

***MUTYH* c.933+3A>C, associated with a severely impaired gene expression, is the first Italian founder mutation in *MUTYH*-Associated Polyposis**

Elisa Pin¹, Chiara Pastrello¹, Rossella Tricarico², Laura Papi², Michele Quaia¹, Mara Fornasarig³, Ileana Carnevali⁴, Cristina Oliani⁵, Alessio Fornasin⁶, Marco Agostini^{7,8,9}, Roberta Maestro¹, Daniela Barana⁵, Stefan Aretz¹⁰, Maurizio Genuardi^{2,11} and Alessandra Viel¹

¹Oncologia Sperimentale 1, Centro di Riferimento Oncologico, IRCCS, Aviano (PN), Italy

²Dipartimento di Fisiopatologia Clinica, Unità di Genetica Medica, Università di Firenze, Florence, Italy

³Unità di Gastroenterologia, Centro di Riferimento Oncologico, IRCCS, Aviano (PN), Italy

⁴U.O. Anatomia e Istologia Patologica, Presidio Ospedaliero di Varese, Varese, Italy

⁵U.O.C. di Oncologia ULSS5 Ovest Vicentino, Ospedale di Montebelluna, Montebelluna Maggiore (VI), Italy

⁶Dipartimento di Scienze Economiche e Statistiche, Università di Udine, Udine, Italy

⁷Clinica Chirurgica II, Dipartimento di Scienze Chirurgiche Oncologiche e Gastroenterologiche, Università di Padova, Padua, Italy

⁸Department of Nanomedicine, The Methodist Hospital Research Institute, Houston, TX, USA

⁹Istituto di Ricerca Pediatrica - Città della Speranza, Padua, Italy

¹⁰Institute of Human Genetics, University of Bonn, Bonn, Germany

¹¹FiorGen Foundation for Pharmacogenomics, Sesto Fiorentino (FI), Italy

***MUTYH* variants are differently distributed in geographical areas of the world. In *MUTYH*-associated polyposis (MAP) patients from North-Eastern Italy, c.933+3A>C (IVS10+3A>C), a transversion causing an aberrant splicing process, accounts for nearly 1/5 of all mutations. The aim of this study was to verify whether its high frequency in North-Eastern Italy is due to a founder effect and to clarify its impact on *MUTYH* transcripts and protein. Haplotype analysis and age estimate performed on members of eleven Italian MAP families and cancer-free controls provided evidence that c.933+3A>C is a founder mutation originated about 83 generations ago. In addition, the Italian haplotype associated with the c.933+3A>C was also found in German families segregating the same mutation, indicating it had a common origin in Western Europe. Altogether c.933+3A>C and the two common Caucasian mutations p.Tyr179Cys and p.Gly396Asp represent about 60% of *MUTYH* alterations in MAP patients from North-Eastern Italy, suggesting the opportunity to perform targeted molecular screening for these variants in the diagnostic setting. Expression analyses performed on lymphoblastoid cell lines supported the notion that *MUTYH* c.933+3A>C alters splicing causing the synthesis of a non functional protein. However, some primary transcripts escape aberrant splicing, producing traces of full-length transcript and wild-type protein in a homozygote; this is in agreement with clinical findings that suggest a relatively mild phenotypic effect for this mutation. Overall, these data, that demonstrate a founder effect and further elucidate the splicing alterations caused by the *MUTYH* c.933+3A>C mutation, have important implications for genetic counseling and molecular diagnosis of MAP.**

Key words: *MUTYH*, splicing, colorectal cancer, founder mutation, lymphoblastoid cell lines

Grant sponsors: ACC (Alleanza Contro il Cancro)—INTEF Project, Ministero della Salute—Rare Disease Project, Istituto Toscano Tumori, Ente Cassa di Risparmio di Firenze

DOI: 10.1002/ijc.27761

History: Received 15 Jun 2012; Accepted 20 Jul 2012; Online 3 Aug 2012

Correspondence to: Alessandra Viel, Oncologia Sperimentale 1, Centro di Riferimento Oncologico, IRCCS, Via Franco Gallini 2, 33081 Aviano, Italy, Tel.: +39 0434 659671, Fax: +39 0434 659659, E-mail: aviel@cro.it

Introduction

It is estimated that 5% of colorectal cancer (CRC) cases can be explained by inherited CRC syndromes.¹ Several tumor suppressor genes (*e.g.*, *APC*, *TP53* and *CDKN2A*), proto-oncogenes (*e.g.*, *KRAS*) and DNA repair genes² are involved in CRC carcinogenesis. The latter includes genes of the mismatch repair system and *MUTYH*, a gene belonging to the base excision repair system.

MUTYH (OMIM 604933) encodes a DNA glycosylase involved in the repair of 8-oxoguanine (8-oxoG), one of the typical mutagenic lesions generated by oxidative stress. Since 8-oxoG mispairs with adenine, *MUTYH* excises the mispaired adenine, thus preventing G:C>T:A transversions. Biallelic germline mutations in *MUTYH* were shown to result in multiple colorectal adenomas and CRC, likely as a result of

What's new?

Polyposis-associated variations in the base excision repair gene *MUTYH* are distributed unevenly across different ethnicities, suggesting diverse mutation origins and founder effects. This haplotype analysis successfully documented the first *MUTYH* founder mutation—a splice variant known as c.933+3A>C, frequently detected in polyposis patients from northeastern Italy. Knowledge of the mutation's origins in this population could have important implications for genetic counseling and diagnosis.

the accumulation of G>T somatic mutations in *APC*, *KRAS*, and presumably other target genes.^{3,4}

MUTYH-Associated Polyposis (MAP; OMIM 608456) is an inherited polyposis syndrome with autosomal recessive transmission, accounting for about 1% of all CRC cases.¹ The significance of monoallelic *MUTYH* mutations is still debated; our previous study on Italian patients, as well as larger case-control studies, provided evidence that the condition of heterozygosity for *MUTYH* mutations contributes to the development of some CRCs.^{5,6}

The distribution of *MUTYH* mutations shows ethnic differences.^{7,8} p.Tyr179Cys (c.536A>G) and p.Gly396Asp (c.1187G>A) are the most frequent mutations in Caucasians,⁹ while other population-specific mutations have been reported in Japan (p.Arg245Cys and c.934-2A>C),¹⁰ Portugal (c.1227_1228dup)¹¹ and Finland (p.Ala473Asp).¹² As expected, in Italy the most common mutations are p.Tyr179Cys and p.Gly396Asp, associated with classical MAP phenotypes with some degree of variability. An Italian study found a relatively high frequency of the c.1437delGGA mutation,¹³ but its effective contribution to MAP in our country seems relatively marginal.

We recently detected c.933+3A>C, a *MUTYH* mutation already reported in the Leiden Open Variation Database (LOVD),¹⁴ (http://chromium.liacs.nl/LOVD2/colon_cancer/home.php?select_db=MUTYH) in 11 unrelated cases, including one homozygote. This splicing variant accounts for about 15% of the *MUTYH* mutations identified in our cohort of MAP patients (12 out of 82 mutant alleles), coming mainly from the North-East of Italy. This finding appears particularly relevant if compared to the frequencies of the most common p.Tyr179Cys and p.Gly396Asp, each of which accounted for 22% of mutant alleles on the same series of cases, a relatively smaller fraction compared with other European regions. Of note, c.933+3A>C has been reported also in other populations, but at lower frequencies (1–8%).^{7,9,15–26} Since these data suggest the possible presence of a founder effect in the North-Eastern Italian population, we undertook haplotype analysis to verify this hypothesis. We also investigated in detail the molecular effects of this mutation.

Material and Methods**Cases and controls**

Since 2003 we have sequenced the *MUTYH* gene in 128 *APC* mutation-negative patients with polyposis referred to the Centro di Riferimento Oncologico in Aviano. We found 17 different *MUTYH* nonpolymorphic variants in 35 cases

(27%). This study was carried out on 11 apparently unrelated families of probands with the c.933+3A>C mutation: 1 homozygote, 1 simple heterozygote and 9 compound heterozygotes (Table 1). Ten patients had attenuated polyposis, one had diffuse polyposis, and seven had developed CRC. Detailed family histories, including details on geographic origins were obtained for all patients. Genealogic investigations did not reveal any relationship between individuals from different families. All patients were resident in North-Eastern Italy and/or had North-Eastern Italian origin, specifically from the Veneto and Friuli-Venezia Giulia regions (Fig. 1).

Seventy healthy controls, native from North-Eastern Italy, were investigated to estimate allele frequencies in the general population: these included 58 individuals with negative colonoscopy (clean colon), and 12 subjects belonging to nuclear pedigrees without history of CRC (parents from 6 father–mother–child trios). In addition, 8 noncarrier chromosomes from healthy relatives belonging to the Italian MAP families segregating the *MUTYH* recurrent mutation, were used to estimate control haplotypes.

DNA samples for haplotype analysis were also collected from members of 11 German MAP families segregating c.933+3A>C, ascertained through two homozygous and 9 compound heterozygous probands.

Overall, 37 individuals (22 probands and 15 relatives) from Italian and German families segregating the c.933+3A>C *MUTYH* mutation were investigated. One to three relatives were genotyped in 10 families. In the remaining families, only the index case was available for genotyping.

Samples from two healthy individuals with wild type (WT) *MUTYH* genotype were used as controls for mRNA and protein expression.

Informed consent was obtained from all study participants. The genetic testing protocol has been evaluated and approved by the Local Independent Ethical Committee (CRO-15-1997).

Sample collection and lymphoblastoid cell line establishment

Lymphocytes were collected from 20 ml of peripheral blood using LymphoprepTM solution (Axis-Shield, Dundee, Scotland) according to manufacturer's instructions and resuspended in 90% FBS, 10% DMSO, frozen at –80°C and then stored in liquid nitrogen. Lymphoblastoid cell lines (LCLs) were established from one c.933+3A>C homozygote, four c.933+3A>C compound heterozygotes (Table 1), and two

Table 1. Genotypes and clinical characteristics of the families segregating the c.933+3A>C *MUTYH* mutation

FAM	Case ID	Genotype ¹	Family relationship	Phenotype ²	LCL ³
1	FAP82	c.[536A>G]; c.[933+3A>C]	PROBAND	AP; CRC 50 y	No
2	FAP117	c.[933+3A>C]; c.[1437delGGA]	PROBAND	DP	Yes
3	FAP181	c.[933+3A>C]; c.[1187G>A]	PROBAND	AP; CRC 39 y	Yes
4	FAP278	c.[933+3A>C]; c.[1147delC]	PROBAND	AP; CRC 45 y	Yes
4	FAP341	c.[933+3A>C]; [=]	BROTHER	–	No
4	FAP345	c.[933+3A>C]; [=]	BROTHER	–	No
5	FAP294	c.[933+3A>C]; c.[734G>A]	PROBAND	AP	Yes
6	FAP349	c.[933+3A>C]; c.[933+3A>C]	PROBAND	AP	Yes
6	FAP359	c.[933+3A>C]; [=]	BROTHER	–	No
7	FAP352 ⁴	c.[933+3A>C]; [=]	PROBAND	AP; CRC 45 y	No
7	FAP411	WT	SON	–	No
7	FAP412	c.[933+3A>C]; [=]	SON	–	No
8	FAP426	c.[933+3A>C]; c.[536A>G]	PROBAND	AP; CRC 44 y	No
8	FAP536	c.[933+3A>C]; [=]	SISTER	–	No
8	FAP537	c.[536A>G]; [=]	BROTHER	–	No
8	FAP538	c.[933+3A>C]; [=]	SISTER	–	No
9	FAP443	c.[933+3A>C]; c.[421delC]	PROBAND	AP	No
9	FAP446	c.[933+3A>C]; c.[421delC]	SISTER	–	No
10	CFS167	c.[933+3A>C]; c.[1147delC]	PROBAND	AP; CRC 40 y	No
10	FAP214	c.[933+3A>C]; [=]	BROTHER	–	No
11	CFS443	c.[933+3A>C]; c.[1437delGGA]	PROBAND	AP; CRC 54 y	No
11	FAP271	c.[1437delGGA]; [=]	SON	–	No
11	FAP272	c.[1437delGGA]; [=]	SON	–	No

¹GenBank reference sequence *MUTYH*: NM_001128425.1. ²DP, diffuse polyposis; AP, attenuated polyposis; CRC, colorectal cancer; –, no tumor.

³LCL, lymphoblastoid cell line, available (Yes) or not (No). ⁴A single *MUTYH* mutation was identified in this proband, who had a typical MAP phenotype. Presence of an additional undetected mutation cannot be excluded.

MUTYH WT controls. Lymphocytes from patients and controls were thawed rapidly at 37°C and immortalized through EBV infection in presence of cyclosporin A. LCLs were grown in RPMI 10% FBS and 1% penicillin-streptomycin at 37°C and 5% CO₂.

Microsatellite analysis

A set of 11 microsatellite markers spanning a region of ~8 Mb covering the *MUTYH* locus was investigated. The following microsatellites, listed in order from telomere to centromere, were analyzed: *DIS2861*, *DIS211*, *DIS421*, *DIS451*, *DIS2677*, *DIS3175*, *DIS2797*, *DIS2874*, *DIS2824*, *DIS197* and *DIS427* (Table 2). PCR primer sequences were designed using Primer Blast software (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Primer sequences and conditions are available on request. PCR product size determination was evaluated by capillary electrophoresis on an ABI3100 Sequencer using Genescan 2.1 software (Applied Biosystems, Foster city, CA).

Haplotyping and estimate of mutation age

Haplotypes including microsatellites *DIS2861*, *DIS211*, *DIS421*, *DIS451*, *DIS2677*, *DIS3175*, *DIS2797* and *DIS2874*

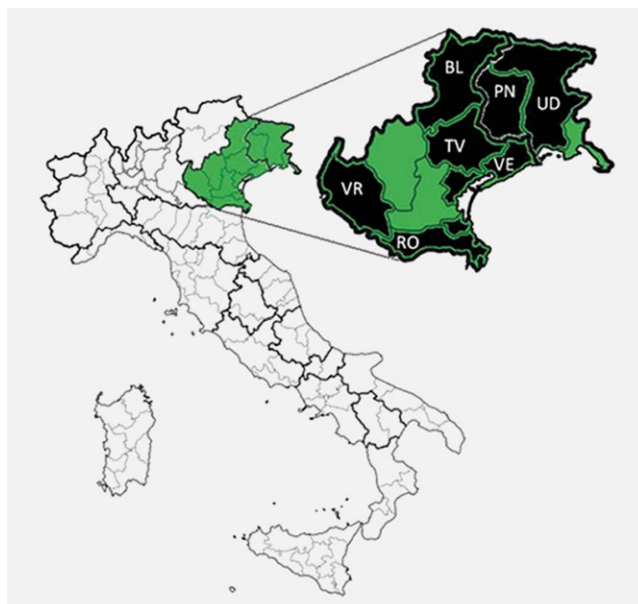


Figure 1. Map of Italy showing the geographic origin (areas in black) of the families segregating the c.933+3A>C mutation in North-Eastern Italy. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 2. Microsatellite alleles in the *MUTYH* region most frequently observed in Italian MAP cases and controls

	D1S2861 ¹	D1S211	D1S211	D1S421	D1S451	D1S2677	MUTYH	D1S3175	D1S2797	D1S2874	D1S2824	D1S197	D1S427
	43106861bp	44231355bp	45256821bp	45757240bp	45792083bp	45794914bp	46017631bp	46933592bp	47908231bp	48762721bp	50750509bp	51245558bp	
Allele size range (bp)	191–207	164–190	142–150	246–262	159–179	153–165	142–174	228–250	232–254	101–131	241–265		
Most frequent alleles in cases	197 199	180	148	246	167	157	164	230	238	117	243		
Cases Observed (Total number tested)	8 8 (22)	7 (22)	20 (22)	10 (22)	20 (22)	21 (22)	10 (22)	8 (22)	9 (22)	7 (22)	12 (22)		
Controls Observed (Total number tested)	51 16 (108)	6 (109)	69 (112)	0 (107)	52 (116)	68 (116)	37 (116)	45 (116)	49 (116)	29 (116)	59 (116)		
<i>p</i> value ² (fisher's test)	0.49 0.05	0.003	0.007	<0.0001	<0.0001	<0.0001	0.23	1	1	0.7	0.9		

¹Physical map positions (bp) are reported under each *locus*.

²Two alleles were found at equal frequencies in the case population for this marker. ³*p* value < 0.01 (in bold type) was considered as cut-off for statistical significance.

were manually constructed to minimize the number of recombinations both in Italian and German families.

The DMLE+2.2 software developed by Rannala and Reeve²⁷ was used to estimate the age of c.933+3A>C in North-Eastern Italy. The program, freely available at URL: <http://www.dmlc.org>, uses a Bayesian approach to compare differences in linkage disequilibrium between the mutation and flanking markers in DNA samples from mutation carriers and controls. The software generates the marginal posterior probability density of mutation age based on the following parameters: (a) observed haplotypes or genotypes in normal and affected chromosomes; (b) map distances between markers and mutation site; (c) population growth rates and (d) an estimated proportion of the mutation bearing chromosomes sampled.

Map distances were estimated on the basis of the Marshfield genetic map (URL:<http://www.ncbi.nlm.nih.gov/mapview/>) or inferred on the basis of physical distances (in Mb) given in UCSC Genome Browser (GRCh37/hg19 assembly) considering 1 Mb = 1cM.

The population growth rate (^{gen}r) in Northern-eastern Italy was estimated with the following formula: $^{gen}r = \frac{\log \frac{P_t}{P_0}}{g}$ in which ^{gen}r is the population growth rate per generation, P_t is the estimated present population size, P_0 is the estimated size of the population at reference time and g is the number of generations between these two time points. The total population of the Veneto region and of the provinces of Pordenone and Udine in Friuli currently comprises 5,737,630 people (<http://demo.istat.it/>). Historical and demographic data indicate that about 800,000 people lived in this area in year 1200.²⁸ Accordingly, the average ^{gen}r of this population was estimated to be 0.0599 from 1200 to the present time, assuming 25 years/generation. The latter value was used for mutation age estimates.

The proportion of mutation bearing chromosomes sampled was estimated to be 0.0677 and was calculated considering a population of 3,688,558 inhabitants living in the area of the North-Eastern Italy from which the patients originated (provinces of Udine, Pordenone, Venezia, Rovigo, Verona, Treviso and Belluno), and assuming that the prevalence of MAP (clinical plus subclinical biallelic carriers) is about 1:10,000 among Europeans.²⁹

MUTYH mRNA analysis

Totally, 2–3 x 10⁶ lymphoblastoid cells from five patients and two controls were pelleted and RNA was extracted with the EZ1 QIAgen RNA extraction Mini-kit (QIAgen, Germantown, MD). Genomic DNA contamination was avoided by the addition of DNase during the extraction phase. Reverse transcription was performed using the Reverse Transcription System kit (Promega, Madison, WI) on 1 µg of total RNA.

Quantitative analysis of *MUTYH* transcripts was performed on cDNA by real-time duplex PCR, with home-made and predeveloped TaqMan assays (Applied Biosystems) on a

Biorad CFX 96 thermal cycler, using the TaqMan Gene Expression Master Mix (Applied Biosystems). Three different *MUTYH* TaqMan assays were designed and used for relative quantification. *MUTYH* wild-type and mutated signals were detected using two FAM-TAMRA probes, specific for exon 9-exon 10 and exon 9-exon 11 junctions, respectively. The total amount of *MUTYH* mRNA (nuclear and mitochondrial forms, normal and mutated) was detected with a predeveloped TaqMan Gene Expression Assay (Applied Biosystems) containing a FAM-MGB probe directed at the junction between exons 5 and 6. Each *MUTYH* assay signal was normalized by quantitation of housekeeping genes in duplex condition. Two different housekeeping genes (β -actin and β -tubulin) were analyzed through predeveloped assays containing a VIC-MGB probe (Applied Biosystems). Quantitative evaluation was carried out by comparing the mutated samples with the two WT controls by the $\Delta\Delta C_t$ method.

Western blotting

Lymphoblastoid cells (5×10^6) were centrifuged to eliminate the culture medium and washed once with PBS 1X. Pelleted cells were then resuspended in lysis buffer (Tris-HCl 20 mM, NaCl 100 mM, MgCl₂ 5 mM, EDTA 0.2 mM, NP40 0.1%) containing protease and phosphatase inhibitors (Sigma, St.Luis, MO) and incubated 30 min on ice. Lysed cells were then centrifuged in a microfuge and the supernatant, containing whole cell protein extract, was collected in a new vial. Protein concentration was determined by Bradford assay (BioRad, Hercules, CA) and 50 μ g of whole cell extracts were analyzed by SDS-PAGE and western blotting. The primary antibody used was a mouse monoclonal anti-MUTYH (Ab 55551; Abcam, Cambridge, UK) specific for a C-terminal protein epitope (aa 435-535). The signal was amplified by an HRP-conjugated secondary antibody and revealed by Enhanced ChemiLuminescence (ECL; GE Healthcare, Little Chalfont, UK). *MUTYH* signals were normalized using the β -tubulin housekeeping protein (rabbit polyclonal anti-tubulin β , Santa Cruz Biotechnology, Santa Cruz, CA).

Results were confirmed in three independent experiments performed on LCL whole extracts produced in different extraction sessions.

Statistical analyses

Allele frequency distributions in wild type and Italian c.933+3A>C chromosomes were compared by Fisher's exact tests; $p < 0.01$ was considered as a cut-off for statistical significance.

Western blot and real-time PCR results are reported as mean \pm SD of signals obtained in three experiments. Comparisons of expression levels between patient and control samples were performed with a two-tailed paired Student's *t* test ($*p < 0.05$, $**p < 0.01$ and $***p < 0.001$).

Results

Haplotype analysis and age estimation

To verify the presence of a possible founder effect, allele and haplotype analyses were first performed on the 11 Italian families segregating the c.933+3A>C mutation and on 148 Italian control chromosomes (140 from the healthy control individuals and 8 from noncarrier relatives in MAP families). Statistically significant differences in allele frequencies between control and c.933+3A>C chromosomes were observed at *loci* *DIS211*, *DIS421*, *DIS451*, *DIS2677* and *DIS3175* (Table 2), the markers located closer to *MUTYH*, strongly suggesting a founder effect.

Haplotype analysis using eight microsatellite markers flanking the *MUTYH* locus was performed on probands and, when possible, on additional family members. Five families/individuals, including the c.933+3A>C homozygote, shared a common haplotype at *loci* *DIS421*, *DIS451*, *DIS2677* and *DIS3175* (148-246-167-157) spanning a region of ~ 0.8 Mb (Table 3). This common haplotype was not found in control chromosomes. In addition, although molecular analysis was not completely informative for *DIS451*, the same haplotype combination appeared to be the most likely in four additional families/individuals (FAP82, FAP294, FAP117, FAP443).

In family CFS167, the *DIS451* allele on the c.933+3A>C chromosome differs from that found in the other patients investigated. This finding is compatible with either a recombination event or a mutation that occurred in *DIS451* on the founder chromosome. Another possible recombination between markers *DIS2677* and *DIS451* was observed in proband FAP181; however, his haplotypes could not be defined with certainty because no other family members were available for analysis.

Mutation age estimate based on DMLE+2.2 software was ~ 83 generations (95% credible set: 57–137) (Fig. 2), using haplotype data from probands and controls.

In addition, haplotype analysis using eight microsatellites was performed on eleven German probands and three family members with the c.933+3A>C mutation. Although segregation could not be investigated in detail, the allele combinations observed in all probands, including two unrelated homozygotes, were compatible with the shared Italian haplotype at *loci* *DIS421*, *DIS451*, *DIS2677* and *DIS3175* (148-246-167-157) (Table 3).

MUTYH expression analyses

Analysis of splicing alteration showed that all samples with c.933+3A>C had exon 10 skipping and traces of additional transcripts retaining intron 11, with and without exon 10, which probably represent alternatively spliced isoforms (data not shown). The full-length and skipped transcripts were then quantified by real-time TaqMan PCR assays (Fig. 3a). Mean results of three different experiments showed that all c.933+3A>C positive samples expressed equal or slightly reduced amounts of total *MUTYH* transcripts (exons 5-6

Table 3. Haplotype analysis in the families segregating the c.933+3A>C *MUTYH* mutation

CASE ID ¹	D1S2861	D1S211	D1S421	D1S451	D1S2677	MUTYH	D1S3175	D1S2797	D1S2874
FAP349	197/199	180	148	246	167	c.933+3A>C	157	150/164	232
	197/199	184	148	246	167	c.933+3A>C	157	150/164	230
CFS443	199	180	148	246	167	c.933+3A>C	157	150	232
	197	184	148	252	167	c.1437delGGA	157	160	230
FAP352	195	180	148	246	167	c.933+3A>C	157	150	230
	199	174	148	252	159	WT	157	160	234
FAP278	197	180/184	148	246	167	c.933+3A>C	157	150/164	228/230
	195	180/184	148	252	167	c.1147delC	157	150/164	228/230
FAP426	197/199	166	148	246	167	c.933+3A>C	157	172	230/232
	197/199	168	146	252	167	c.536A>G	157	164	230/232
FAP 82	195	180/184	148	246/252	167	c.933+3A>C	157	164	228/234
	195	180/184	148	246/252	167	c.536A>G	157	164	228/234
FAP294	197/199	180/186	148	246/252	167	c.933+3A>C	157	150/164	230/232
	197/199	180/186	148	246/252	167	c.734G>A	157	150/164	230/232
FAP117	197/199	172/184	148	246/252	167	c.933+3A>C	157	164/174	228
	197/199	172/184	148	246/252	167	c.1437delGGA	157	164/174	228
FAP443	197/201	184	148	246/252	167	c.933+3A>C	157	166/172	234/236
	197/201	184	148	246/252	167	c.421delC	157	166/172	234/236
CFS167	197	164	148	248	167	c.933+3A>C	157	164	230
	199	170	148	252	167	c.1147delC	157	164	228
FAP181	199/201	180/182	148/146	250/252	167/159	c.933+3A>C	157/161	160/164	230/238
	199/201	180/182	148/146	250/252	167/159	c.1187G>A	157/161	160/164	230/238
16530	199	172	148	246	167	c.933+3A>C	157	172	236/238
	199	172	148	246	167	c.933+3A>C	157	172	236/238
25045	197	186	148	246	167	c.933+3A>C	157	172	238
	197	186	148	246	167	c.933+3A>C	157	172	238
354	197/203	168/186	148	246/252	167	c.933+3A>C	157	164/172	234/238
	197/203	168/186	148	246/252	167	c.1437delGGA	157	164/172	234/238
3345	197/199	170/186	148	246/252	167	c.933+3A>C	157	168/172	228/232
	197/199	170/186	148	246/252	167	c.536A>G	157	168/172	228/232
4017	199	168/186	148	246/252	167	c.933+3A>C	157/161	164/172	232/238
	199	168/186	148	246/252	167	c.536A>G	157/161	164/172	232/238
4022	197	174/184	148	246/252	167	c.933+3A>C	157	164/166	228/230
	197	174/184	148	246/252	167	c.536A>G	157	164/166	228/230
4704	195	170/186	148	246/252	167	c.933+3A>C	157	164/172	230/238
	195	170/186	148	246/252	167	c.536A>G	157	164/172	230/238
5537	197/199	172/184	148	246/252	167	c.933+3A>C	157	164/172	230/238
	197/199	172/184	148	246/252	167	c.536A>G	157	164/172	230/238
12287	197	184	148	246/252	167	c.933+3A>C	157/161	166/172	228/238
	197	184	148	246/252	167	c.536A>G	157/161	166/172	228/238
17203	197/199	172/184	148	246/252	167	c.933+3A>C	157	172	228/238
	197/199	172/184	148	246/252	167	c.536A>G	157	172	228/238
30457	199	166/184	148	246/252	167	c.933+3A>C	157	166	234
	199	166/184	148	246/252	167	c.536A>G	157	166	234

Numbers indicate the alleles observed on each chromosome of Italian (upper part) and German (lower part) families.

The “/” symbol indicates that the phase of the disease haplotype could not be established.

Alleles present in the shared haplotype are represented in bold type.

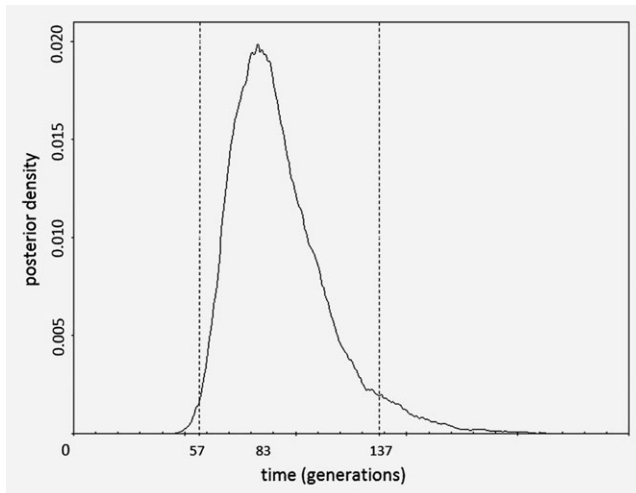


Figure 2. Posterior probability distribution plot of the c.933+3A>C mutation age (in generations), as estimated by DMLE+2.2 software ($g^{enr} = 0.0599$). The vertical broken lines indicate the 95% credible set of value for the calculation. Peak: 83 generations (95% credible set 57–137).

probe) compared to controls. In addition, based on the exon 9-10-specific assay, all mutated samples had a level of transcript containing exon 10 significantly lower compared to controls (<50%). Interestingly, we observed that the homozygote for c.933+3A>C produces traces of correctly spliced *MUTYH* transcript (about 10% of the total amount). To integrate data of the exon 9-10 assay, we also quantified c.933+3A>C transcripts lacking exon 10 using an exon 9-exon 11 probe. As expected, both WT samples were negative for this assay, whereas all LCLs from compound heterozygotes expressed the –exon 10 transcript, in amounts lower than 50% compared to the FAP349 homozygote.

FAP349 expressed traces of full-length protein, whereas compound heterozygous LCLs expressed variable amounts of *MUTYH*, ranging from 1 to 90% of the WT signal, depending on the nature of the second mutation and on the affinity of the primary antibody used that fails to recognize C-terminal truncated protein products (Fig. 3b).

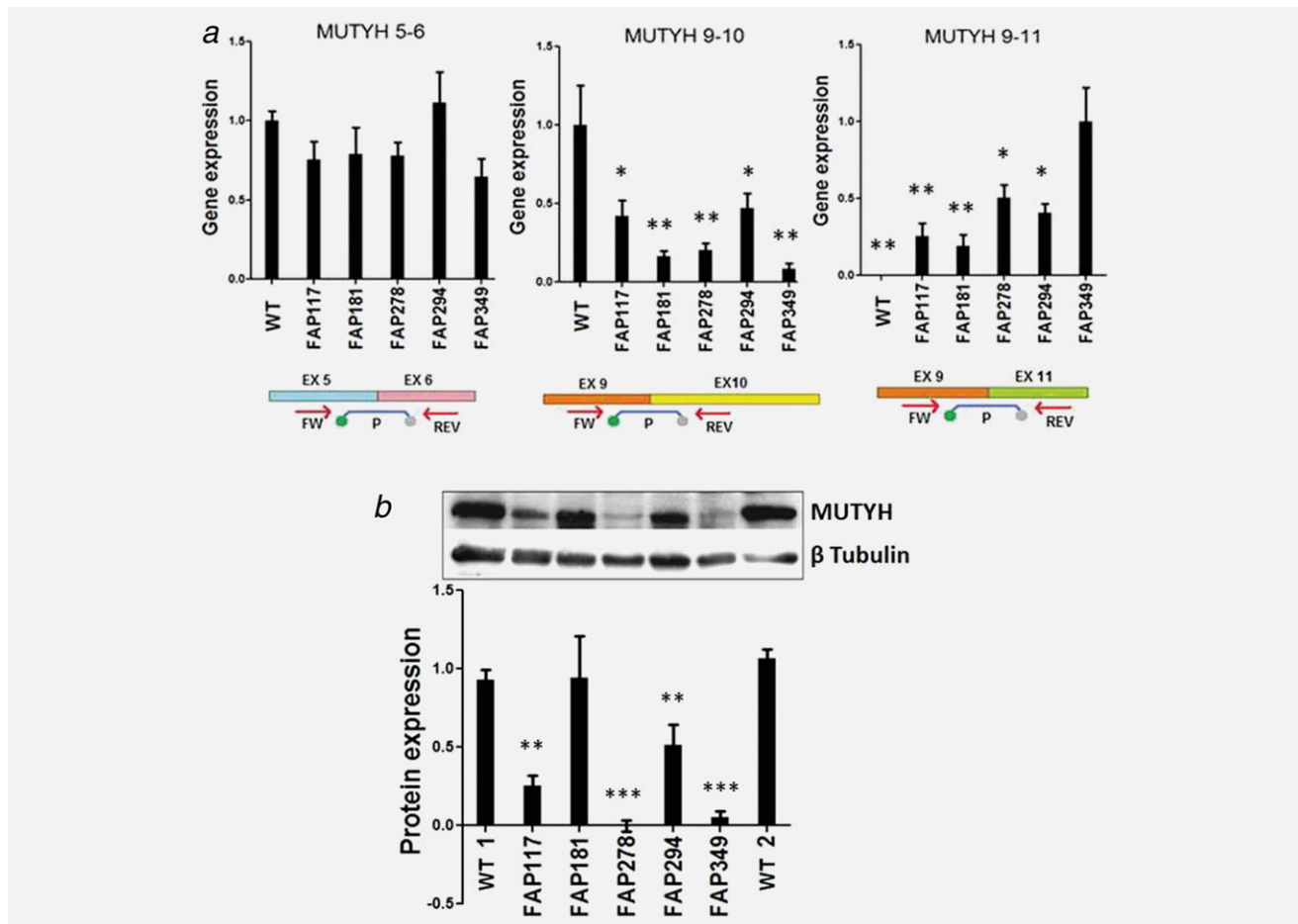


Figure 3. Quantitation of *MUTYH* gene and protein expression in LCLs (a) mRNA expression analysis by real-time PCR. Graphs report the mean ratio of *MUTYH* exon 5–6, 9–10 and 9–11 assay signals normalized by β -tubulin and β -actin genes (mean of three experiments with standard deviations). Control bars are WT (mean results obtained on two wild-type LCLs) for the 5–6 and 9–10 assays, FAP349 (c.933+3A>C homozygote) for the 9–11 assay. An assay scheme is reported below each graph (P, probe; FW, forward primer; REV, reverse primer). * $p < 0.05$ and ** $p < 0.01$ relative to controls. (b) Western blot analysis of *MUTYH* protein. Bars with standard deviations represent the mean ratio between *MUTYH* and β -tubulin for three experiments on each sample. ** $p < 0.01$ and *** $p < 0.001$ relative to WT controls. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Discussion

During the last 8 years the functional effects of different *MUTYH* mutations have been characterized,^{30–34} and the results of mutation analyses have revealed the existence of ethnic differences in the distribution of *MUTYH* variants.^{7,8} However, so far, founder effects have not been documented for *MUTYH* variants.

The c.933+3A>C mutation is frequently found in patients from the North-East of Italy. It involves the 3rd base of intron 10, altering the consensus sequence of the donor splice site. Skipping of exon 10 with insertion of a premature stop codon has been documented.²¹ The LOVD database reports the predicted effect on the protein as p.Gly264TrpfsX7.¹⁴ In our cohort of patients, it accounted for about 15% of all *MUTYH* mutations so far identified. c.933+3A>C is present in other populations, although its proportion among *MUTYH* mutations has been found to be lower (1–8%) in studies on Caucasian patients from North America and other European countries.^{7,9,15–26} In particular, it has been identified in several German and in a few Swiss patients (LOVD).

In this study, we identified a common haplotype spanning a region of approximately 0.8 Mb surrounding the *MUTYH* gene, in North-Eastern Italian and German families segregating the c.933+3A>C. According to the DMLE+2.2 software, the age of the *MUTYH* c.933+3A>C mutation was estimated as 83 generations (95% credible set: 57–137), corresponding to ~2,075 (95% credible set: 1,425–3,425) years. Therefore its origin in North-Eastern Italy should date back to when the Venetics (Venetian population) lived in this area. However, it is worth noting that the method used estimates the time of origin of the mutation and not the time from the most recent common ancestor in Italy.³⁵ In addition, the results obtained with the DMLE, like with other Bayesian methods for the determination of age of mutation origin or age of the most recent common ancestor, are influenced by the input parameters used, namely the $g^{en}r$ value. The $g^{en}r$ used in this study was calculated based on the most reliable historical estimates of population size that are available for North-Eastern Italy.

The evidence that the Italian and German c.933+3A>C chromosomes share the same haplotype suggests that this mutation has originated only once in the past, at least in Europe. From available data it cannot be determined if the c.933+3A>C originated in Italy and then spread to other areas where it is now found at appreciable frequencies, or alternatively, if it entered in the Italian population from somewhere else, possibly Germany, and then became frequent in Italy. This issue could possibly be solved when data on the c.933+3A>C frequency in different parts of Europe will become available.

Analyses of mRNA extracted from LCLs confirmed the presence of transcripts lacking exon 10 in 5 patients with the c.933+3A>C mutation (1 homozygote and 4 compound heterozygotes). This aberrant mRNA, if translated, determines the synthesis of a truncated protein lacking a nuclear localization signal and the binding site for APE1 (human Apur-

inic/Pyrimidinic Endonuclease) and PCNA (Proliferating Cells Nuclear Antigen). A *MUTYH* protein devoid of these functional domains most likely would lose nuclear localization and DNA repair activity. As previously described for c.934-2A>G,³⁰ the truncated *MUTYH* protein could accumulate in the cytoplasm.

A real-time quantitative assay for *MUTYH* transcripts was set up to compare transcript levels between patients and controls and to measure the proportion of altered transcript on the total amount of *MUTYH* mRNA produced. Results showed that the overall levels of *MUTYH* transcript were equal, or slightly lower, in all c.933+3A>C LCLs compared to WT controls. As expected, all c.933+3A>C samples had significantly lower levels of correctly spliced transcript in comparison with controls. Interestingly, also the FAP349 LCL homozygous for c.933+3A>C revealed traces of normal transcript. This finding indicates that a very small amount of transcript produced from mutant alleles can be correctly spliced, generating traces of full-length functional protein, as revealed by Western blot analysis. Quantitative real-time analysis confirmed the presence of significantly lower amounts of –exon 10 transcripts in heterozygotes compared to FAP349, and that WT LCLs did not produce the aberrantly spliced transcript.

Protein amounts produced by LCLs with the c.933+3A>C mutation ranged between 1 and 100% of WT LCLs. This wide variability was dependent on the effects of the second mutation. The homozygous mutant gave less than 10% of WT signal. When the second mutation event caused another frameshift with insertion of a premature stop codon (c.1147delC in FAP278), no significant signal was observed, according to the predicted loss of *MUTYH* C-terminal target epitope for the antibody used in this study. On the other hand, when the second mutation event determined loss (FAP117) or substitution (FAP181 and FAP294) of a single aminoacid, protein band intensities were variable. FAP117 and FAP294 express about 50% of the *MUTYH* protein levels observed in WT cells; likely, the c.1437delGGA (p.Glu480del) and c.734G>A (p.Arg245Hys) alleles present in these LCLs are able to drive the synthesis of almost normal protein amounts, indicating that these mutations do not grossly affect mRNA or protein stability. Instead, the compound heterozygote FAP181 containing c.1187G>A (p.Gly396Asp) expresses protein amounts equal to WT, suggesting that the produced protein is stable but not functional, in accordance with previous findings.³¹ Unfortunately, the lack of reliable antibodies directed against the N-terminus of *MUTYH* prevented us to investigate the expression and stability of the putatively truncated protein. However, if the c.933+3A>C encoded a truncated p.Gly264TrpfsX7 protein, this would likely not be functional due to the loss of nuclear localization and impaired binding to key components of DNA repair.

On the basis of these findings, subjects carrying a homozygous *MUTYH* c.933+3A>C mutation should have a reduced ability to correct nuclear DNA damages induced by

reactive oxygen species in their colon mucosa, and this would increase the risk of adenomas and CRC. The extent of risk in compound heterozygotes is likely to be dependent on the effects of the second mutation. Almost all c.933+3A>C carriers in our series had an attenuated form of polyposis, the most common phenotype in MAP patients.⁹ Interestingly, the homozygous FAP349 patient underwent hemicolectomy following a diagnosis of attenuated polyposis (less than 30 polyps) at 51 years, in absence of CRC. Conversely, CRC was diagnosed before 55 years of age in 7 of the other 11 compound heterozygotes. The observation that traces of full-length transcripts and proteins are expressed in c.933+3A>C LCLs suggest that the mutant allele maintains a minimal residual MUTYH function, thus producing an attenuated phenotype. On the other hand, it has been observed that mutations that completely inactivate the MUTYH protein, such as frameshift variants and the Caucasian p.Tyr179Cys mutation, are usually associated with a more severe phenotype.⁹

The molecular and clinical data obtained in this study and the demonstration of a founder effect for c.933+3A>C have implications for MUTYH genetic testing. To optimize MUTYH mutation analysis, a rapid screening for

c.933+3A>C and for the two most common Caucasian mutations (p.Gly396Asp and p.Tyr179Cys) could be undertaken as a first step for molecular analysis in polyposis patients of ascertained North-Eastern Italy ancestry. Since altogether these three mutations account for almost 60% of the MUTYH alterations detected in this area, a simple multiplex ad hoc test would allow to rapidly establish a genetic diagnosis of MAP in at least one third of patients with this condition. Moreover, this test could be also applied for genetic pre-screening of polyposis/CRC patients not fully complying with current criteria for referral to MUTYH genetic testing. In addition, the data here presented suggest that genotype knowledge may be useful for phenotype prediction and consequently to establish tailored surveillance. However, evaluation of larger series will be necessary to allow a better definition of genotype/phenotype correlations, in order to optimize surveillance in at risk individuals.

Acknowledgements

This study was partially supported by grants from ACC (Alleanza Contro il Cancro) – INTEF Project and Ministero della Salute – Rare Disease Project (to AV), and from Istituto Toscano Tumori (to MG) and Ente Cassa di Risparmio di Firenze (to FiorGen).

References

1. Tops CM, Wijnen JT, Hes FJ. Introduction to molecular and clinical genetics of colorectal cancer syndromes. *Best Pract Res Clin Gastroenterol* 2009;23:127–46.
2. Jo WS, Chung DC. Genetics of hereditary colorectal cancer. *Semin Oncol* 2004;32:11–23.
3. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C>T:A mutations in colorectal tumors. *Nat Genet* 2002;30:227–32.
4. Lipton L, Halford SE, Johnson V, et al. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res* 2003;63:7595–9.
5. Avezù A, Agostini M, Pucciarelli S, et al. The role of MYH gene in genetic predisposition to colorectal cancer: another piece of the puzzle. *Cancer Lett* 2008;268:308–13.
6. Cleary SP, Cotterchio M, Jenkins MA, et al. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. *Gastroenterology* 2009;136:1251–60.
7. Sampson JR, Dolwani S, Jones S, et al. Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet* 2003;362:39–41.
8. Poulsen ML, Bisgaard ML. MUTYH Associated Polyposis (MAP). *Curr Genomics* 2008;9:420–35.
9. Nielsen M, Joerink-van de Beld MC, Jones N, et al. Analysis of MUTYH genotypes and colorectal phenotypes in patients with MUTYH-associated polyposis. *Gastroenterology* 2009;136:471–6.
10. Miyaki M, Iijima T, Yamaguchi T, et al. Germline mutations of the MYH gene in Japanese patients with multiple colorectal adenomas. *Mutat Res* 2005;578:430–3.
11. Isidro G, Laranjeira F, Pires A, et al. Germline MUTYH (MYH) mutations in Portuguese individuals with multiple colorectal adenomas. *Hum Mutat* 2004;24:353–4.
12. Alhopuro P, Parker AR, Lehtonen R, et al. A novel functionally deficient MYH variant in individuals with colorectal adenomatous polyposis. *Hum Mutat* 2005;26:393.
13. Gismondi V, Meta M, Bonelli L, et al. Prevalence of the Y165C, G382D and 1395delGGA germline mutations of the MYH gene in Italian patients with adenomatous polyposis coli and colorectal adenomas. *Int J Cancer* 2004;109:680–4.
14. Out AA, Tops CM, Nielsen M, et al. Leiden open variation database of the MUTYH gene. *Hum Mutat* 2010;31:1205–15.
15. Croitoru ME, Cleary SP, Di Nicola N, et al. Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. *J Natl Cancer Inst* 2004;96:1631–4.
16. Eliason K, Hendrickson BC, Judkins T, et al. The potential for increased clinical sensitivity in genetic testing for polyposis colorectal cancer through the analysis of MYH mutations in North American patients. *J Med Genet* 2005;42:95–6.
17. Jenkins MA, Croitoru ME, Monga N, et al. Risk of colorectal cancer in monoallelic and biallelic carriers of MYH mutations: a population-based case-family study. *Cancer Epidemiol Biomarkers Prev* 2006;15:312–4.
18. Kairupan CF, Meldrum CJ, Crooks R, et al. Mutation analysis of the MYH gene in an Australian series of colorectal polyposis patients with or without germline APC mutations. *Int J Cancer* 2005;116:73–7.
19. Nielsen M, Franken PF, Reinards TH, et al. Multiplicity in polyp count and extracolonic manifestations in 40 Dutch patients with MYH associated polyposis coli (MAP). *J Med Genet* 2005;42:e54.
20. Aretz S, Uhlhaas S, Goergens H, et al. MUTYH-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. *Int J Cancer* 2006;119:807–14.
21. Kanter-Smoler G, Bjork J, Fritzell K, et al. Novel findings in Swedish patients with MYH-Associated Polyposis: mutation detection and clinical characterization. *Clin Gastroenterol Hepatol* 2006;4:499–506.
22. Bouguen G, Manfredi S, Blayau M, et al. Colorectal adenomatous polyposis associated with MYH mutations: genotype and phenotype characteristics. *Dis Colon Rectum* 2007;50:1612–17.
23. Olschwang S, Blanché H, de Moncuit C, et al. Similar colorectal cancer risk in patients with monoallelic and biallelic mutations in the MYH gene identified in a population with adenomatous polyposis. *Genet Test* 2007;11:315–20.
24. O'Shea AM, Cleary SP, Croitoru MA, et al. Pathological features of colorectal carcinomas in MYH-associated polyposis. *Histopathology* 2008;53:184–94.
25. Vogt S, Jones N, Christian D, et al. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. *Gastroenterology* 2009;137:1976–85.
26. Morak M, Laner A, Bacher U, et al. MUTYH-associated polyposis—variability of the clinical phenotype in patients with biallelic and monoallelic MUTYH mutations and report on novel mutations. *Clin Genet* 2010;78:353–63.
27. Rannala B, Reeve JP. High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *Am J Hum Genet* 2001;69:159–78.

28. Fornasin A, Zannini A. Crisi e ricostruzione demografica nel Seicento veneto. In: *La popolazione italiana nel Seicento*, Bologna: Clueb, 1999. 103–22.
29. Aretz S, Hes FJ. Clinical utility gene card for: MUTYH-associated polyposis (MAP), autosomal recessive colorectal adenomatous polyposis. *Eur J Hum Genet* 2010;18:doi: 10.1038/ejhg.2010.77. Epub 2010 May 26.
30. Tao H, Shinmura K, Hanaoka T, et al. A novel splice-site variant of the base excision repair gene MYH is associated with production of an aberrant mRNA transcript encoding a truncated MYH protein not localized in the nucleus. *Carcinogenesis* 2004;25: 1859–66.
31. Parker AR, Sieber OM, Shi C, et al. Cells with pathogenic biallelic mutations in the human MUTYH gene are defective in DNA damage binding and repair. *Carcinogenesis* 2005;26: 2010–8.
32. Ali M, Kim H, Cleary S, et al. Characterization of mutant MUTYH proteins associated with familial colorectal cancer. *Gastroenterology* 2008;135: 499–507.
33. Molatore S, Russo MT, D'Agostino VG, et al. MUTYH mutations associated with familial adenomatous polyposis: functional characterization by a mammalian cell-based assay. *Hum Mutat* 2010;31:159–66.
34. D'Agostino VG, Minoprio A, Torreri P, et al. Functional analysis of MUTYH mutated proteins associated with familial adenomatous polyposis. *DNA Repair* 2010;9:700–7.
35. Greenwood CM, Sun S, Veenstra J, et al. How old is this mutation?—a study of three Ashkenazi Jewish founder mutations. *BMC Genet* 2010;11:39.