Somatic mosaicism

ORIGINAL ARTICLE

Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair

Gülşah M Dal,[1] Bekir Ergüner,[2] Mahmut S Sağıroğlu,[2] Bayram Yüksel,[3] Onur Emre Onat,[1] Can Alkan,[4] Tayfun Özçelik[1,5]

[1]Department of Molecular Biology and Genetics, Faculty of Science, Bilkent University, Ankara, Turkey
[2]TÜBİTAK-BİLGEM-UEKAE, Gebze, Kocaeli, Turkey
[3]TÜBİTAK-MAM-GEBI, Gebze, Kocaeli, Turkey
[4]Faculty of Engineering, Department of Computer Engineering, Bilkent University, Ankara, Turkey
[5]Institute of Materials Science and Nanotechnology (UNAM), Bilkent University, Ankara, Turkey

Correspondence to
Dr Tayfun Özçelik, Faculty of Science, Department of Molecular Biology and Genetics, Bilkent University, Ankara 06800, Turkey; tozcelik@bilkent.edu.tr

![CrossMark]

ABSTRACT

Background Human de novo single-nucleotide variation (SNV) rate is estimated to range between $0.82–1.70\times10^{-8}$ mutations per base per generation. However, contribution of early postzygotic mutations to the overall human de novo SNV rate is unknown.

Methods We performed deep whole-genome sequencing (more than 30-fold coverage per individual) of the whole-blood-derived DNA samples of a healthy monozygotic twin pair and their parents. We examined the genotypes of each individual simultaneously for each of the SNVs and discovered de novo SNVs regarding the timing of mutagenesis. Putative de novo SNVs were validated using Sanger-based capillary sequencing.

Results We conservatively characterised 23 de novo SNVs shared by the twin pair, 8 de novo SNVs specific to twin I and 1 de novo SNV specific to twin II. Based on the number of de novo SNVs validated by Sanger sequencing and the number of callable bases of each twin, we calculated the overall de novo SNV rate of 1.31×10^{-8} and 1.01×10^{-8} for twin I and twin II, respectively. Of these, rates of the early postzygotic de novo SNVs were estimated to be 0.34×10^{-8} for twin I and 0.04×10^{-8} for twin II.

Conclusions Early postzygotic mutations constitute a substantial proportion of de novo mutations in humans. Therefore, genome mosaicism resulting from early mitotic events during embryogenesis is common and could substantially contribute to the development of diseases.

INTRODUCTION

Mutations provide the means of genetic diversity on which natural selection operates. Characterising the patterns and rate of de novo mutations is crucial to the understanding of mutational processes that may lead to evolution and human disease.[1] Recent studies based on the direct observation of de novo single-nucleotide variation (SNV) in parent–offspring trios revealed a rate that spans a wide range ($0.82–1.70\times10^{-8}$ mutations per base per generation).[2][3] However, the proportion of prezygotic versus early postzygotic mutations within the de novo mutation rate is not yet known.

In an attempt to further elucidate the origins of de novo mutations, we performed deep whole-genome sequencing of a healthy monozygotic twin pair and their parents (figure 1A). Our working hypothesis is that those de novo mutations observed in only one twin would be of mitotic origin, occur early during development possibly in the post-twinning embryo and be detectable in whole-blood-derived constitutive DNA (figure 1B).

We indeed observed, following a conservative characterisation and validation, 23 shared, 8 twin I-specific and 1 twin II-specific de novo SNVs (table 1). These results implicate that a substantial proportion of de novo SNVs in the genome of an individual could be of early postzygotic origin.

STUDY SAMPLES AND METHODS

Study samples

A quad family consisting of a male 25-year-old healthy monozygotic twin pair and their parents (59 and 49 years old) participated in this study. They were recruited to the control group of a study on movement disorders, which was approved by the institutional review boards (IRB) at Bilkent, Hacettepe, Başkent and Çukurova Universities (decisions: BEK02, 28.08.2008; TBK08/4, 22.04.2008; KA07/47, 02.04.2007; and 21/3, 08.11.2005, respectively). Zygosity of the twin pair was assessed by comparing the SNVs of each twin (99.8% similar). A Mendelian disorder has not been documented in the family. The participants had no exposure to chemotherapeutics. Written informed consent prepared according to the guidelines of the Ministry of Health in Turkey was attained from all individuals prior to the initiation of this study. Blood, buccal smear and urine samples were obtained from all participants. DNA isolation from whole blood of each individual was done using NucleoSpin Blood (Macherey-Nagel) kit, according to the manufacturer's protocol. Buccal smear and urine cell-derived DNA were isolated using DNeasy Tissue Kit (QIAGEN).

Whole-genome sequencing

Whole-genome sequencing data were generated from blood DNA using Illumina HiSeq 2000 instrument. DNA samples were prepared for whole-genome sequencing using Illumina TruSeq DNA sample preparation kit according to the manufacturer's recommended protocols. For each of the parent single library and for each of the twins two libraries were prepared. For library preparation, 1–3 μg of genomic DNA was fragmented through sonication. The resulting DNA fragments were end-repaired and adapter ligation was done. End-repaired and adapter-ligated DNA fragments were run on 2% agarose gel for size selection. Following the isolation of the fragments of approximately 400 bp in size from the gel using QIAGEN MinElute Gel Extraction Kit, size-selected libraries were subjected to PCR amplification. Enriched libraries were purified and their
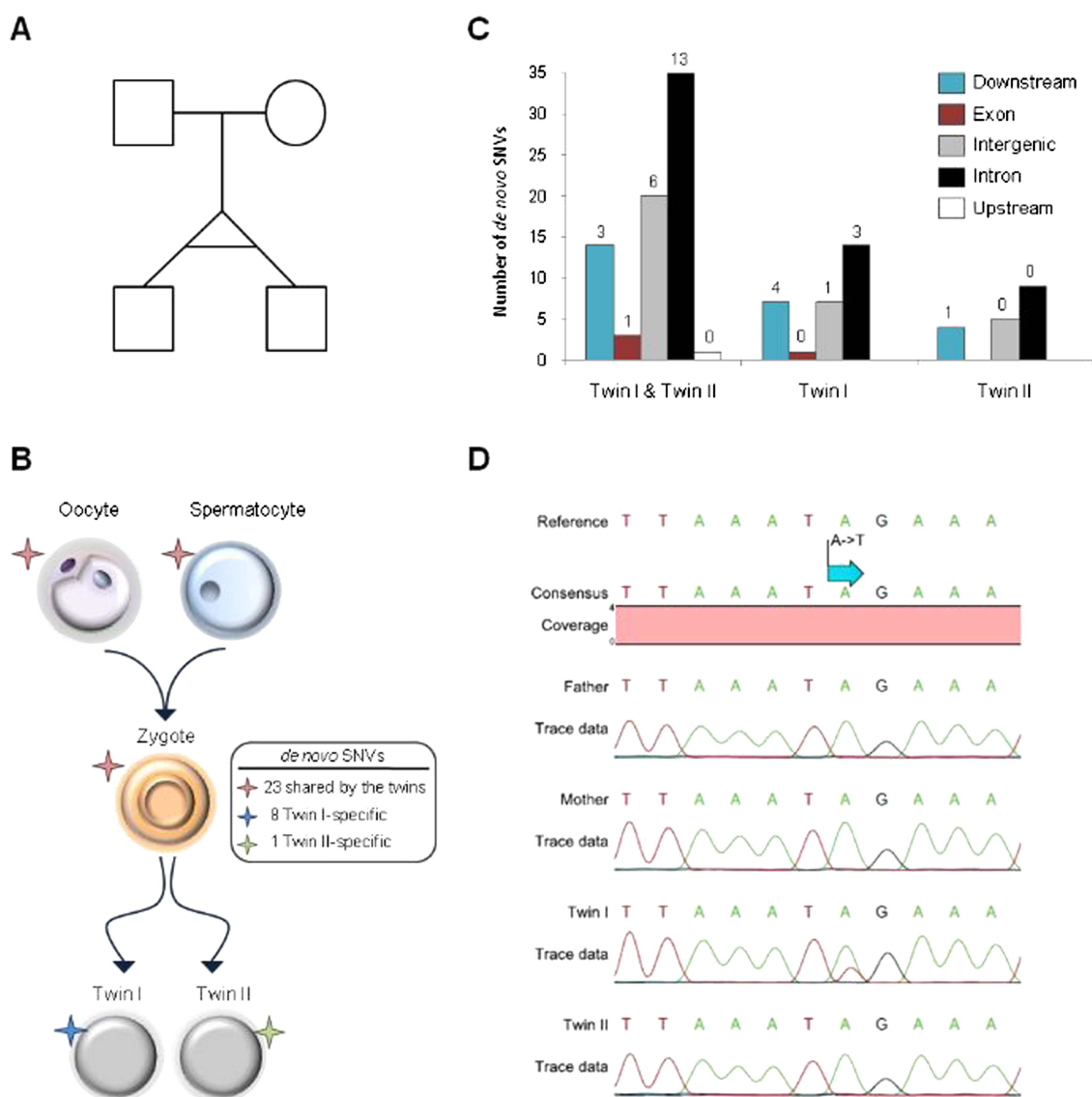
Figure 1 Discovery and characterisation of de novo single-nucleotide variations (SNVs) in a monozygotic twin pair. (A) Pedigree of the whole-genome sequenced quad family. (B) Prezygotic versus postzygotic de novo SNVs found in twins I and II. (C) Annotation of the putative de novo SNVs against human database GRCh37.68 using SnpEff,[13] twin I and twin II: de novo SNVs shared between the twin pair. Twin I, twin II: twin I-specific and twin II-specific de novo SNVs, respectively. The number of validated SNVs is shown above each bar. The only validated exonic SNV shared by the twins is a synonymous mutation in PRSS58. (D) Sanger sequencing chromatogram for a twin I-specific intronic de novo SNV (XPO1 gene, chr2: 61728351, A>T).

Table 1 De novo mutations of the twin pair and the mutation rate estimation

	All SNVs*		Novel SNVs†		De novo SNVs‡			Validated de novo SNVs§			Mutation rate (×10⁻⁸)¶		
Sample	dbSNP137	Novel	Inherited	De novo**	Group I††	Group II‡‡	Total	Group I	Group II	Total	Group I	Group II	Total
Twin I	3 446 849	49 800	48 405	1396	159	83	242	23	8	31	0.97	0.34	1.31
Twin II	3 443 828	49 182	47 810	1372		48	207		1	24		0.04	1.01

*Number of all SNVs based on the variants calling through GATK.
†Number of SNVs that are not reported dbSNP137.
‡Number of de novo SNVs after applying filters (see online supplementary tables S5–S7).
§De novo SNVs validated through Sanger sequencing.
¶Mutation rate calculated based on the validated de novo SNVs and callable bases of each twin.
**Number of novel SNVs detected as de novo before applying filters (see online supplementary tables S5–S7).
††De novo SNVs classified as occurred in parental gametes or before separation of the zygote into two embryos.
‡‡De novo SNVs classified as occurred after separation of the zygote into two embryos but very early during development.
GATK, Genome Analysis Toolkit; SNV, single-nucleotide variation.

quality was controlled with respect to the recommendations provided by Illumina. Resulting libraries were sequenced on eight lanes for each twin and four lanes for each parent using an Illumina Hiseq 2000 instrument, employing 2×101–104 cycles of incorporation. Imaging was performed using Illumina SBS kits TruSeq V.3. Image analysis and base calling were done using Illumina's Real Time Analysis software V.1.13 with default parameters.

Analysis of the whole-genome sequencing data

We converted base calling data into FASTQ files by using Illumina's CASAVA V.1.8.2 software package. The paired-end reads were aligned to the NCBI Build 37 version (GRCh37) of the human reference genome using Burrows-Wheeler Aligner (BWA)[4] V.0.6.1 with standard parameters. BAM files were generated from the resulting alignment files and merged into a single BAM file using SAMtools[5] V.0.1.18. We also used SAMtools to mark and remove the PCR duplicates found in the resulting BAM file. SNVs and insertions and/or deletions (indels) were discovered using Genome Analysis Toolkit (GATK)[6] software V.1.6–13 with standard filtering parameters. In order to minimise spurious SNV calls due to misalignment of bases in regions around indels, the reads were locally realigned using RealignerTargetCreator and IndelRealigner tools of GATK. UnifiedGenotyper tool of GATK was used to generate an initial list of raw variants. Finally, variant recalibration with the variant quality score recalibration was applied to the SNVs/indels in order to generate the final list of variants of four individuals in a single VCF file. GATK-recommended hard filtering for very low quality call-sets was applied with 'AB <0.2 ‖ MQ0 >50 ‖ SB >−0.10 ‖ QUAL <10' parameters.

Discovery of the putative de novo SNVs shared by the twins (group I)

We used the following approaches to discover putative de novo SNVs and to classify them according to the timing of mutagenesis. We selectively extracted SNVs (both novel and reported in dbSNP137) using GATK's 'SelectVariants' module. We compared the genotypes of the parents and the twin pair for each of the SNVs concurrently using SnpSift[7] V.1.1.3. We evaluated all possible genotype combinations, including those that could be classified as genotyping errors (such as those SNVs in a homozygous state in one of the twins yet not present in one of the parents). Using BEDtools,[8] we first filtered the SNVs reported in SNP database dbSNP137. Novel putative de novo SNVs were removed when they were in non-variant sites or in sites for which the read depth is less than 3 in each individual simultaneously. We filtered the variants intersecting segmental duplications and/or simple repeats in the genome using the intersectBed tool within the BEDtools[8] suite. Finally, we excluded de novo SNV candidates classified as genotyping errors. Remaining de novo SNV candidates were selected for further analysis. We considered the timing of mutagenesis for the shared de novo SNVs as parental or postzygotic yet prior to twining.

Discovery of twin-specific putative de novo SNVs (group II)

This class of de novo SNVs was discovered using the above-mentioned strategy including the comparison of the genotypes of the mother, father and the twin pair for all SNVs simultaneously. SNVs observed in one of the twins and not observed in the other twin as well as both parents were considered as candidate de novo SNVs. Same set of filters described above was applied to the resulting candidate de novo SNVs to generate a list of high-confidence putative twin-specific de novo SNVs. We

considered the timing of mutagenesis for the twin-specific de novo SNVs as early postzygotic after the twinning event.

Evaluation of the putative de novo SNVs based on visual examination

We used Integrative Genomics Viewer (IGV)[9] to examine the Sequence Alignment/Map (BAM) files to further evaluate the putative de novo SNVs. We classified the variants in our initial call-set as (a) inherited, (b) no variant, (c) located on a low-quality read and (d) de novo. In the presence of more than three reads supporting the variant in one or both parents, the variant was classified as inherited and filtered. When the SNV or indel density was high around the region flanking 40 bp the putative de novo SNVs, we classified the sequencing reads supporting the variants as 'low-quality reads' and filtered them. Additionally, the putative de novo SNVs that are not supported by at least three reads and located on the 'low-quality reads' were classified as 'no variant', and thus filtered. We included the remaining putative de novo SNVs in our final call-set and selected them for validation by Sanger-based capillary sequencing (see online supplementary table S1).

Validation of de novo SNVs with Sanger sequencing

We selected all 120 high-confidence putative de novo SNVs for validation. Of these, 37 could not be tested due to unavailability of appropriate primers or PCR products of sufficient quality for capillary sequencing. We designed primer pairs using Primer3[10] and PerlPrimer[11] (see online supplementary table S2). PCR was performed using 25 ng genomic DNA, One *Taq* DNA Polymerase (New England Biolabs), 25 mM dNTPs (Fermentas), 5X One *Taq* standard reaction buffer and 10 pmol of each forward and reverse primers. Thermocycling conditions were as recommended by the manufacturer (94C° for 30 s; 30–35 cycles of 94C° for 30 s, 58–60°C for 40 s and 68C° for 40 s; and 68C° for 5 min). PCR products were analysed through 0.8% agarose gel electrophoresis. Following the purification of PCR products, Sanger sequencing was performed using forward primer, reverse primer or both. CLC Main Workbench software package (CLCBio Inc) was used to analyse Sanger sequencing data.

Calculation of the reference allele frequency

We used reference allele frequency (RAF) to evaluate the putative de novo SNVs. RAF is the total number of reads supporting the reference allele as a fraction of total number of reads supporting both reference and alternative allele.[12] For the putative de novo SNVs that are present in a heterozygous state, RAF threshold was determined as 0.5–0.7 for the twins and 0.95–1.0 for the parents.

RESULTS

Mapping of the whole-genome sequencing data and discovery of the sequence variants

We sequenced the whole genomes of the monozygotic twin pair and their parents using the Illumina HiSeq2000 platform and generated a total of 538 Gb genomic sequence with a mean coverage depth of more than 30-fold per individual (see online supplementary table S3). Using BWA,[4] we mapped the paired-end sequencing reads to the human reference genome (GRCh37) and discovered genomic variants using the GATK.[6]

Herein, we discovered genomic variants using the multisample calling options of UnifiedGenotyper (a total of 170-fold coverage). We identified approximately 3.4 million SNVs and 750 000 small insertions and deletions (indels) per genome (see online supplementary table S4).

Identification of the putative de novo SNVs

Examination of the genotypes of the twin pair and their parents synchronously for each of the SNVs identified an initial set of putative de novo SNVs. To distinguish true de novo SNVs from sequencing and variant calling artefacts, we progressively applied the following set of filters: (a) novelty vis-á-vis dbSNP137, (b) having a coverage depth of at least three for the site encompassing the variant in each individual, (c) presence outside of the segmental duplications and/or repetitive DNA and (d) genotyping error (see 'Methods'). With this filtering strategy, we identified a total of 290 high-confidence putative de novo SNVs. Of these, 159 were shared by the twin pair (group I); and 83 were specific for twin I and 48 for twin II (group II) (see online supplementary tables S5–S7).

Selection of the high-confidence putative de novo SNVs for validation

Subsequent to the identification of the initial call-set, we visually examined the BAM files for all of the 290 high-confidence putative de novo SNVs using IGV.[9] We also considered the allelic depth for each SNV in each individual. This resulted in the filtering of 86/159 shared, 54/83 twin I-specific and 30/48 twin II-specific de novo SNVs (see 'Methods' and online supplementary figure S1 and tables S8–S11).

We selected the remaining de novo SNV candidates for Sanger sequencing (73 shared, 29 twin I-specific and 18 twin II-specific) and also classified them by genomic context, which revealed a biased distribution towards non-coding regions of the genome (figure 1C).[13] This is expected because of two reasons: (a) non-coding regions constitute most of the genome, and (b) they are subjected to less stringent evolutionary pressure.

We calculated the average RAF and coverage depth values for all of the putative de novo SNVs (a) shared by the twin pair, (b) specific to twin I and (c) specific to twin II (see 'Methods' and online supplementary table S12). As expected, average RAF is between 0.95 and 0.97 for the parents. For the twin pair, average RAF for shared de novo SNVs is 0.59 for twin I and 0.55 for twin II, supporting the heterozygous presence of the variants. And for the twin-specific de novo SNV candidates, RAF values are consistent with mosaic presence (0.68–0.69) in the twin that carries the variant.

Experimental validation of de novo SNV candidates

Of the 73 shared de novo SNV candidates, 49 could be tested using capillary sequencing, and 23 confirmed to be de novo SNV in heterozygous state (see online supplementary table S8). We considered the timing of mutagenesis as parental or pre-twinning zygote for these 23 shared de novo SNVs.

Of the 29 twin I-specific and 18 twin II-specific high-confidence putative de novo SNVs (group II), 19 and 15, respectively, could be analysed by capillary sequencing; and 8/19 for twin I and 1/15 for twin II were confirmed (see online supplementary tables S9 and S10).

In order to control the validity of the filtering process that targeted the putative de novo SNVs classified as genotyping error, we selected all 34 for capillary sequencing. Of these, 19 could be tested, and none confirmed as a de novo SNV (see online supplementary table S13).

We confirmed the exclusive representation of twin-specific de novo SNVs by simultaneous Sanger sequencing of the whole family. The mutant allele for these twin-specific de novo SNVs was clearly visible along with the wild-type allele in the capillary chromatograms (figure 1D).

We calculated one more time the average RAF and coverage depth values for the confirmed de novo SNVs as described above (see online supplementary table S14). Average RAF is 0.96–1.00 for the parents. For the twin pair, it is 0.53 (twin I) and 0.46 (twin II) for the shared; and 0.65 (twin I) and 0.78 (twin II) for the twin-specific de novo SNVs. These results are consistent with the Sanger sequencing chromatograms, which indicate the mosaic presence of twin-specific de novo SNVs.

Mosaic representation of a mutation in a parent could be interpreted as a de novo mutation in the twins.[14–16] We therefore extended our de novo mutation validation in blood (mesoderm) to the analysis of DNA samples isolated from the buccal (ectoderm) and the urine (endoderm) cells of the parents. We did not observe the de novo SNVs in either cell type, which suggests that parental mosaicism does not contribute to the shared de novo SNVs in the twins (see online supplementary figure S2).

Calculation of the mutation rate

Mutation rate was calculated based on the number of validated de novo SNVs and callable bases. Bases with a minimum read depth of 3 and do not intersect segmental duplications were considered as callable. We calculated 2 703 445 479 bases for twin I and 2 702 251 568 bases for twin II. We used the formula below to calculate the mutation rates[3]:

$$\mu = (N \times (1 + FNR)/L$$

At this point, N represents the number of de novo SNVs validated by capillary sequencing, FNR is the false-negative rate (see online supplementary table S15) and L is the total number of callable bases.

We calculated a rate of 0.97×10^{-8} bp per generation for de novo SNVs shared by the twin pair. For twin-specific de novo SNVs, we calculated the rate of 0.34×10^{-8} bp per generation for twin I and 0.04×10^{-8} bp per generation for twin II, respectively.

DISCUSSION

The rate of de novo SNVs in healthy individuals has been investigated through whole-genome sequencing of mother–father–offspring trios and yielded rates ranging between $0.82–1.70 \times 10^{-8}$ per base pair.[17] Among these overall rates, the early postzygotic de novo SNVs have not been distinguished from those which have parental origin. Here we present overall de novo SNV rate of 1.01 and 1.31×10^{-8} in a healthy monozygotic twin pair consistent with previous direct estimations. We also observe that parental mosaicism does not contribute to the shared de novo SNVs in the twins. Furthermore, we report the contribution of early postzygotic mutations to the average de novo SNV burden with a rate of $0.04–0.34 \times 10^{-8}$. Our deep whole-genome sequencing-based approach combined with effective filtering and validation enabled the direct measurement of early postzygotic versus prezygotic de novo SNV rate.

The very first studies concerned with the estimation of somatic mutation rate were restricted to coding or non-coding regulatory regions and mainly focused on disease phenotypes.[18 19] The average somatic mutation rate has been reported to be 7.7×10^{-10}.[20] A more recent study that surveyed the whole genome estimated the somatic mutation rate for common SNPs in healthy monozygotic twins to be 1.2×10^{-7} per base pair.[12] Conrad and his colleagues distinguished the germline and non-germline (somatic and cell line derived) mutations through whole-genome sequencing of a CEU trio and provided a rate of 2.52×10^{-7} per nucleotide for non-germline mutations. This estimate, which is greater than the estimate presented in this study, is probably influenced by mutagenicity and age of the cell culture.[21]

De novo mutations that are not shared by monozygotic twins could be cell line derived.[22] We excluded this possibility by directly sequencing whole-blood-derived DNA. We, however, point out that the rate of postzygotic mutations is likely to be underestimated. This is because mutations that occur later in embryo development will be observed as very low-frequency mosaic SNVs that will be regarded as sequencing errors by mapping and variant calling algorithms.[23]

High levels of allelic, locus and phenotypic heterogeneity have important implications for gene discovery in complex human diseases.[24] Genomic microarrays and next-generation sequencing technologies are now enabling researchers to dissect the molecular basis of complex phenotypes that arise from de novo mutations.[25] This has been demonstrated in common neurodevelopmental diseases such as schizophrenia[26] and autistic-spectrum disorders,[27] pointing to a monogenic basis of disease with the mutation representing a single event of large effect. However, even after the spectacular success of these modern genetic studies, a majority of cases remain unsolved. An added level of complexity, consistent with this 'de novo model', could be the timing of mutational events whereby early postzygotic de novo mutations could be critically important. This is supported by the observations that somatic mosaicism has been well documented in Mendelian phenotypes,[14 28 29] including monozygotic twins discordant for a given disorder,[30] and by demonstration of extensive genetic variation in human tissues[31] including brain.[32] It will be crucial to expand genome sequencing studies to the next level of tissue or even cell type-specific interrogation to better delineate causal mutations especially in sporadic forms of Mendelian disorders.

REFERENCES

1 Haldane JBS. The rate of spontaneous mutation of a human gene. *J Genet* 1935;31:317–26.

2 Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WS, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U, Stefansson K. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012;488:471–5.

3 Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, Han L, Vives L, O'Roak BJ, Sudmant PH, Shendure J, Abney M, Ober C, Eichler EE. Estimating the human mutation rate using autozygosity in a founder population. *Nat Genet* 2012;44:1277–80.

4 Li H, Durbin L. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.

5 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.

6 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.

7 Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front Genet* 2012;15:3.

8 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.

9 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.

10 Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 2007;23:1289–91.

11 Marshall OJ. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* 2004;20:2471–7.

12 Li R, Montpetit A, Rousseau M, Wu SY, Greenwood CM, Spector TD, Pollak M, Polychronakos C, Richards JB. Somatic point mutations occurring early in development: a monozygotic twin study. *J Med Genet* 2014;51:28–34.

13 Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80–92.

14 Darras BT, Francke U. A partial deletion of the muscular dystrophy gene transmitted twice by an unaffected male. *Nature* 1987;329:556–8.

15 Biesecker LG, Spinner NB. A genomic view of mosaicism an human disease. *Nat Rev Genet* 2013;14:307–20.

16 Ozcelik T. Uncovering the complex mysteries of mosaicism. *Nature* 2002;417:588.

17 Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. *Trends Genet* 2013;29:575–84.

18 Araten DJ, Golde DW, Zhang RH, Thaler HT, Gargiulo L, Notaro R, Luzzatto L. A quantitative measurement of the human somatic mutation rate. *Cancer Res* 2005;65:8111–17.

19 Iwama T. Somatic mutation rate of the APC gene. *Jpn J Clin Oncol* 2001;31:185–7.

20 Lynch M. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci USA* 2010;107:961–8.

21 Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P; 1000 Genomes Project. Variation in genome-wide mutation rates within and between human families. *Nat Genet* 2011;43:712–14.

22 Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, McCarroll SA. Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am J Hum Genet* 2012;91:1033–40.

23 Baranzini SE, Mudge J, van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW, May GD, Woodward JE, Caillier SJ, McElroy JP, Gomez R, Pando MJ, Clendenen LE, Ganusova EE, Schilkey FD, Ramaraj T, Khan OA, Huntley JJ, Luo S, Kwok PY, Wu TD, Schroth GP, Oksenberg JR, Hauser SL, Kingsmore SF. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 2010;464:1351–6.

24 McClellan J, King MC. Genetic heterogeneity in human disease. *Cell* 2010;141:210–17.

25 Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet* 2012;13:565–75.

26 Gulsuner S, Walsh T, Watts AC, Lee MK, Thornton AM, Casadei S, Rippey C, Shahin H, Consortium on the Genetics of Schizophrenia (COGS); PAARTNERS Study GroupNimgaonkar VL, Go RC, Savage RM, Swerdlow NR, Gur RE, Braff DL, King MC, McClellan JM. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 2013;154:518–29.

27 Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M. De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012;74:285–99.

28 Stern C. Genetic mosaics in animals and man. In: Stern C. eds. *Genetic mosaics and other essays*. Cambridge, Massachusetts: Harvard University Press, 1968:27–129.

29 Lupski JR. Genome mosaicism—one human, multiple genomes. *Science* 2013;341:358–9.

30 Vadlamudi L, Dibbens LM, Lawrence KM, Iona X, McMahon JM, Murrell W, Mackay-Sim A, Scheffer IE, Berkovic SF. Timing of de novo mutagenesis—a twin study of sodium-channel mutations. *N Engl J Med* 2010;363:1335–40.

31 O'Huallachain M, Karczewski KJ, Weissman SM, Urban AE, Snyder MP. Extensive genetic variation in somatic human tissues. *Proc Natl Acad Sci USA* 2012;109:18018–23.

32 Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, Talbot RT, Gustincich S, Freeman TC, Mattick JS, Hume DA, Heutink P, Carninci P, Jeddeloh JA, Faulkner GJ. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 2011;479:534–7.

JMG

Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair

Gülsah M Dal, Bekir Ergüner, Mahmut S Sagiroglu, et al.

Updated information and services can be found at:

http://jmg.bmj.com/content/51/7/455.full.html

These include:

Data Supplement	*"Supplementary Data"* http://jmg.bmj.com/content/suppl/2014/04/25/jmedgenet-2013-102197.DC1.html
References	This article cites 31 articles, 11 of which can be accessed free at: http://jmg.bmj.com/content/51/7/455.full.html#ref-list-1
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections	Articles on similar topics can be found in the following collections Epidemiology (575 articles) Reproductive medicine (490 articles)

Notes

To request permissions go to:

http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:

http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:

http://group.bmj.com/subscribe/