mouse. *Nucleic Acids Res.* 46, 8888–8897. doi: 10.1093/nar/gky613

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 da Silva Francisco Junior, dos Santos Ferreira, Santos e Silva, Terra Machado, Côrtes Martins, Ramos, Simões Carnivari, Garcia and Medina-Acosta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.