ORIGINAL ARTICLE

Somatic point mutations occurring early in development: a monozygotic twin study

Rui Li,[1,2] Alexandre Montpetit,[3] Marylène Rousseau,[4] Si Yu Margaret Wu,[4] Celia M T Greenwood,[2,5,6] Timothy D Spector,[7] Michael Pollak,[8] Constantin Polychronakos,[4] J Brent Richards[1,2,7]

[1]Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada
[2]Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada
[3]McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada
[4]Departments of Pediatrics and Human Genetics, McGill University, Montreal, Quebec, Canada
[5]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[6]Department of Oncology, McGill University, Montreal, Quebec, Canada
[7]Twin Research and Genetic Epidemiology, King's College London, London, UK
[8]Departments of Medicine and Oncology, McGill University, Montreal, Quebec, Canada

Correspondence to
Dr Brent Richards, Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, mcGill university, Montreal, Quebec, Canada H3A 2T5; brent.richards@mcgill.ca

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2013-101712).

ABSTRACT

The identification of somatic driver mutations in cancer has enabled therapeutic advances by identifying drug targets critical to disease causation. However, such genomic discoveries in oncology have not translated into advances for non-cancerous disease since point mutations in a single cell would be unlikely to cause non-malignant disease. An exception to this would occur if the mutation happened early enough in development to be present in a large percentage of a tissue's cellular population. We sought to identify the existence of somatic mutations occurring early in human development by ascertaining base-pair mutations present in one of a pair of monozygotic twins, but absent from the other and assessing evidence for mosaicism. To do so, we genome-wide genotyped 66 apparently healthy monozygotic adult twins at 506 786 high-quality single nucleotide polymorphisms (SNPs) in white blood cells. Discrepant SNPs were verified by Sanger sequencing and a selected subset was tested for mosaicism by targeted high-depth next-generation sequencing (20 000-fold coverage) as a surrogate marker of timing of the mutation. Two de novo somatic mutations were unequivocally confirmed to be present in white blood cells, resulting in a frequency of 1.2×10^{-7} mutations per nucleotide. There was little evidence of mosaicism on high-depth next-generation sequencing, suggesting that these mutations occurred early in embryonic development. These findings provide direct evidence that early somatic point mutations do occur and can lead to differences in genomes between otherwise identical twins, suggesting a considerable burden of somatic mutations among the trillions of mitoses that occur over the human lifespan.

INTRODUCTION

Mutation is an important source of genetic variation in the human genome. It can introduce deleterious nucleotide changes to genes or provide fuel for phenotypic evolution. It is well accepted that such mutations may lead to cancer,[1] but it is unlikely that *all* base-pair mutations would cause malignancy. Rather, such stochastic events could also lead to possible disruption of an organ's function, particularly so if the mutation were present in a large proportion of the cells within a tissue. Such mutations could exist if they were introduced early in embryogenesis and their identification could identify important control points in disease aetiology, such as has happened for malignant disease.

Indeed, somatic mutations have previously been demonstrated to cause non-malignant disease. Rapid advances in molecular genetics have demonstrated the importance of somatic mutation in a great variety of human diseases other than cancer (reviewed elsewhere).[2] For instance, a recent proof-of-concept study has revealed the existence of early embryonic somatic mutations causing Dravet syndrome,[3] as well as another novel finding of mosaic *AKT1* mutation in a Proteus syndrome patient.[4] A more recent publication identified somatic mutations in individuals with clonal haematopoiesis but without haematopoietic malignancies.[5]

In addition, the identification of early somatic mutations would provide preliminary insights into somatic mutation rates, which have been previously estimated in in vitro cell models[6–10] or disease–gene data.[11 12] Recent advances in sequencing technology have provided such rates in tumour samples,[13–16] and genotyping arrays have led to the detection of copy number variation in population studies.[17–19] However, the burden of point mutations in non-malignant tissue is not known, and further we are not aware of previous reports identifying the existence of somatic point mutations in otherwise apparently healthy individuals.

Monozygotic twins provide a natural experiment to address these questions since any differences between monozygotic co-twins would arise due to somatic changes. In the current study, we genome-wide genotyped 66 monozygotic twins (33 pairs) to identify somatic point mutations. In addition, we sought validation of the candidate somatic mutations through Sanger sequencing. Given that mutations occurring later in organ development would lead to somatic mosaicism within a cell line, we assessed the degree of mosaicism of identified mutations using next-generation sequencing. Together, these data provide direct evidence of the existence of somatic mutations occurring early in human development and a preliminary estimate of the rate at which they occur in an otherwise normal population.

RESULTS

Data generation

By genome-wide genotyping with Illumina 610K single nucleotide polymorphism (SNP) arrays, in the same laboratory at the same time, we obtained genotype calls on 506 821 SNPs in 33 pairs of monozygotic twins (66 individuals) after quality

control (figure 1). We only included high-quality (per SNP missing rate <0.01, Hardy–Weinberg Equilibrium p>10^{-6}) common SNPs (minor allele frequency >0.05). The cut-off of genotyping call rate per sample was 5%.

The call rate in the 66 individuals included in this study was >99.9%. Zygosity of the twins was determined by standardised questionnaire and confirmed by genotype concordance on a genome-wide level (>99.9%) in the 33 pairs of twins. To identify potential somatic mutations, we first assessed the genome-wide genotypes and identified 1087 discordant SNPs across all twin pairs. Of these, 1019 were present once across all twin pairs, 64 were present in two pairs of twins and 4 SNPs were discordant in three pairs of twins. Excluding these four SNPs as a potential source of genotyping error (since some SNPs are more difficult to genotype than others), we observed a median of 10 discordant SNPs per pair of monozygotic twins ranging from 1 to 282 (mean=34.76).

Raw data inspection to identify candidate somatic mutations for external verification

Despite precautions to select only the highest quality SNPs, the majority of the discordant SNPs were likely to be due to genotyping errors. To resolve this possible source of error, we visually inspected the cluster plots of the 1087 discordant SNPs between 66 monozygotic twins (33 pairs). Examination of the data (cluster separation, tightness, intensity and outsider genotype) suggested that 93% of these discordances most probably represented base-calling errors. Next we selected only SNPs that were called as being in the centre of a fluorescent cluster on the genotyping cluster plots, so as to preferentially identify mutations that were not mosaic, since mosaic mutations would more likely be between two clusters rather than at the centre of a cluster. We identified 80 SNPs as potentially early somatic mutations and carried out validation experiments on all these SNPs.

Verification by Sanger sequencing

We next amplified the region spanning each candidate somatic mutation from each pair of genotype discordant monozygotic twins and sequenced them by conventional Sanger sequencing for 80 discordant SNPs in 33 monozygotic twins from initial data set. Only 2 out of 80 (2.5%) candidate somatic mutations were confirmed from two distinct pairs, with one twin being homozygous and the other twin being heterozygous (figure 2). This low verification rate suggested that few additional mutations would be identified if we relaxed the quality control criteria in the analysis on genome-wide genotyping data.

Mutation validation and somatic mosaicism exploration by next-generation sequencing

Somatic mutations occurring at a late stage in the exponential cellular expansion phase of development would be only present in a small proportion of cells and would result in somatic mosaicism.[20][21] It has been suggested that Sanger sequencing has limited sensitivity to detect such mosaic somatic mutations,[22–24] and therefore, in order to interrogate mutations for mosaicism, we further applied high-depth next-generation sequencing as a complementary approach for 18 SNPs that showed evidence of possible somatic mosaicism in Sanger sequencing and the two somatic mutations verified by Sanger sequencing as positive controls. Equal amounts of PCR–amplicon DNA from 20 different genomic regions in 20 distinct twin pairs were pooled in two samples while DNA from co-twins was placed in separate pools randomly. Using the Illumina HiSeq, we generated 23 Gb of sequence for each of the pooled samples with an average of ~20 000-fold coverage per amplified region (see online supplementary table S1).

Aligning the sequencing reads to the human reference genome (build 37), the two somatic mutations identified by Sanger sequencing were confirmed. The reference allele over all alleles ratio (RAF), as determined by allelic reads, was 1 for the homozygous twin while the RAF was 0.5 and 0.6 for each of the heterozygous twin, respectively (table 1). The reads counted prior to PCR duplicates removal gave the same ratio as that subjected to duplicates removal. We did not identify additional evidence for somatic mutations from the 18 other candidate mutant single nucleotide variants (SNVs) (see online supplementary table S1).

Preliminary measurement of somatic mutation frequency

The total number of sites investigated was determined from the intersection of high-quality SNPs called in both co-twins that met the quality control criteria as shown in online supplementary table S2. The somatic mutation frequency was calculated as the number of mutation events divided by the sum of total number of SNP sites in each twin pair. The estimated mutation frequency was 1.2×10^{-7} per base pair.

We observed little evidence for striking mosaicism from the next-generation sequencing results based on RAF (see online supplementary table S1). Since we observed an RAF of 0.5–0.6 for the heterozygous twin at mutant sites, it is highly probable that the mutations we observed were generated at the cell division when two monozygotic twins were separated, which results in the almost 50:50 ratio of two alleles if the mutation causing one reference allele become the alternative allele (see online supplementary figure S1). Given that the mutation may occur in either co-twin, we estimate the somatic mutation rate as 6.0×10^{-8} per nucleotide per cell division. We note that this estimate is preliminary and is discussed below.

Given that the human genome is estimated to contain approximately 3.0×10^9 bp[25] (human genome build 37 V.2: 3156 Mb), it is likely that each individual carries approximately over 300 postzygotic mutations in the nuclear genome of white blood cells that occurred in the early development.

Mutations observed

Of the two somatic mutations we observed, one is a G to A transition (rs32598) located on chromosome 5q14, which is an intronic SNP in the *EGF-like repeats and discoidin I-like domains 3* gene (*EDIL3*, [MIM XXX 606018]). The other mutation is represented by a transversion from G to T (rs17197101) in a CpG dinucleotide site located on chromosome 6p21.33. This SNP resides in 5′UTR of the *Transcription factor 19* gene (*TCF19*, [MIM 600912]). In the Catalogue Of Somatic Mutations In Cancer (COSMIC),[26] multiple types of somatic mutations, including nonsense substitutions and missense substitutions, have been found for *EDIL3* in cancer cells, and two mutations have been found for *TCF19*. In spite of this, no mutations have been found in haematopoietic or lymphoid cancers for these genes. Thus, it is unlikely that either change conferred a proliferative or survival advantage that would interfere with the above calculations.

DISCUSSION

In this study, we have identified and verified somatic point mutations occurring in otherwise healthy individuals early in development. Our preliminary estimates of the rate of such mutations suggest that this phenomenon occurs relatively frequently. Whether such mutations may rarely give rise to non-malignant disease requires further investigation, but such exploration may
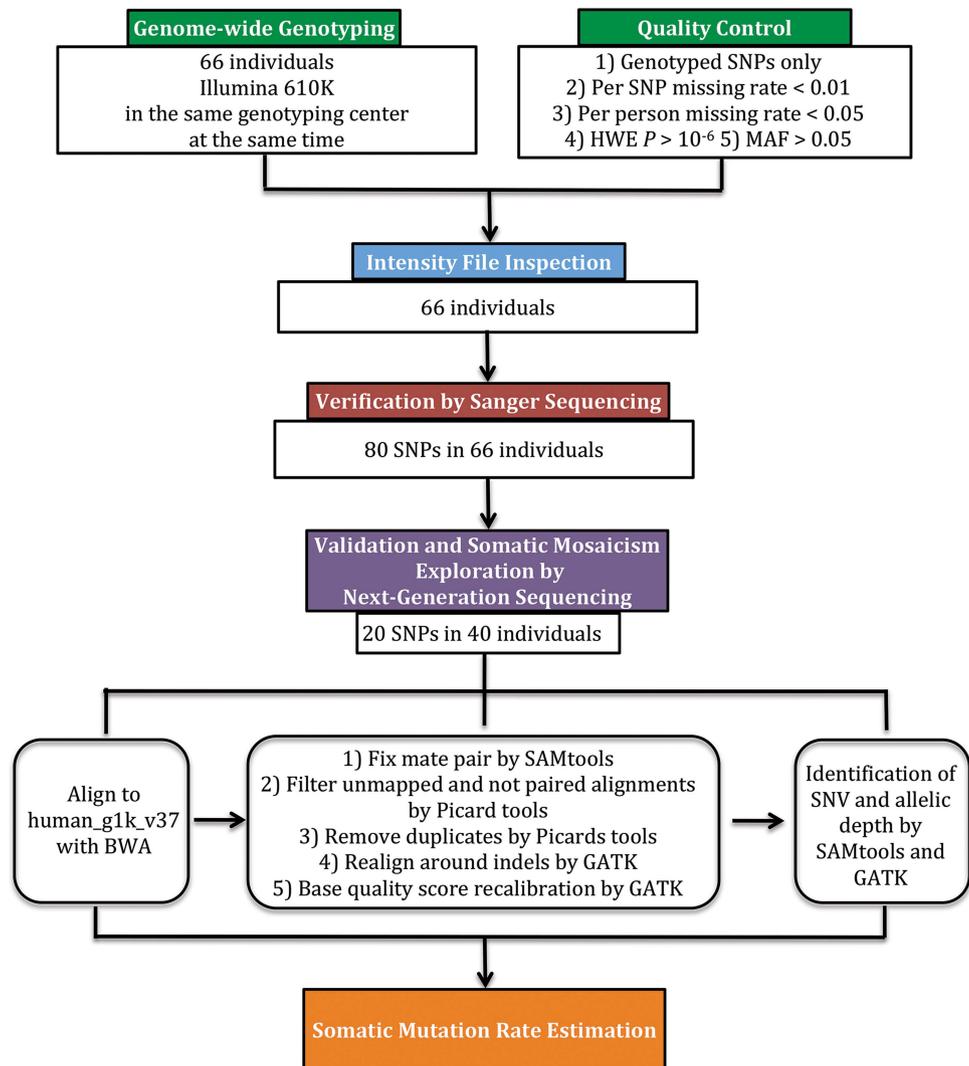
Figure 1 Flowchart of study design. MZ, monozygotic; HWE, Hardy–Weinberg Equilibrium; MAF, minor allele frequency.

identify important physiologic control points in disease aetiology rendering novel drug targets, much like the identification of driver mutations in cancer has led to important therapeutic advances in oncology.

Driver mutations in cancer have highlighted critical drug targets whose manipulation has led to changes in disease survival. For example, somatic mutations in vascular endothelial growth factor (*VEGF*) leading to non-small cell lung cancer[27]
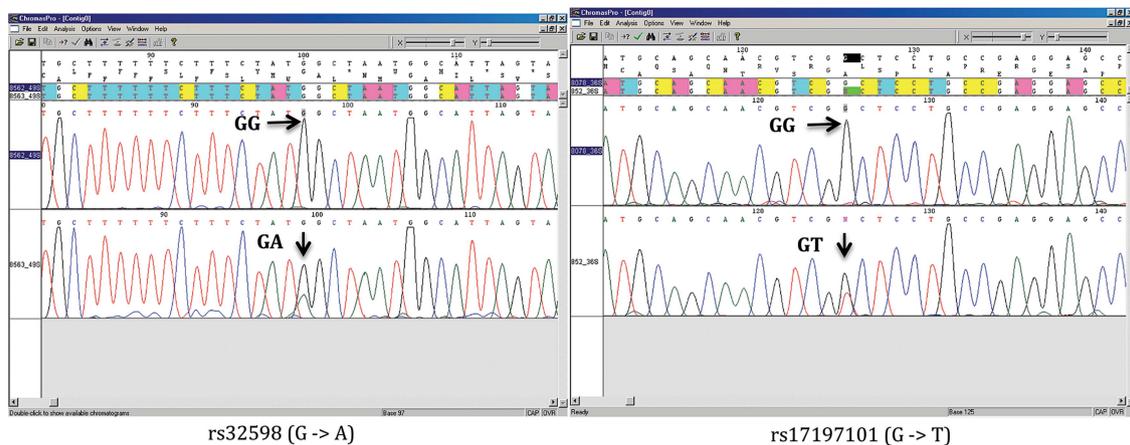
Figure 2 Sanger sequencing chromatography of identified somatic mutations.

Table 1 Somatic mutations verified by Sanger sequencing and next-generation sequencing

Genotype	rs32598		rs17197101	
	Twin1-1 RR	Twin1-2 RA	Twin2-1 RR	Twin2-2 RA
Allelic depth* (R/A)	59509/46	47601/31967	59474/54	39900/39141
RAF (N_R/N_A+N_R)	1	0.6	1	0.5
Age at sampling of blood	18	18	57	54

R, reference allele in reference genome hum_g1k_v37; A, alternative allele; RAF, reference allele frequencies; N_A, the number of reads containing alternative allele; N_R, the number of reads containing reference allele.
*Depth after duplicates were removed.

respond to bevacizumab, which neutralises *VEGF*.[28] And likewise somatic mutations in *ERBB2* (also called *HER2*) are present in lung and breast cancers,[29] and successful targeting of this tyrosine kinase with trastuzumab improves outcomes.[30] Despite these advances in malignant disease, point mutations as a cause of non-malignant disease have received little attention since such mutations would only rarely underlie disease causation and they would need to happen early enough in development, or confer a non-malignant proliferative advantage, to lead to organ disruption and consequent disease. However, despite their rarity, their identification would render a drug target that would be sufficient to cause disease.

Despite these limitations, there is evidence that somatic mutations can lead to rare diseases. McCune–Albright syndrome is due to mosaicism for an activating somatic mutation.[31] Somatic spontaneous reversions of disease-causing mutations to the wild-type allele occur in Fanconi anaemia[32] and other autosomal recessive disorders.[33–36] Thus, somatic mutations can rarely cause non-malignant disease.

Monozygotic twins are considered an attractive model for studying somatic variations. Previous studies identified large structural variants such as different copy number profiles[17 37] and chromosomal aneuploidies.[38] Genetic differences in monozygotic twins have also been demonstrated in epigenetic marks[39] and DNA changes.[3] However, in most cases, the somatic variations were observed in a later stage of life[17 39] or related to disease phenotype,[3] leaving the burden of single-point mutation arising in early life poorly described. Our study surveyed the magnitude of these kind of mutations across the entire genome. Using normal adult monozygotic twins, we provide evidence of relatively frequent mutations originating in early development.

Most early studies on somatic mutation rate only captured the ones causing impaired function in coding region or non-coding regulatory region,[6 11 12] which excludes the majority of the genome. In addition, several other factors could have an effect on mutation estimate derived from gene-based studies, such as gene structure and sensitivity of mutation detection methods. Accounting for all these factors, the average somatic point mutation rate has been estimated as 7.7×10^{-10} per nucleotide per cell division,[40] which is much lower than our estimate (6.0×10^{-8}). Moreover, a recent study sequencing whole genomes of a CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection) trio detected a somatic mutation frequency of 2.52×10^{-7} per nucleotide[41] by distinguishing germline and somatic mutation through validation from a third generation of the same family. Although this estimate is two times ours (1.2×10^{-7}), it is nevertheless influenced by unavoidable additional mutations arising in cell lines.

Our measure of mutation rate has a number of limitations, which will be overcome by future studies. First, the uncertainty of mutation rate could be improved substantially by more accurate cell depth estimate using advanced technology, for instance, a well-established method estimating cell depth based on somatic mutations in microsatellites.[42] Second, we only assessed common SNPs, which may not fully represent the nature of mutation at sites of very low alternate allele frequencies.[43] Large-scale sequencing projects on monozygotic twins, for example, the UK10K project, will be able to assess the mutation rate at rare SNPs, but will always be encumbered by the relatively high error rates of next-generation sequencing. Third, comparing other cell types of different lineages than leucocytes would better assess the timing of these somatic mutations. However, only blood samples were collected from these subjects and both twin pairs did not permit sampling of other tissues.

Our data provide evidence that early somatic mutations do occur, and based on the mutation frequency observed in our study, it can be inferred that each individual carries, on the average, greater than ~300 postzygotic mutations that occurred early enough to be present in the majority of blood cells. These mutations are not constrained by the need to be compatible with whole-organism survival when they first occur. However, they provide fuel for phenotypic variation, which may lead to tissue disruption and disease in later life. It has been speculated that, for any gene, a small number of individuals in the population carry a somatic mutation in a large majority of cells that have occurred in early development.[20] Our data provide evidence to support this concept.

In summary, our findings support the emerging consensus that somatic point mutations arising in development are a relatively common phenomenon. The mutation frequency we observed may serve as the first step in refining important parameters influencing somatic non-malignant disease.

MATERIALS AND METHODS
Data generation
The monozygotic twins for this study were collected from 3512 individuals in the TwinsUK cohort that have undergone a genome-wide scan using DNA sample extracted from blood. The TwinsUK cohort includes approximately 12 000 volunteer twins from all over the UK,[44] who have no differences in age-matched characteristics compared with the general population.[45] We performed the genome-wide genotyping by using the Infinium 610k assay (Illumina, San Diego, USA) at two facilities: the Center for Inherited Diseases Research (CIDR) in Baltimore, Maryland, USA, and the Wellcome Trust Sanger Institute in UK. Co-twins that were not genotyped in the same centre at the same time were excluded to avoid the possible sources of spurious mutation findings caused by lab-based systematic differences in genotyping. We used the Illuminus calling algorithm[46] to assign SNP genotypes. Null genotypes were assigned if an

individual's most likely genotype was called with less than a posterior probability threshold of 0.95. Only the highest quality common SNPs (minor allele frequency >0.05) that had a call rate greater than 0.99 were retained for analysis. In addition, we filtered SNPs demonstrating deviation from Hardy–Weinberg Equilibrium ($p<10^{-6}$). The zygosity of the twins was confirmed based on concordance rates. SNP genotypes were compared between monozygotic co-twins to identify putative somatic mutations for the following analysis. We obtained informed consent from all participants before they entered the studies. Ethics approval was obtained from the local research ethics committee.

Data filtering to identify candidate somatic mutations for verification

We trimmed the list of putative early somatic mutations via visual inspection of intensity data. The normalisation and base calling were done using Illumina Genome Studio Genotyping Module V.1.9.4. SNPs with discordant calls between co-twins were manually reviewed.

Verification by Sanger sequencing

For the filtered candidate somatic mutations, we performed PCR amplification using 30 ng of genomic DNA from each of the twins. We designed PCR primers by using Primer-3 (http://frodo.wi.mit.edu/) to amplify PCR products in size from 260 to 572 bp (supplementary note). We used High Fidelity Taq Polymerase (error rate is 4×10^{-6}) from Roche in PCR. PCR products were sent for Sanger sequencing and also used to form two pools for next-generation sequencing. Each twin from each pair was randomly assigned to one of the two pools, A or B.

Mutation verification and somatic mosaicism exploration by next-generation sequencing

Before pooling, PCR products were quantified by the PicoGreen kit (Life Technologies Canada, Burlington ON) and 160 ng of each PCR fragment was added to the pool. The pooled DNA was then seared in a Covaris S200 sonicator at peak incident power set at 175 and 200 cycles per burst for 430 s to obtain a fragment distribution with a peak around 150 bp (see online supplementary figure S2). Each pool was separately barcoded. Libraries were constructed by following Illumina pair-end library protocol. The resulting library for each pool was sequenced in a single lane of an Illumina HiSeq GAII analyser flow cell to generate 101 bp paired-end reads according to the manufacturer's instructions.

Sequencing data analysis and somatic mutation identification

For Sanger sequencing, somatic mutations were identified by directly comparing genotypes of co-twins from chromatograms. For next-generation sequencing, the paired-end reads were aligned to reference genome (build 37) using Burrows-Wheeler Aligner (BWA, V.0.6.1-r104) with standard parameters.[47] Further postalignment steps were undertaken to minimise false-positive findings from error-prone sequencing results: (1) SAMtools (V.0.1.18-r982:295)[48] was used to synchronise the mate pair information. Alignments were filtered for mapping and pairing. (2) PCR duplicates were removed by Picard tools (V.1.77) (http://picard.sourceforge.net/). (3) Genome Analysis Toolkit (GATK, V.2.0)[49] was used to realign the reads on the edges of potential insertions and/or deletions to reduce SNV artefacts. (4) Recalibration was also applied to base quality scores by using GATK. SAMtools was used to identify single-nucleotide

variations and obtain allelic depth. Allelic depth was used to generate RAF. RAF was given by $N_R/(N_A+N_R)$, in which N_R and N_A are the number of reads containing reference allele and alternative allele, respectively.

Calculations of somatic mutation frequency

The somatic mutation frequency, f, is calculated as the number of mutation events (m) divided by the sum of total number of SNP sites (S) in each twin pair (T):

$$f = m/\sum\nolimits_{i=1}^{T} S_i$$

All 33 twin pairs were considered in the calculation. The total number of sites investigated was determined from the intersection of high-quality SNPs called in both co-twins that met the quality control criteria as shown in online supplementary table S2. We note that somatic mutations could possibly occur in the hybridisation probe; however, genotypes were generated for SNP sites on the Illumina assay. Thus, only the total number of SNPs assayed (not including those on the probe) was considered for calculation of mutation frequency.

We estimated the mutation rate by the following formula[6]:

$$\mu = \frac{f}{2n}$$

where μ is the single-nucleotide mutation rate, 2 is used to quantify the fact that the mutation may occur in either co-twin, f is the frequency of somatic mutations confirmed by Sanger sequencing experiments and n is the number of cell division that had occurred up to the time of the mutation.

Our study was only able to detect mutations occurring in the exponential growth phase of development rather than tissue renewal[20] since we employed genome-wide genotyping for mutation identification. Since the RAF was approximately 0.5 and 1 in heterozygous and homozygous twins, respectively, such a mutation most likely has occurred at the cell division when two twins were separated. Thus, n was estimated to be 1.

REFERENCES

1 Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;194:23–8.
2 Erickson RP. Somatic gene mutation and human disease other than cancer: an update. *Mutat Res* 2010;705:96–106.
3 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.
4 Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, Turner J, Cannons JL, Bick D, Blakemore L, Blumhorst C, Brockmann K, Calder P, Cherman N, Deardorff MA, Everman DB, Golas G, Greenstein RM, Kato BM, Keppler-Noreuil KM, Kuznetsov SA, Miyamoto RT, Newman K, Ng D, O'Brien K, Rothenberg S, Schwartzentruber DJ, Singhal V, Tirabosco R, Upton J, Wientroub S, Zackai EH, Hoag K, Whitewood-Neal T, Robey PG, Schwartzberg PL, Darling TN, Tosi LL, Mullikin JC, Biesecker LG. A mosaic activating mutation in AKT1 associated with the Proteus syndrome. *N Engl J Med* 2011;365:611–19.

5   Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, Mollica L, Li J, Viale A, Heguy A, Hassimi M, Socci N, Bhatt PK, Gonen M, Mason CE, Melnick A, Godley LA, Brennan CW, Abdel-Wahab O, Levine RL. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* 2012;44:1179–81.

6   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.

7   Araten DJ, Luzzatto L. The mutation rate in PIG-A is normal in patients with paroxysmal nocturnal hemoglobinuria (PNH). *Blood* 2006;108:734–6.

8   Lichtenauer-Kaligis EG, Thijssen J, den Dulk H, van de Putte P, Tasseron-de Jong JG, Giphart-Gassler M. Comparison of spontaneous hprt mutation spectra at the nucleotide sequence level in the endogenous hprt gene and five other genomic positions. *Mutat Res* 1996;351:147–55.

9   Umar A, Risinger JI, Glaab WE, Tindall KR, Barrett JC, Kunkel TA. Functional overlap in mismatch repair by human MSH3 and MSH6. *Genetics* 1998;148:1637–46.

10   Glaab WE, Tindall KR. Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects. *Carcinogenesis* 1997;18:1–8.

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

12   Hornsby C, Page KM, Tomlinson I. The in vivo rate of somatic adenomatous polyposis coli mutation. *Am J Pathol* 2008;172:1062–8.

13   Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, Tomsho LP, Peters BA, Pujara K, Cordes S, Davis DP, Carlton VE, Yuan W, Li L, Wang W, Eigenbrot C, Kaminker JS, Eberhard DA, Waring P, Schuster SC, Modrusan Z, Zhang Z, Stokoe D, de Sauvage FJ, Faham M, Seshagiri S. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* 2010;466:869–73.

14   Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, Koker M, Dahmen I, Muller C, Di Cerbo V, Schildhaus HU, Altmuller J, Baessmann I, Becker C, de Wilde B, Vandesompele J, Bohm D, Ansen S, Gabler F, Wilkening I, Heynck S, Heuckmann JM, Lu X, Carter SL, Cibulskis K, Banerji S, Getz G, Park KS, Rauh D, Grutter C, Fischer M, Pasqualucci L, Wright G, Wainer Z, Russell P, Petersen I, Chen Y, Stoelben E, Ludwig C, Schnabel P, Hoffmann H, Muley T, Brockmann M, Engel-Riedel W, Muscarella LA, Fazio VM, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman DA, Snijders PJ, Cappuzzo F, Ligorio C, Damiani S, Field J, Solberg S, Brustugun OT, Lund-Iversen M, Sanger J, Clement JH, Soltermann A, Moch H, Weder W, Solomon B, Soria JC, Validire P, Besse B, Brambilla E, Brambilla C, Lantuejoul S, Lorimier P, Schneider PM, Hallek M, Pao W, Meyerson M, Sage J, Shendure J, Schneider R, Buttner R, Wolf J, Nurnberg P, Perner S, Heukamp LC, Brindle PK, Haas S, Thomas RK. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet* 2012;44:1104–10.

15   Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansen S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Janne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;150:1107–20.

16   Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennemer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang Z. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 2010;465:473–7.

17   Forsberg LA, Rasi C, Razzaghian HR, Pakalapati G, Waite L, Thilbeault KS, Ronowicz A, Wineinger NE, Tiwari HK, Boomsma D, Westerman MP, Harris JR, Lyle R, Essand M, Eriksson F, Assimes TL, Iribarren C, Strachan E, O'Hanlon TP, Rider LG, Miller FW, Giedraitis V, Lannfelt L, Ingelsson M, Piotrowski A, Pedersen NL, Absher D, Dumanski JP. Age-related somatic structural changes in the nuclear genome of human blood cells. *Am J Hum Genet* 2012;90:217–28.

18   Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, Ling H, Hetrick KN, Pugh EW, Amos C, Wei Q, Wang LE, Lee JE, Barnes KC, Hansel NN, Mathias R, Daley D, Beaty TH, Scott AF, Ruczinski I, Scharpf RB, Bierut LJ, Hartz SM, Landi MT, Freedman ND, Goldin LR, Ginsburg D, Li J, Desch KC, Strom SS, Blot WJ, Signorello LB, Ingles SA, Chanock SJ, Berndt SI, Le Marchand L, Henderson BE, Monroe KR, Heit JA, de Andrade M, Armasu SM, Regnier C, Lowe WL, Hayes MG, Marazita ML, Feingold E, Murray JC, Melbye M, Feenstra B, Kang JH, Wiggs JL, Jarvik GP, McDavid AN, Seshan VE, Mirel DB, Crenshaw A, Sharopova N, Wise A, Shen J, Crosslin DR, Levine DM, Zheng X, Udren JI, Bennett S, Nelson SC, Gogarten SM, Conomos MP, Heagerty P, Manolio T, Pasquale LR, Haiman CA, Caporaso N, Weir BS. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nature genetics* 2012;44:642–50.

19   Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, Hutchinson A, Deng X, Liu C, Horner MJ, Cullen M, Epstein CG, Burdett L, Dean MC, Chatterjee N, Sampson J, Chung CC, Kovaks J, Gapstur SM, Stevens VL,

Teras LT, Gaudet MM, Albanes D, Weinstein SJ, Virtamo J, Taylor PR, Freedman ND, Abnet CC, Goldstein AM, Hu N, Yu K, Yuan JM, Liao L, Ding T, Qiao YL, Gao YT, Koh WP, Xiang YB, Tang ZZ, Fan JH, Aldrich MC, Amos C, Blot WJ, Bock CH, Gillanders EM, Harris CC, Haiman CA, Henderson BE, Kolonel LN, Le Marchand L, McNeill LH, Rybicki BA, Schwartz AG, Signorello LB, Spitz MR, Wiencke JK, Wrensch M, Wu X, Zanetti KA, Ziegler RG, Figueroa JD, Garcia-Closas M, Malats N, Marenne G, Prokunina-Olsson L, Baris D, Schwenn M, Johnson A, Landi MT, Goldin L, Consonni D, Bertazzi PA, Rotunno M, Rajaraman P, Andersson U, Beane Freeman LE, Berg CD, Buring JE, Butler MA, Carreon T, Feychting M, Ahlbom A, Gaziano JM, Giles GG, Hallmans G, Hankinson SE, Hartge P, Henriksson R, Inskip PD, Johansen C, Landgren A, McKean-Cowdin R, Michaud DS, Melin BS, Peters U, Ruder AM, Sesso HD, Severi G, Shu XO, Visvanathan K, White E, Wolk A, Zeleniuch-Jacquotte A, Zheng W, Silverman DT, Kogevinas M, Gonzalez JR, Villa O, Li D, Duell EJ, Risch HA, Olson SH, Kooperberg C, Wolpin BM, Jiao L, Hassan M, Wheeler W, Arslan AA, Bueno-de-Mesquita HB, Fuchs CS, Gallinger S, Gross MD, Holly EA, Klein AP, LaCroix A, Mandelson MT, Petersen G, Boutron-Ruault MC, Bracci PM, Canzian F, Chang K, Cotterchio M, Giovannucci EL, Goggins M, Hoffman Bolton JA, Jenab M, Khaw KT, Krogh V, Kurtz RC, McWilliams RR, Mendelsohn JB, Rabe KG, Riboli E, Tjonneland A, Tobias GS, Trichopoulos D, Elena JW, Yu H, Amundadottir L, Stolzenberg-Solomon RZ, Kraft P, Schumacher F, Stram D, Savage SA, Mirabello L, Andrulis IL, Wunder JS, Patino Garcia A, Sierrasesumaga L, Barkauskas DA, Gorlick RG, Purdue M, Chow WH, Moore LE, Schwartz KL, Davis FG, Hsing AW, Berndt SI, Black A, Wentzensen N, Brinton LA, Lissowska J, Peplonska B, McGlynn KA, Cook MB, Graubard BI, Kratz CP, Greene MH, Erickson RL, Hunter DJ, Thomas G, Hoover RN, Real FX, Fraumeni JF Jr, Caporaso NE, Tucker M, Rothman N, Perez-Jurado L, Chanock SJ. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet* 2012;44:651–8.

20   Frank SA. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc Natl Acad Sci USA* 2010;107(Suppl 1):1725–30.

21   Cairns J. Mutation selection and the natural history of cancer. *Nature* 1975;255:197–200.

22   Aretz S, Stienen D, Friedrichs N, Stemmler S, Uhlhaas S, Rahner N, Propping P, Friedl W. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). *Hum Mutat* 2007;28:985–92.

23   Hes FJ, Nielsen M, Bik EC, Konvalinka D, Wijnen JT, Bakker E, Vasen HF, Breuning MH, Tops CM. Somatic APC mosaicism: an underestimated cause of polyposis coli. *Gut* 2008;57:71–6.

24   Rohlin A, Wernersson J, Engwall Y, Wiklund L, Bjork J, Nordling M. Parallel sequencing used in detection of mosaic mutations: comparison with four diagnostic DNA screening techniques. *Hum Mutat* 2009;30:1012–20.

25   Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931–45.

26   Ndegwa N, Cote RG, Ovelleiro D, D'Eustachio P, Hermjakob H, Vizcaino JA, Croft D. Critical amino acid residues in proteins: a BioMart integration of Reactome protein annotations with PRIDE mass spectrometry data and COSMIC somatic mutations. *Database: J Biolo Databases Curation* 2011;2011:bar047.

27   Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, Wendl MC, Lawrence MS, Larson DE, Chen K, Dooling DJ, Sabo A, Hawes AC, Shen H, Jhangiani SN, Lewis LR, Hall O, Zhu Y, Mathew T, Ren Y, Yao J, Scherer SE, Clerc K, Metcalf GA, Ng B, Milosavljevic A, Gonzalez-Garay ML, Osborne JR, Meyer R, Shi X, Tang Y, Koboldt DC, Lin L, Abbott R, Miner TL, Pohl C, Fewell G, Haipek C, Schmidt H, Dunford-Shore BH, Kraja A, Crosby SD, Sawyer CS, Vickery T, Sander S, Robinson W, Winckler W, Baldwin J, Chirieac LR, Dutt A, Fennell T, Hanna M, Johnson BE, Onofrio RC, Thomas RK, Tonon G, Weir BA, Zhao X, Ziaugra L, Zody MC, Giordano T, Orringer MB, Roth JA, Spitz MR, Wistuba II, Ozenberger B, Good PJ, Chang AC, Beer DG, Watson MA, Ladanyi M, Broderick S, Yoshizawa A, Travis WD, Pao W, Province MA, Weinstock GM, Varmus HE, Gabriel SB, Lander ES, Gibbs RA, Meyerson M, Wilson RK. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 2008;455:1069–75.

28   Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, Lilenbaum R, Johnson DH. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med* 2006;355:2542–50.

29   Stephens P, Hunter C, Bignell G, Edkins S, Davies H, Teague J, Stevens C, O'Meara S, Smith R, Parker A, Barthorpe A, Blow M, Brackenbury L, Butler A, Clarke O, Cole J, Dicks E, Dike A, Drozd A, Edwards K, Forbes S, Foster R, Gray K, Greenman C, Halliday K, Hills K, Kosmidou V, Lugg R, Menzies A, Perry J, Petty R, Raine K, Ratford L, Shepherd R, Small A, Stephens Y, Tofts C, Varian J, West S, Widaa S, Yates A, Brasseur F, Cooper CS, Flanagan AM, Knowles M, Leung SY, Louis DN, Looijenga LH, Malkowicz B, Pierotti MA, Teh B, Chenevix-Trench G, Weber BL, Yuen ST, Harris G, Goldstraw P, Nicholson AG, Futreal PA, Wooster R, Stratton MR. Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* 2004;431:525–6.

30   Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, Slamon DJ, Murphy M, Novotny WF, Burchmore M, Shak S, Stewart SJ, Press M. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20:719–26.

31 Happle R. The McCune-Albright syndrome: a lethal gene surviving by mosaicism. *Clin Genet* 1986;29:321–4.

32 Waisfisz Q, Morgan NV, Savino M, de Winter JP, van Berkel CG, Hoatlin ME, Ianzano L, Gibson RA, Arwert F, Savoia A, Mathew CG, Pronk JC, Joenje H. Spontaneous functional correction of homozygous fanconi anaemia alleles reveals novel mechanistic basis for reverse mosaicism. *Nat Genet* 1999;22:379–83.

33 Kvittingen EA, Rootwelt H, Brandtzaeg P, Bergan A, Berger R. Hereditary tyrosinemia type I. Self-induced correction of the fumarylacetoacetase defect. *J Clin Invest* 1993;91:1816–21.

34 Ellis NA, Lennon DJ, Proytcheva M, Alhadeff B, Henderson EE, German J. Somatic intragenic recombination within the mutated locus BLM can correct the high sister-chromatid exchange phenotype of Bloom syndrome cells. *Am J Hum Genet* 1995;57:1019–27.

35 Stephan JL, Vlekova V, Le Deist F, Blanche S, Donadieu J, De Saint-Basile G, Durandy A, Griscelli C, Fischer A. Severe combined immunodeficiency: a retrospective single-center study of clinical presentation and outcome in 117 patients. *J Pediatr* 1993;123:564–72.

36 Jonkman MF, Scheffer H, Stulp R, Pas HH, Nijenhuis M, Heeres K, Owaribe K, Pulkkinen L, Uitto J. Revertant mosaicism in epidermolysis bullosa caused by mitotic gene conversion. *Cell* 1997;88:543–51.

37 Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, Diaz de Stahl T, Menzel U, Sandgren J, von Tell D, Poplawski A, Crowley M, Crasto C, Partridge EC, Tiwari H, Allison DB, Komorowski J, van Ommen GJ, Boomsma DI, Pedersen NL, den Dunnen JT, Wirdefeldt K, Dumanski JP. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* 2008;82:763–71.

38 Razzaghian HR, Shahi MH, Forsberg LA, de Stahl TD, Absher D, Dahl N, Westerman MP, Dumanski JP. Somatic mosaicism for chromosome X and Y aneuploidies in monozygotic twins heterozygous for sickle cell disease mutation. *Am J Med Genet* 2010;152A:2595–8.

39 Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, Heine-Suner D, Cigudosa JC, Urioste M, Benitez J, Boix-Chornet M, Sanchez-Aguilera A, Ling C, Carlsson E, Poulsen P, Vaag A, Stephan Z, Spector TD, Wu YZ, Plass C, Esteller M. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci USA* 2005;102:10604–9.

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

41 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. Variation in genome-wide mutation rates within and between human families. *Nat Genet* 2011;43:712–14.

42 Wasserstrom A, Frumkin D, Adar R, Itzkovitz S, Stern T, Kaplan S, Shefer G, Shur I, Zangi L, Reizel Y, Harmelin A, Dor Y, Dekel N, Reisner Y, Benayahu D, Tzahor E, Segal E, Shapiro E. Estimating cell depth from somatic mutations. *PLoS Comput Biol* 2008;4:e1000058.

43 Messer PW. Measuring the rates of spontaneous mutation from deep and large-scale polymorphism data. *Genetics* 2009;182:1219–32.

44 Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int J Epidemiol* 2013;42:76–85.

45 Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* 2001;4:464–77.

46 Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, Kwiatkowski DP, Clark TG. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 2007;23:2741–6.

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

48 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.

49 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.

Somatic point mutations occurring early in development: a monozygotic twin study

Rui Li, Alexandre Montpetit, Marylène Rousseau, et al.

Updated information and services can be found at:

http://jmg.bmj.com/content/early/2013/10/11/jmedgenet-2013-101712.full.html

These include:

Data Supplement	*"Supplementary Data"* http://jmg.bmj.com/content/suppl/2013/10/11/jmedgenet-2013-101712.DC1.html
References	This article cites 47 articles, 16 of which can be accessed free at: http://jmg.bmj.com/content/early/2013/10/11/jmedgenet-2013-101712.full.html#ref-list-1
P<P	Published online October 11, 2013 in advance of the print journal.
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 (557 articles) Immunology (including allergy) (520 articles) Reproductive medicine (476 articles)

Advance online articles have been peer reviewed, accepted for publication, edited and typeset, but have not not yet appeared in the paper journal. Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

Notes

Advance online articles have been peer reviewed, accepted for publication, edited and typeset, but have not not yet appeared in the paper journal. Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

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/