



RESEARCH ARTICLE

A characterization of postzygotic mutations identified in monozygotic twins

Klaasjan G. Ouwens^{1,2} | Rick Jansen³ | Bas Tolhuis² | P. Eline Slagboom⁴ |
Brenda W.J.H. Penninx³ | Dorret I. Boomsma¹¹Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands²Genalix Core BV, Nijkerk, The Netherlands³Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands⁴Department of Molecular Epidemiology, Leids Universitair Medisch Centrum, Leiden, The Netherlands**Correspondence**

Klaasjan G. Ouwens, Department of Biological Psychology, VU University Amsterdam, Van der Boerhorstraat 1, 1081 BT Amsterdam, The Netherlands.

Email: klaasjan.ouwens@vu.nl

Funding information

BBMRI-NL, Grant/Award Number: NWO 184.021.007; EMGO Institute for Health and Care Research

Communicating by Bruce R. Gottlieb

Abstract

Postzygotic mutations are DNA changes acquired from the zygote stage onwards throughout the lifespan. These changes lead to differences in DNA sequence among cells of an individual, potentially contributing to the etiology of complex disorders. Here we compared whole genome DNA sequence data of two monozygotic twin pairs, 40 and 100 years old, to detect somatic mosaicism. DNA samples were sequenced twice on two Illumina platforms (13X and 40X read depth) for increased specificity. Using differences in allelic ratios resulted in sets of 1,720 and 1,739 putative postzygotic mutations in the 40-year-old twin pair and 100-year-old twin pair, respectively, for subsequent enrichment analysis. This set of putative mutations was strongly ($p < 4.37e-91$) enriched in both twin pairs for regulatory elements. The corresponding genes were significantly enriched for genes that are alternatively spliced, and for genes involved in GTPase activity. This research shows that somatic mosaicism can be detected in monozygotic twin pairs by using allelic ratios calculated from DNA sequence data and that the mutations which are found by this approach are not randomly distributed throughout the genome.

KEYWORDS

genetics, mutations, next-generation sequencing, postzygotic mutations, somatic mosaicism

1 | INTRODUCTION

When the DNA of a person does not encompass the same sequence in every cell of the body, but contains de novo postzygotic genetic mutations in a fraction of the cells only, the person is considered a mosaic. This is different from genetic chimerism, where a single organism is composed from different zygotes, for example, after transplacental exchange between mother and child or between twins (vanDijk, Boomsma, & deMan, 1996). In the literature, mosaicism has been described in phenotypically discordant monozygotic twin pairs (Zwijnenburg, Meijers-Heijboer, & Boomsma, 2010) and postzygotic mutations are considered as a possible cause for such twin discordance. Recent research found that a fraction of presumed germline de novo mutations are actually either postzygotic or inherited as a consequence of low-level mosaicism in one of the parents (Acuna-Hidalgo et al., 2015; Rahbari et al., 2016), and that early postzygotic mutations could account for a substantial proportion of de novo single nucleotide variants (SNVs) in the genome of an individual (Dal et al.,

2014). Postzygotic mosaic mutations are usually associated with cancer development (Abyzov et al., 2017; Biesecker & Spinner, 2013; Buntinx, Campbell, & van den Akker, 2014; Cohen, Wilson, Trinh, & Ye, 2015; Forsberg, Absher, & Dumanski, 2013; Iourov, Vorsanova, & Yurov, 2010; Jacobs et al., 2012; Laurie et al., 2012), but they are also an important confounder in medical genetic testing (Forsberg, Gisselsson, & Dumanski, 2017). A recent study suggests that somatic mosaicism in the brain might represent a potential mechanism contributing to neuronal diversity and the etiology of neuropsychiatric disorders (McConnell et al., 2017). The number of clonal SNVs has been estimated at 1,000–1,500 per neuronal genome which are enriched in coding exons (Lodato et al., 2015).

Monozygotic (MZ) twins are often called identical twins, since they are presumed to have no differences at the level of the DNA sequence. MZ twins arise from one fertilized oocyte, and there is a chance of a somatic mutation at each subsequent mitosis. The moment in life at which these mutations occur determines whether they are present in both twins, in only one twin, or even in only a fraction of the cells in one

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Human Mutation* published by Wiley Periodicals, Inc.

twin (Martin, Boomsma, & Machin, 1997). Throughout life, limitations in somatic cell maintenance lead to accumulation of mutations. Aging is thought to be a consequence of this accumulation (Jacobs et al., 2012; Veitia, Govindaraju, Bottani, & Birchler, 2017). Since mechanisms of cellular and molecular aging are inherently stochastic, this will cause MZ twins to diverge (Kirkwood, 2005).

An example of large-scale genomic mosaicism is the phenomenon of early postzygotic mitotic nondisjunction resulting in MZ twins having different numbers of chromosomes. There have been documented cases where twins are mosaic 45,X/46,XY and discordant for phenotypic sex (Reindollar, Byrd, Hahn, Haseltine, & McDonough, 1987). The same is possible for MZ discordances for autosomal trisomies (e.g., Down syndrome [MIM: 190685]). Compared to germline aneuploidies, many more mosaic aneuploidies have been found to be compatible with life, including monosomy 7 and 18, and trisomies 7, 8, 9, 12, 14, 15, 16, 17, and 20 (Machin, 1996). This is an indication that postzygotic mutations causative of heritable disease may result in a milder phenotype. Small-scale genomic mosaicism is also possible, including single nucleotide substitutions. For example, somatic mosaicism for a mutation in the *COL4A5* gene (HGNC: 2207) is the cause of a milder phenotype of male Alport syndrome (Krol et al., 2008). Similarly, a somatic mutation partially rescuing a child with Hutchinson–Gilford progeria syndrome was recently reported (Bar et al., 2017).

Given the number of mitoses required for human development, it is plausible that every human has some cells harboring a mutation causative of genetic disease (Behjati et al., 2014; Gong, Gu, & Woodruff, 2005; Iourov et al., 2010; Seshadri, Kutlaca, Trainor, Matthews, & Morley, 1987). However, the level of mosaicism can be very low (i.e., with a postzygotic mutation visible in only a small subset of somatic cells) which will make mosaicism difficult to detect, especially using Sanger sequencing (Beicht et al., 2013). Next-Generation Sequencing (NGS) has facilitated faster sequencing with a lower per-base cost; however, restricted budgets still limit the maximum read depth of available data. While a medium read-depth (i.e., below 30X) could be sufficient for sensitivity and specificity regarding constitutional mutations, this might not be the case for detecting mosaicism.

Ye et al. reported on a sequencing project in DNA samples from peripheral blood of two MZ twin pairs of differing ages (40 and 100 years old) using two different sequencing and variant calling pipelines (Ye et al., 2013). One pipeline used Illumina 40X sequencing, alignment with Burrows–Wheeler Aligner (BWA) (H. Li & Durbin, 2009) and compared the four nucleotide base counts per genomic location between co-twins using the CaVEMAN pipeline developed by the Wellcome Trust Sanger Institute (Stephens et al., 2012). The second pipeline used Complete Genomics whole-genome sequencing, where between-twin variant calls were determined using the Complete Genomics tumor variant calling tool (Drmanac et al., 2010). Intersecting the large numbers of putative mosaic mutations for each platform resulted in 13 and 17 potential postzygotic mutations occurring in the 40-year-old and 100-year-old twin pairs, respectively. After validation with Sanger sequencing, Ye et al. found no somatic mutations in the 40-year-old twin pair, and eight validated somatic mutations in the older twin pair, consistent with the theory of aging as accumulation of somatic variants (Ye et al., 2013). However, this might be a conservative estimate of the true rate of mosaicism since mutations might have been missed by the sequence alignment software or by applying tumor-specific analysis software, or may have been detected by one approach and not the other, thereby leading to a very small intersection of variants that were detected by both platforms.

Here we use an alternative method for detecting postzygotic mutations in the same two twin pairs studied by Ye et al., analyzing measurements from two Illumina sources, at read depths 13X and 40X as opposed to Ye et al. (2013). We calculated for each locus in each person the allelic ratio: the fraction of alternative read counts as part of the total read depth at that locus. This is similar to deviation from genotype-expected b-allele frequency (B_{dev}), which is considered for detecting larger structural mosaic events such as copy number variations (King et al., 2017). With this method, having only reference reads generates a zero value, having 50% alternative reads will get value 0.5 (being heterozygous) and having only alternative reads will get value 1 (see Figure 1). However, any value between these values is also possible (e.g., having 30% alternative reads at a locus). This is different from

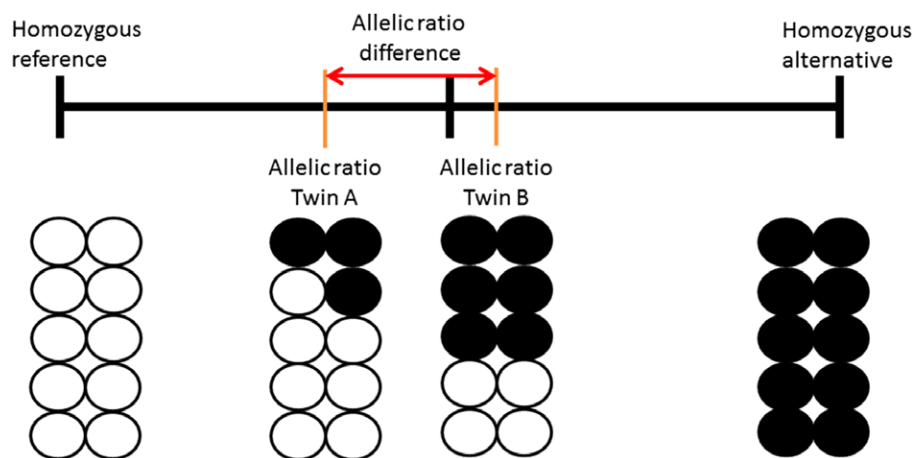


FIGURE 1 The result of a possible postzygotic mutation during early development in a monozygotic twin pair. Circles represent somatic cells, with cells containing a postzygotic mutation in black. A locus that was heterozygous before the twinning event may show different allele fractions between co-twins posttwinning. NGS variant calling software would generally call both co-twins heterozygous at this locus

standard variant calling, where loci are called either heterozygote or homozygote without any in-between values. For each locus, we compared this quantitative measure between the co-twins from the same pair and for the two sequences from the same individual. The larger the difference between the allelic ratios, the more the co-twins differ from each other at that locus, or the twin from herself between the 13X and 40X reads. Any difference between co-twins will represent a posttwinning event. In contrast, any differences within an individual are likely to represent noise. Mutations arising from parental germline mosaicism or during pretwinning stages will be present in both co-twins and therefore not detected by our method. Using this technique, we were able to identify multiple putative mosaic sites, which were then characterized in terms of position and function.

2 | MATERIALS AND METHODS

2.1 | Next-generation sequencing data measurements and analysis

One MZ twin pair aged 40 years was selected from the Netherlands Twin Register (Willemsen et al., 2013), and one MZ twin pair aged 100 years was selected from the Leiden Longevity Study (Westendorp et al., 2009). Both pairs were female. Data collection was approved by the Medical Ethics Committee of the Leiden University Medical Center and the Medical Ethics Committee of the VU Medical Centre, Amsterdam. Informed, written consent was obtained from the twins participating in the study. DNA was extracted from whole blood and sequenced using the Illumina HiSeq 2000 platform (91-basepair [bp] paired-end reads) at read depth 13X with the library preparation protocol developed for the Genome of the Netherlands project (Boomsma et al., 2014). For the same samples, 100-bp paired-end reads were generated using Illumina GAIIx instruments at read depth 40X using the manufacturer standard protocol and library as used in Ye et al. (2013). Including the Illumina 13 × 91 bp paired-end reads should increase the reliability of the read alignments, increasing the probability of placing a read with nonreference variants at the correct genomic location. Using a combination of two different sequencing platforms should reduce the influence of errors from PCR and sequencing.

Both Illumina-based NGS datasets were aligned to GRCh37 using *gaMap* 2.4.0 by Genalice. Variant calling was done with *gaVariant* 2.4.0. Both tools by Genalice were recently benchmarked against *PEMapper*/*PECaller*, *BWA*/*GATK* and *Isaac* (Pluss et al., 2017) showing comparable sensitivity (> 0.998) to *BWA*/*GATK* and outperforming *PEMapper*/*PECaller* and *Isaac*. The calling process included removal of PCR duplicates and removal of low-quality bases and reads, and local realignment to reduce false positive variant calls. Variant calling was optimized by applying a *softclip* to reads, and by filtering on quality- and coverage-induced noise levels (see <https://www.genalice.com/download-whitepapers/>). We restricted the analysis to high-confidence regions of the genome (Rosenbloom et al., 2015; Zook et al., 2014), a collection of regions identified by the Genome In A Bottle consortium (GIAB). By integrating and arbitrating between 14 data sets from five sequencing technologies, seven read

mappers and three variant callers, GIAB published regions in the genome where systematic sequencing errors, local alignment errors, and mapping errors have minimal influence while minimizing bias toward any individual sequencing platform. In practice, about 23% of the genome is discarded for displaying systematic sequencing errors or mapping difficulties (e.g., due to many simple repeats). In addition, we omitted repetitive regions of the genome using the University of California Santa Cruz (UCSC) RepeatMasker track (Karolchik et al., 2004). To ensure validity of our variant calling methods, we compared the resulting Variant Call Format file (VCF) with the results from a Burrows-Wheeler aligner/Genome Analysis Toolkit (*BWA*/*GATK*) pipeline following manufacturers' best practices (Van der Auwera et al., 2013). Using results from *BWA*/*GATK* as a baseline, we found a mean sensitivity, specificity, and F-ratio of 0.9858, 0.9811, and 0.9834 respectively (see Supporting Information Table S1). Further downstream analysis was done using a combination of custom shell, Perl, and R scripts. Loci where both co-twins were homozygous for the reference allele were discarded. These sites have, per definition, not one single reliably detected difference between twins and as a result do not contain information on possible postzygotic mutations. For each co-twins, the allelic ratio was calculated: the fraction of the alternative read count compared to the total read count. Subsequently, the difference in allelic ratio was calculated for each locus between co-twins. This resulted in a number between -1 and 1 , where 0 indicates no difference between co-twins and higher absolute numbers indicate larger differences between co-twins (as shown in Figure 1).

2.2 | Statistical analysis

Under the null hypothesis of absence of mosaic loci, it is still possible to find false positive putative mosaic sites at heterozygous loci due to random sampling. This is in addition to other sources of false positives including sequencing, PCR and mapping errors. Even though both co-twins may be heterozygous for a locus, there is still a chance that the ratio of alternative alleles deviates from 50%. To ascertain the false positive rate because of this sampling noise, we simulated two binomial distributions, with 40 observations (the read depth) and chance of success 0.5 and calculated the number of times a difference between co-twins was observed. We performed 1,000 simulations of 893,581 loci, the average number of heterozygous loci we analyzed per twin pair. We computed that due to random sampling we could expect 0.36% of these loci to have a difference in allelic ratio higher than 0.25 in both 40X and 13X data (see Figure 2 and Supporting Information Table S2).

We performed permutations to test for overabundance of postzygotic mutations within short distance to each other, within twin pairs and between twin pairs. For each permutation ($N = 1,000$), we extracted random loci from the reference set ($n = 1,720$ and $n = 1,739$ for the 40-year-old twin pair and the 100-year-old twin pair, respectively) and computed the number of loci with a distance smaller than the threshold, which we used as null distribution for the permutation test. Finally, loci were annotated using Variant Effect Predictor (VEP) (McLaren et al., 2016) and an enrichment test was performed using Fisher-exact tests. A gene-based enrichment analysis was performed with *DAVID* v6.8 (Huang da, Sherman, & Lempicki, 2009). For an

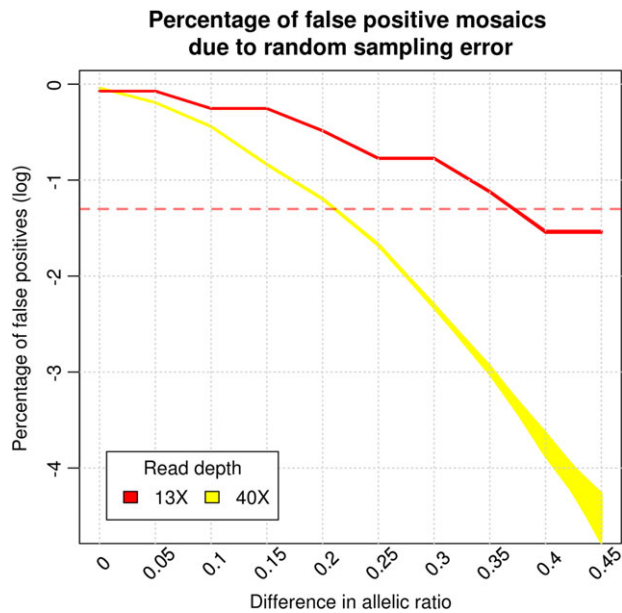


FIGURE 2 The chance of finding a postzygotic mutation because of sampling error in 40X and 13X data, based on 1,000 simulations of 893,581 heterozygote sites. The red dotted line indicates 5% false positives

additional validation of the enrichment analysis, we compared 13X and 40X sequencing data sets from each individual and tested whether loci with an allelic ratio difference were enriched for intronic regions. If differences within an individual truly represent noise in the data, we predict that we will not observe any enrichments.

3 | RESULTS

3.1 | Next-generation sequencing data

DNA sequencing was measured in an MZ twin pair aged 40 years and an MZ twin pair aged 100 years. DNA was extracted from whole blood and sequenced at read depth 13X and at read depth 40X. Variant calling was restricted to high-confidence regions of the genome, and repetitive regions were omitted (see Materials and Methods). After discarding loci where both co-twins were homozygous for the reference allele, and taking the intersection of loci found from both sequencing sets, 881,298 and 905,864 single-nucleotide loci were left for analysis for the 40-year-old and 100-year-old twin pair, respectively (see Supporting Information Table S3 for a flowchart of several filtering steps). We calculated for each locus in each person the allelic ratio: The fraction of alternative read counts as part of the total read depth at that locus and for each twin pair the allelic ratio difference at each site. This was done for both sequencing sets. Subsequently, we looked at the number of matching loci, here defined as sites where the allelic ratio difference has the same sign, according to both sequencing platforms. To increase the number of matching loci, we employed two filtering steps. We applied a threshold for minimum allelic ratio difference in both 40X and 13X data. We subsequently checked the percentage of matching loci at several thresholds and tested if the num-

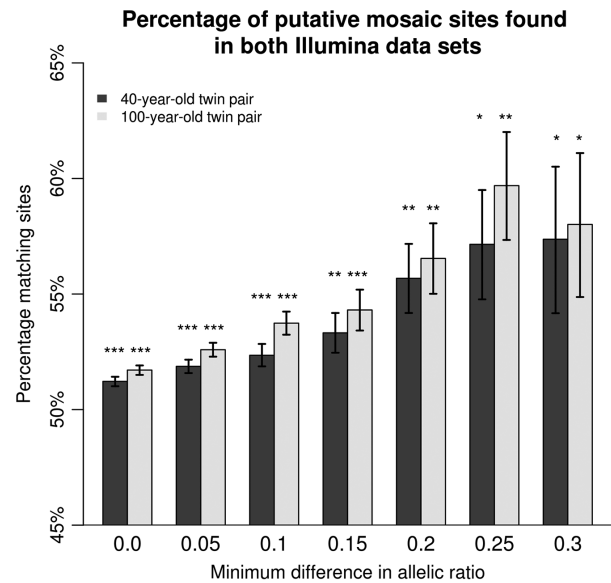


FIGURE 3 The fraction of putative mosaic sites found in both Illumina sets for both twin pairs. Asterisks indicate significance of binomial test of matching sites versus nonmatching sites. * $p < 1e-5$; ** $p < 1e-10$, *** $p < 1e-20$

ber of matching loci was significantly higher than the number of non-matching loci using a binomial test. Second, we found that conditioning on loci where one co-twin is clearly not mosaic (having an allelic ratio that differs less than 0.05 from 0, 0.5, or 1) also improves the percentage of matching loci. At minimum allelic ratio difference zero, this clear-call filter reduced the number of sites in our set from 881,298 to 226,945 (40-year-old twin pair) and from 905,864 to 225,010 (100-year-old twin pair). Increasing the threshold for allelic ratio differences up to 0.25 and using the clear-call filter increased the percentage and significance of matching loci for both twin pairs (see Figure 3). Therefore, we applied the clear-call filter and a threshold of 0.25 for both twin pairs. This resulted in 57.15% matching loci in the 40-year-old twin pair ($N = 1,720$, $p = 3.274e-9$) and 59.69% matching loci in the 100-year-old twin pair ($N = 1,739$, $p = 6.355e-16$). Out of these two sets, 19 loci were found to be simultaneously putatively mosaic in both twin pairs (see Supporting Information Table S3 for a detailed overview of all filtering steps).

Our choice for 0.25 as a threshold for minimal difference in the allelic ratio was supported by our simulations. In our data, the percentage of loci with an allelic ratio higher difference than 0.25 was 0.90% (40-year-old twins) and 0.89% (100-year-old twins; see Supporting Information Table S3). This difference was statistically significant ($p < 0.001$, 1,000 permutations) from what we would expect due to random sampling (0.36%). Sampling heterozygote loci influencing erroneous reporting of mosaics shows that setting a threshold for allelic ratio difference of > 0.25 led to a maximum of 2% false positive mosaics due to random sampling (see Figure 2). Loci identified in both the 13X and 40X data with the same direction of effect of allelic ratio difference are more likely to be true positives. Likewise, nonmatching loci, having opposite-signed effects, are more likely to be false positives. Thus, the difference between the percentage of matching loci (57.15% and 59.69%) and the percentage of nonmatching loci can be

considered as an estimate of the percentage of true positives (14.3% and 19.38%, respectively). Note that this estimate only holds for differences identifiable by both 13X and 40X platforms.

3.2 | Enrichment analyses

Within our respective sets of 1,720 and 1,739 putative mosaic mutations, we found enrichment for mosaics that are within 101–500, 501–1,000, 1,001–5,000, and 5,001–10,000 bp from each other (all nominally significant; see Supporting Information Table S4). Between the twin pairs, we also found that postzygotic mutations seem to cluster in hotspots with genetic distances up to 10,000 bp (10 pairs at 501–1,000 bp, 7 pairs at 501–1,000 bp, 40 pairs at 1,001–5,000 bp, 33 pairs at 5,001–10,000 bp, all $p < 0.001$; see Supporting Information Table S5). This enrichment gets less strong for larger genetic distances. We used VEP to annotate the results of putative mosaic mutations, and tested with a Fisher's exact test whether the 14 single-nucleotide polymorphism (SNP) categories provided by VEP were significantly enriched compared to our full list of heterozygous variants. From the 14 categories, nine were enriched in the 40-year-old twin pair and 10 were enriched in the 100-year-old twin pair. Remarkably, for both twin pairs the strongest enrichments were in the categories *regulatory elements* and *5' untranslated regions* ($p < 4.37e-91$, $p < 7.94e-33$, Table 1). Post hoc analysis showed that the significance of this enrichment increases with the

applied filter steps (Supporting Information Table S3). Additionally, we tested for enrichment of putative mosaic mutations in two genes, *DNMTA3* (HGNC:2978) and *TET2* (HGNC:25941), both linked to clonal expansion of hematopoietic stem cells. Somatic mosaicism in these genes was reported to be common in the elderly (Acuna-Hidalgo et al., 2017; van den Akker et al., 2016; Zink et al., 2017). We found a slight enrichment for low-mutational ratio mosaicism in these genes in both twin pairs (minimal difference in allelic ratio threshold > 0.05 , 40-year-old twin pair $p = 0.0023$, 100-year-old twin pair $p = 0.0031$; Supporting Information Table S3). For an additional validation of the enrichment analysis, we compared two sequencing runs of the same individual using the same procedure as was used for the co-twins. We identified loci with allelic ratio differences in these sequence runs and tested whether these loci were enriched for intronic regions. The only significant enrichment was found when comparing the 40X data between co-twins, but not when comparing different sequencing runs of the same person. This indicates that the identified putative postzygotic mutations may contain a high number of false positives, but these false positives are unlikely to drive the reported enrichment (see Supporting Information Tables S6 and S7).

Using DAVID, we tested whether genes where a mosaic mutation was found ($N = 1023$ for the 40-year-old twins, $N = 1,055$ for the 100-year-old-twins) were enriched in functionally related gene groups. After false discovery rate (FDR) correction, nine and 16 gene groups were found to be significantly enriched in the 40-year-old and

TABLE 1 Enrichment of variant types in the set of putative postzygotic mutations

	40-year-old twin pair ARD > 0	40-year-old twin pair ARD > 0.25	Enrichment <i>p</i> -value	Enrichment <i>p</i> -value, FDR-corrected	100-year-old twin pair ARD > 0	100-year-old twin pair ARD > 0.25	Enrichment <i>p</i> -value	Enrichment <i>p</i> -value, FDR-corrected
Total variants	226,945	1,720			225,010	1,739		
Intronic	55.4%	59.7%	<i>9.46e-05</i>	1.66e-04	55.2%	59.6%	<i>5.39e-05</i>	1.26e-04
Intergenic	35.1%	24.5%	1	1	35.4%	24.6%	1	1
Modifier	99.7%	99.8%	0.338	0.364	99.8%	99.7%	0.840	0.905
Low impact	0.9%	2.6%	<i>1.79e-09</i>	5.01e-09	0.9%	2.0%	<i>1.80e-05</i>	<i>5.04e-05</i>
Moderate impact	0.6%	1.5%	<i>1.57e-05</i>	3.14e-05	0.6%	1.3%	<i>5.15e-04</i>	<i>7.99e-04</i>
High impact	0.03%	0.1%	0.0978	0.114	0.03%	0.1%	0.0915	0.107
Noncoding	32.4%	35.6%	<i>1.52e-03</i>	2.13e-03	32.1%	35.8%	<i>2.80e-04</i>	<i>5.60e-04</i>
Synonymous	0.7%	1.5%	<i>2.06e-04</i>	3.20e-04	0.7%	1.4%	<i>5.71e-04</i>	<i>7.99e-04</i>
Missense	0.6%	1.5%	<i>1.57e-05</i>	3.14e-05	0.6%	1.3%	<i>5.15e-04</i>	<i>7.99e-04</i>
Regulatory	9.1%	25.7%	<i>4.37e-91</i>	6.12e-91	9.1%	25.8%	<i>4.66e-93</i>	<i>6.52e-92</i>
TF binding	0.7%	2.9%	<i>8.10e-18</i>	3.78e-17	0.7%	2.7%	<i>7.68e-15</i>	<i>2.69e-14</i>
Protein coding	49.2%	57.6%	<i>4.52e-13</i>	1.582e-12	49.4%	59.1%	<i>1.42e-16</i>	<i>6.63e-16</i>
3' UTR	1.7%	2.4%	0.0118	0.015	1.7%	2.8%	<i>6.75e-04</i>	<i>8.59e-04</i>
5' UTR	0.4%	4.8%	<i>1.15e-57</i>	8.05e-57	0.4%	3.3%	<i>7.94e-33</i>	<i>5.56e-32</i>
1 KG median	31.6	32.7			31.3%	33.9		
1 KG mean	34.8	35.6			34.6%	36.1		
Freq < 0.1	16.0%	13.4%			16.6%	14.4%		
Percentage in 1000G	96.6%	82.6%			96.7%	82.5%		

Notes. ARD = allelic ratio difference; FDR = false discovery rate. For an explanation of the different annotation terms, see https://www.ensembl.org/info/genome/variation/predicted_data.html#consequences. *p*-values in italics are significant after correction for multiple testing.

TABLE 2 Significant results from functional annotation using DAVID

Database	Term	40-year-old twin pair		100-year-old twin pair	
		<i>p</i> -value	FDR <i>p</i> -value	<i>p</i> -value	FDR <i>p</i> -value
UP	Alternative splicing	7.47e-18	1.05e-14	2.43e-22	3.45e-19
UP	Splice variant	1.09e-11	1.92e-08	7.73e-19	1.39e-15
UP	Polymorphism	3.55e-07	4.992647e-04	1.19e-07	1.69e-04
KEGG	Inflammatory mediator regulation of TRP channels	1.11e-06	1.45e-04	n.s.	n.s.
GO	Intracellular signal transduction	1.53e-06	2.75e-03	n.s.	n.s.
GO	Positive regulation of synapse assembly	7.95e-06	0.0143	n.s.	n.s.
GO	Positive regulation of GTPase activity	1.01e-05	0.0182	n.s.	n.s.
UP	Cell junction	1.50e-05	0.0211	8.96e-07	1.27e-03
UP	Sequence variant	1.36e-05	0.0241	1.19e-05	0.0214
UP	Synapse	n.s.	n.s.	1.74e-06	2.47e-03
UP	Ion channel	n.s.	n.s.	2.00e-06	2.84e-03
UP	Epidermal growth factor-like domain	n.s.	n.s.	5.35e-06	8.92e-03
INTERPRO	Axon guidance	n.s.	n.s.	5.99e-06	0.011
GO	EGF-like domain	n.s.	n.s.	1.26e-05	0.0178
UP	Ig-like C2-type 5	n.s.	n.s.	9.95e-06	0.0179
UP	GTPase activator activity	n.s.	n.s.	1.40e-05	0.0220
GO	Membrane	n.s.	n.s.	1.73e-05	0.0245
UP	Metal-binding	n.s.	n.s.	1.79e-05	0.0254
UP	Metal ion binding	n.s.	n.s.	2.45e-05	0.0386
GO	EGF-like 2	n.s.	n.s.	2.26e-05	0.0405

Notes. FDR = false discovery rate; GO = Gene Ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes; UP = UniProt.

100-year-old twin pairs respectively (see Table 2 and Supporting Information Tables S8 and S9). Of these gene groups, five overlapped: alternative splicing, splice variant, polymorphism, cell junction, sequence variant, and GTPase activity. The latter was annotated for the 40-year-old twin pair as GO-term *Positive regulation of GTPase activity*, while it was annotated for the 100-year-old twin pair as UniProt-term *GTPase activator activity*. GTPase activating proteins are essential modulators of the biological activity of guanine nucleotide binding proteins (G-proteins). G-protein-coupled receptors are crucial players in tumor growth and metastasis (Dorsam & Gutkind, 2007).

4 | DISCUSSION

By using the difference in alternative allele ratio in MZ twin pairs as a measure for mosaicism, we identified and annotated a set of 1,720 and 1,739 loci containing putative mosaic mutations in a 40-year-old MZ twin pair and a 100-year-old MZ twin pair, respectively. The number of putative mutations identified in blood-derived DNA, although with a high false positive rate, is high compared to an earlier approximation of postzygotic mutations, which found that each individual carries about 300 postzygotic mutations also using blood-derived DNA (R. Li et al., 2014). However, since sequencing was done with two separate libraries (40X and 13X) and we limited ourselves to high-confidence regions of the genome, we were able to identify preferential enrichment of postzygotic mutations. The differences between twins we identified point to clustering of postzygotic mutations around

hotspots up to 10 Mb in size. We found similar patterns of strong enrichments for variant types and gene sets in both twin pairs, suggesting that postzygotic mutations follow a nonrandom pattern, confirming recent findings (Vattathil & Scheet, 2016). The enrichment of regulatory elements suggests a relevant role for mosaicism. We replicated the finding by Lodato (2015) that coding exons are enriched for postzygotic mutations (Lodato et al., 2015). The enrichment we found for mosaicism in genes associated with GTPase activity, involved in tumor growth, is in line with recent findings that larger-scale genomic mosaicism in genes are associated with cancer (Laurie et al., 2012; Machiela et al., 2015; Vattathil & Scheet, 2016). The continuing discovery of even more cases of mosaicism provides much-needed insights into postzygotic mutational signatures (see, e.g., Ju et al. 2017; Martincorena and Campbell, 2016). The current research cannot speak to the role of postzygotic mutations in disease pathogenesis, as it was based on results from healthy MZ twins. The literature suggests that notion of one person not having the same genome in every cell throughout the body is of important clinical relevance as mosaicism is involved in diseases that would be lethal in constitutional state, as well as in organ-specific diseases since mosaicism can remain site-specific (Biesecker & Spinner, 2013).{Biesecker, 2013 #45} In a 115-year-old woman, somatic mutations detected in blood were not detected in other tissues or tumor tissue, indicating that these mutations were not derived from tumor cells (Holstege et al., 2014). Mosaicism has been observed in several Mendelian diseases, for example, as previously mentioned in Alport Syndrome, where somatic mosaicism resulted in an unusually mild phenotype (Bruttini et al., 2000; Krol et al., 2008).

However, the level of mosaicism does not necessarily correlate with the severity of clinical manifestation, and mosaicism may even not have any visible effects (Cohen et al., 2015). Future research should take advantage of new technologies, for example, single cell sequencing, for high-resolution detection and localization of genetic mosaicism.

In spite of the high number of false positives in the identified mutations, these false positives are unlikely to be preferentially enriched in functional elements or gene pathways, as was confirmed by additional analysis comparing 13X and 40X measurements from the same samples. We are aware that by selecting mutations that are present both in 40X and 13X data we may discard some true positive data present in 40X but not in 13X data. However, this removes more false positives than true positives. This is exemplified with the enrichment analyses, which show stronger effects after more stringent filtering. In three of the eight loci previously validated through Sanger sequencing (Ye et al., 2013), we found evidence in our data for mosaicism. Out of the 22 false-positive mutations in Ye et al. that were shown to be false positives by Sanger sequencing, we identified 21 true-negative and one false-positive indication of mosaicism (see Supporting Information Table S10).

Our stringent filtering might also explain why we do not see age differences in the number of mosaic mutations. Age-related somatic mosaicism results in very low mutational ratios (Laurie et al., 2012), whereas higher mutational ratios should occur only early after fertilization or by clonal expansion. This suggests that the mutations we extracted in our procedure likely occurred early in life, instead of being accumulated with age. It is important to note that postzygotic mutations resulting in differences in allelic ratio between MZ twins cannot be a result of parental germline mosaicism, but rather occur during embryogenesis. Parental mosaicism would result in mutations in both co-twins, which would not result in a difference in allelic ratio between twins and would therefore not be detectable by our method.

Although Huang et al. mention that accurate identification of postzygotic mutations provides insights into finding the “missing heritability” (Huang et al., 2014), the number of expected postzygotic mutations is so low that the impact on heritability estimates in twin studies is negligible. Twin-based heritability is based on the assumption that the additive genetic correlations of MZ and dizygotic (DZ) twins equal $r_{MZ} = 1$ and $r_{DZ} = 0.5$, respectively. If a model with $r_{MZ} = 1$ and $r_{DZ} = 0.5$ is fitted to twin data, while both r_{MZ} and r_{DZ} are marginally lower due to mosaicism, the estimate of the additive genetic variance will be only marginally biased. For instance: let mosaicism cause the r_{MZ} to be 0.99 and r_{DZ} to be 0.495, and let the true additive genetic variance be 0.35. The misspecified model (assuming $r_{MZ} = 1$ and $r_{DZ} = 0.5$) would then yield an estimate of 0.3465. The true rate of mosaicism, and therefore its contribution to the “missing heritability,” is many magnitudes lower. While the effects on the heritability are expected to be negligible, this does not rule out that nonheritable genetic variation may be an important factor in the development of sporadic diseases (Forsberg et al., 2013).

In this paper, we showed the value of the monozygotic twin design as a method to identify mosaicism. Even with a limited sample size, we established with this design that mosaic mutations are not randomly

distributed across the genome but rather are highly enriched for specific genomic hotspot locations, transcript location, and gene groups.

ACKNOWLEDGMENTS

We thank prof. Dr. Conor Dolan for input in the discussion section. This study makes use of data generated in the Leiden Longevity Study with Prof. P. E. Slagboom as principle investigator and of data generated in the Netherlands Twin Register with prof. D. I. Boomsma as principle investigator.

Klaasjan Ouwens is financially supported by the EMGO Institute for Health and Care Research (EMGO+). We acknowledge BBRMI-NL (NWO 184.021.007).

CONFLICTS OF INTERESTS

Bas Tolhuis is a paid employee of Genallice Core BV. Klaasjan Ouwens is an embedded Ph.D. candidate at Genallice Core BV.

ORCID

Klaasjan G. Ouwens  <http://orcid.org/0000-0002-3864-7710>

REFERENCES

- Abyzov, A., Tomasini, L., Zhou, B., Vasmatzis, N., Coppola, G., Amenduni, M., ... Vaccarino, F. M. (2017). One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin. *Genome Research*, 27(4), 512–523. <https://doi.org/10.1101/gr.215517.116>
- Acuna-Hidalgo, R., Bo, T., Kwint, M. P., van de Vorst, M., Pinelli, M., Veltman, J. A., ... Gilissen, C. (2015). Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. *American Journal of Human Genetics*, 97(1), 67–74. <https://doi.org/10.1016/j.ajhg.2015.05.008>
- Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S. H., Kiemeny, L., ... Hoischen, A. (2017). Ultra-sensitive sequencing identifies high prevalence of clonal hematopoiesis-associated mutations throughout adult life. *American Journal of Human Genetics*, 101(1), 50–64. <https://doi.org/10.1016/j.ajhg.2017.05.013>
- Bar, D. Z., Arlt, M. F., Brazier, J. F., Norris, W. E., Campbell, S. E., Chines, P., ... Gordon, L. B. (2017). A novel somatic mutation achieves partial rescue in a child with Hutchinson-Gilford progeria syndrome. *Journal of Medical Genetics*, 54(3), 212–216. <https://doi.org/10.1136/jmgenet-2016-104295>
- Behjati, S., Tarpey, P. S., Sheldon, H., Martincorena, I., Van Loo, P., Gundem, G., ... Campbell, P. J. (2014). Recurrent PTPRB and PLCG1 mutations in angiosarcoma. *Nature Genetics*, 46(4), 376–379. <https://doi.org/10.1038/ng.2921>
- Beicht, S., Strobl-Wildemann, G., Rath, S., Wachter, O., Alberer, M., Kaminsky, E., ... Hoefele, J. (2013). Next generation sequencing as a useful tool in the diagnostics of mosaicism in Alport syndrome. *Gene*, 526(2), 474–477. <https://doi.org/10.1016/j.gene.2013.05.045>
- Biesecker, L. G., & Spinner, N. B. (2013). A genomic view of mosaicism and human disease. *Nature Reviews Genetics*, 14(5), 307–320. <https://doi.org/10.1038/nrg3424>
- Boomsma, D. I., Wijmenga, C., Slagboom, E. P., Swertz, M. A., Karssen, L. C., Abdellaoui, A., ... van Duijn, C. M. (2014). The Genome of the Netherlands: Design, and project goals. *European Journal of Human Genetics*, 22(2), 221–227. <https://doi.org/10.1038/ejhg.2013.118>

- Bruttini, M., Vitelli, F., Meloni, I., Rizzari, G., Volpe, M. D., Mazzucco, G., ... Renieri, A. (2000). Mosaicism in Alport syndrome with genetic counselling. *Journal of Medical Genetics* 37(9), 717–719.
- Buntinx, F., Campbell, C., & van den Akker, M. (2014). Cancer in the elderly. *Journal of Cancer Epidemiology*, 2014, 872029. <https://doi.org/10.1155/2014/872029>
- Cohen, A. S., Wilson, S. L., Trinh, J., & Ye, X. C. (2015). Detecting somatic mosaicism: Considerations and clinical implications. *Clinical Genetics*, 87(6), 554–562. <https://doi.org/10.1111/cge.12502>
- Dal, G. M., Erguner, B., Sagiroglu, M. S., Yuksel, B., Onat, O. E., Alkan, C., & Ozcelik, T. (2014). Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. *Journal of Medical Genetics* 51(7), 455–459. <https://doi.org/10.1136/jmedgenet-2013-102197>
- Dorsam, R. T., & Gutkind, J. S. (2007). G-protein-coupled receptors and cancer. *Nature Reviews Cancer*, 7(2), 79–94. <https://doi.org/10.1038/nrc2069>
- Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., ... Reid, C. A. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, 327(5961), 78–81. <https://doi.org/10.1126/science.1181498>
- Forsberg, L. A., Absher, D., & Dumanski, J. P. (2013). Non-heritable genetics of human disease: Spotlight on post-zygotic genetic variation acquired during lifetime. *Journal of Medical Genetics* 50(1), 1–10. <https://doi.org/10.1136/jmedgenet-2012-101322>
- Forsberg, L. A., Gisselsson, D., & Dumanski, J. P. (2017). Mosaicism in health and disease - clones picking up speed. *Nature Reviews Genetics*, 18(2), 128–142. <https://doi.org/10.1038/nrg.2016.145>
- Gong, Y., Gu, S., & Woodruff, R. C. (2005). The influence of premeiotic clusters of mutation on indirect estimations of mutation rate. *Human Heredity*, 60(3), 150–155. <https://doi.org/10.1159/000089955>
- Holstege, H., Pfeiffer, W., Sie, D., Hulsman, M., Nicholas, T. J., Lee, C. C., ... Sistermans, E. A. (2014). Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis. *Genome Research*, 24(5), 733–742. <https://doi.org/10.1101/gr.162131.113>
- Huang, A. Y., Xu, X., Ye, A. Y., Wu, Q., Yan, L., Zhao, B., ... Wei, L. (2014). Postzygotic single-nucleotide mosaicism in whole-genome sequences of clinically unremarkable individuals. *Cell Research*, 24(11), 1311–1327. <https://doi.org/10.1038/cr.2014.131>
- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>
- Iourov, I. Y., Vorsanova, S. G., & Yurov, Y. B. (2010). Somatic genome variations in health and disease. *Current Genomics*, 11(6), 387–396. <https://doi.org/10.2174/138920210793176065>
- Jacobs, K. B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., ... Chanock, S. J. (2012). Detectable clonal mosaicism and its relationship to aging and cancer. *Nature Genetics*, 44(6), 651–658. <https://doi.org/10.1038/ng.2270>
- Ju, Y. S., Martincorena, I., Gerstung, M., Petljak, M., Alexandrov, L. B., Rahbari, R., ... Stratton, M. R. (2017). Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature*, 543(7647), 714–718. <https://doi.org/10.1038/nature21703>
- Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, 32(Database issue), D493–496. <https://doi.org/10.1093/nar/gkh103>
- King, D. A., Sifrim, A., Fitzgerald, T. W., Rahbari, R., Hobson, E., Homfray, T., ... Deciphering Developmental Disorders, S. (2017). Detection of structural mosaicism from targeted and whole-genome sequencing data. *Genome Research*, 27(10), 1704–1714. <https://doi.org/10.1101/gr.212373.116>
- Kirkwood, T. B. (2005). Understanding the odd science of aging. *Cell*, 120(4), 437–447. <https://doi.org/10.1016/j.cell.2005.01.027>
- Krol, R. P., Nozu, K., Nakanishi, K., Iijima, K., Takeshima, Y., Fu, X. J., ... Yoshikawa, N. (2008). Somatic mosaicism for a mutation of the COL4A5 gene is a cause of mild phenotype male Alport syndrome. *Nephrology Dialysis Transplantation*, 23(8), 2525–2530. <https://doi.org/10.1093/ndt/gfn005>
- Laurie, C. C., Laurie, C. A., Rice, K., Doheny, K. F., Zelnick, L. R., McHugh, C. P., ... Weir, B. S. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature Genetics*, 44(6), 642–U658. <https://doi.org/10.1038/ng.2271>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, R., Montpetit, A., Rousseau, M., Wu, S. Y., Greenwood, C. M., Spector, T. D., ... Richards, J. B. (2014). Somatic point mutations occurring early in development: A monozygotic twin study. *Journal of Medical Genetics* 51(1), 28–34. <https://doi.org/10.1136/jmedgenet-2013-101712>
- Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., ... Walsh, C. A. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science*, 350(6256), 94–98. <https://doi.org/10.1126/science.aab1785>
- Machiela, M. J., Zhou, W. Y., Sampson, J. N., Dean, M. C., Jacobs, K. B., Black, A., ... Chanock, S. J. (2015). Characterization of Large Structural Genetic Mosaicism in Human Autosomes. *American Journal of Human Genetics*, 96(3), 487–497. <https://doi.org/10.1016/j.ajhg.2015.01.011>
- Machin, G. A. (1996). Some causes of genotypic and phenotypic discordance in monozygotic twin pairs. *American Journal of Medical Genetics*, 61(3), 216–228. [https://doi.org/10.1002/\(SICI\)1096-8628\(19960122\)61:3<216::AID-AJMG5>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-8628(19960122)61:3<216::AID-AJMG5>3.0.CO;2-S)
- Martin, N., Boomsma, D., & Machin, G. (1997). A twin-pronged attack on complex traits. *Nature Genetics*, 17(4), 387–392. <https://doi.org/10.1038/ng1297-387>
- Martincorena, I., & Campbell, P. J. (2016). Somatic mutation in cancer and normal cells (vol 351, aaf5401, 2016). *Science*, 353(6295), 132–132.
- McConnell, M. J., Moran, J. V., Abyzov, A., Akbarian, S., Bae, T., Cortes-Ciriano, I., ... Brain Somatic Mosaicism, N. (2017). Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science*, 356(6336). <https://doi.org/10.1126/science.aal1641>
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... Cunningham, F. (2016). The Ensembl variant effect predictor. *Genome Biology*, 17(1), 122. <https://doi.org/10.1186/s13059-016-0974-4>
- Pluss, M., Kopps, A. M., Keller, I., Meienberg, J., Caspar, S. M., Dubacher, N., ... Matyas, G. (2017). Need for speed in accurate whole-genome data analysis: GENALICE MAP challenges BWA/GATK more than PEMapper/PECall and Isaac. *Proceedings of National Academy of Sciences United States of America*, <https://doi.org/10.1073/pnas.1713830114>
- Rahbari, R., Wuster, A., Lindsay, S. J., Hardwick, R. J., Alexandrov, L. B., Turki, S. A., ... Hurles, M. E. (2016). Timing, rates and spectra of human germline mutation. *Nature Genetics*, 48(2), 126–133. <https://doi.org/10.1038/ng.3469>
- Reindollar, R. H., Byrd, J. R., Hahn, D. H., Haseltine, F. P., & McDonough, P. G. (1987). A cytogenetic and endocrinologic study of a set of monozygotic isokaryotic 45,X/46,Xy twins discordant for phenotypic sex - mosaicism versus chimerism. *Fertility and Sterility*, 47(4), 626–633.
- Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., ... Kent, W. J. (2015). The UCSC Genome Browser

- database: 2015 update. *Nucleic Acids Research*, 43(D1), D670–D681. <https://doi.org/10.1093/nar/gku1177>
- Seshadri, R., Kutlaca, R. J., Trainor, K., Matthews, C., & Morley, A. A. (1987). Mutation-rate of normal and malignant human-lymphocytes. *Cancer Research*, 47(2), 407–409.
- Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., ... Osbreac (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, 486(7403), 400–404. <https://doi.org/10.1038/nature11017>
- van den Akker, E. B., Pitts, S. J., Deelen, J., Moed, M. H., Potluri, S., van Rooij, J., ... Consortium, G. N. (2016). Uncompromised 10-year survival of oldest old carrying somatic mutations in DNMT3A and TET2. *Blood*, 127(11), 1512–1515. <https://doi.org/10.1182/blood-2015-12-685925>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43, 11 10, 11–33. <https://doi.org/10.1002/0471250953.bi1110s43>
- vanDijk, B. A., Boomsma, D. I., & deMan, A. J. M. (1996). Blood group chimerism in human multiple births is not rare. *American Journal of Medical Genetics*, 61(3), 264–268. [https://doi.org/10.1002/\(Sici\)1096-8628\(19960122\)61:3<264::Aid-Ajmg11>3.0.Co;2-R](https://doi.org/10.1002/(Sici)1096-8628(19960122)61:3<264::Aid-Ajmg11>3.0.Co;2-R)
- Vattathil, S., & Scheet, P. (2016). Extensive hidden genomic mosaicism revealed in normal tissue. *American Journal of Human Genetics*, 98(3), 571–578. <https://doi.org/10.1016/j.ajhg.2016.02.003>
- Veitia, R. A., Govindaraju, D. R., Bottani, S., & Birchler, J. A. (2017). Aging: Somatic mutations, epigenetic drift and gene dosage imbalance. *Trends in Cell Biology*, 27(4), 299–310. <https://doi.org/10.1016/j.tcb.2016.11.006>
- Westendorp, R. G. J., van Heemst, D., Rozing, M. P., Frolich, M., Mooijaart, S. P., Blauw, G. J., ... Grp, L. L. S. (2009). Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *Journal of the American Geriatrics Society*, 57(9), 1634–1637. <https://doi.org/10.1111/j.1532-5415.2009.02381.x>
- Willemsen, G., Vink, J. M., Abdellaoui, A., den Braber, A., van Beek, J. H. D. A., Draisma, H. H. M., ... Boomsma, D. I. (2013). The Adult Netherlands Twin Register: Twenty-Five Years of Survey and Biological Data Collection. *Twin Research and Human Genetics*, 16(1), 271–281. <https://doi.org/10.1017/thg.2012.140>
- Ye, K., Beekman, M., Lameijer, E. W., Zhang, Y., Moed, M. H., van den Akker, E. B., ... Slagboom, P. E. (2013). Aging as accelerated accumulation of somatic variants: Whole-genome sequencing of centenarian and middle-aged monozygotic twin pairs. *Twin Research and Human Genetics*, 16(6), 1026–1032. <https://doi.org/10.1017/thg.2013.73>
- Zink, F., Stacey, S. N., Norddahl, G. L., Frigge, M. L., Magnusson, O. T., Jonsdottir, I., ... Stefansson, K. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*, 130(6), 742–752. <https://doi.org/10.1182/blood-2017-02-769869>
- Zook, J. M., Chapman, B., Wang, J., Mittelman, D., Hofmann, O., Hide, W., & Salit, M. (2014). Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nature Biotechnology*, 32(3), 246–251. <https://doi.org/10.1038/nbt.2835>
- Zwijnenburg, P. J. G., Meijers-Heijboer, H., & Boomsma, D. I. (2010). Identical But Not the Same: The Value of Discordant Monozygotic Twins in Genetic Research. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 153(6), 1134–1149. <https://doi.org/10.1002/ajmg.b.31091>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ouwens KG, Jansen R, Tolhuis B, Slagboom PE, Penninx BWJH, Boomsma DI. A characterization of postzygotic mutations identified in monozygotic twins. *Hum Mutat.* 2018;39:1393–1401. <https://doi.org/10.1002/humu.23586>