Cellular and Molecular Biology ™ **55**, N°2, 102-110
DOI 10.1170/T859

ISSN 1165-158X
2009 Cell. Mol. Biol.™

VARIATIONS IN THE LENGTH OF POLY-C AND POLY-T TRACTS IN INTRON 3 OF THE HUMAN FERROCHELATASE GENE

J. BARMAN[1], X. SCHNEIDER-YIN[1], R. MAMET[2], N. SCHOENFELD [2,3], E. I. MINDER[1]

[1]Zentrallabor, Stadtspital Triemli, Birmensdorferstrasse 497, 8063 Zurich, Switzerland
Fax: +41 44 466 2744; E-mail: elisabeth.minder@triemli.stzh.ch
[2] National Laboratory for the Biochemical Diagnoses of Porphyrias, Rabin Medical Center, Beilinson Hospital, Petah-Tikva, Israel
[3] Department of Human Molecular Genetics and Biochemistry, The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
E.L. Minder : Zentrallabor, Stadtspital Triemli, Birmensdorferstrasse 497, 8063 Zurich, Switzerland Fax: +41 44 466 2744; Email:elisabeth.minder@triemli.stzh.ch

Received, March 30th 2009; Accepted May 5th, 2009; Published July 1st, 2009

Abstract – The third intron of human ferrochelatase (FECH) gene contains according to NCBI, a poly-C (11) and a poly-T (24) tracts which are located approximately 900 bp upstream from the known splice modulating SNP IVS3-48 c/t. Ferrochelatase catalyses the last step in heme biosynthesis and a deficiency of this enzyme results in the hereditary disorder of erythropoietic protopoprhyria (EPP). During the course of mutation analysis in the *FECH* gene among EPP patients, we observed variations in the length of the poly-C and poly-T tracts. To study these variations, we analyzed a total of 54 individuals of Swiss and Israeli origins. Among them, 37 were control subjects (23 individuals with the genotype t/t and 14 with the genotype c/t), 10 were unrelated EPP patients (genotype c/M) and 7 were unrelated asymptomatic mutation carriers (genotype t/M). The length of poly-C tract varied from 10 to 16, that of poly-T tract from 22 to 24 in the study cohort. Statistic analysis showed that the low-expressed FECH allele (IVS3-48c) is associated with poly-C12, C13 and C15 and poly-T22. In addition, the segregation of poly-C and poly-T tracts was studied in two Israeli EPP families. Instabilities, as seen by both insertion and deletion of one nucleotide between two generations, were observed only in the poly-T tract. The function of the poly-C and poly-T tracts are yet to be explored.

INTRODUCTION

Erythropoietic protoporphyria (EPP, OMIM#177000) is a hereditary disorder of the heme biosynthetic pathway. The affected enzyme ferrochelatase (FECH, EC 4.99.1.1) catalyzes the insertion of iron into protoporphyrin IX to form heme. As the result of ferrochelatase deficiency, a large amount of protoporphyrin accumulates in the body and is responsible for the clinical symptoms of cutaneous photosensitivity in EPP patients (1). In approximately 93% of the cases, EPP patients carry a c/M genotype i.e., a mutated FECH allele (M) *in trans* to a low-expressed allele. This low allele is characterized by a c-variant (c) at a polymorphic site in intron 3 (IVS3-48 c/t). The resulting enzyme activity in patients is ~35% of that of normal individuals. Individuals with a t/M genotype i.e., a combination of a normal IVS3-48t allele (t) with an M allele, show a 50% reduction in the enzyme activity, without clinical symptoms of EPP. Both t/t and c/t genotypes present neither biochemical nor clinical signs of EPP (4). About 4% of the EPP cases are recessive with one mutation on each of the two FECH alleles. Another 3% of the cases are the so called variant EPP in which deletions in the 5-aminolevulinate synthase 2 (ALAS2) gene were found (11).

The third intron of human ferrochelatase (FECH) gene contains a poly-C and a poly-T tract which are located approximately 900 bp upstream from IVS3-48. According to the sequence data from NCBI, the length of these tracts are 11 for poly-C and 24 for poly-T (accession number NT025028). During the course of mutation analysis in the *FECH* gene

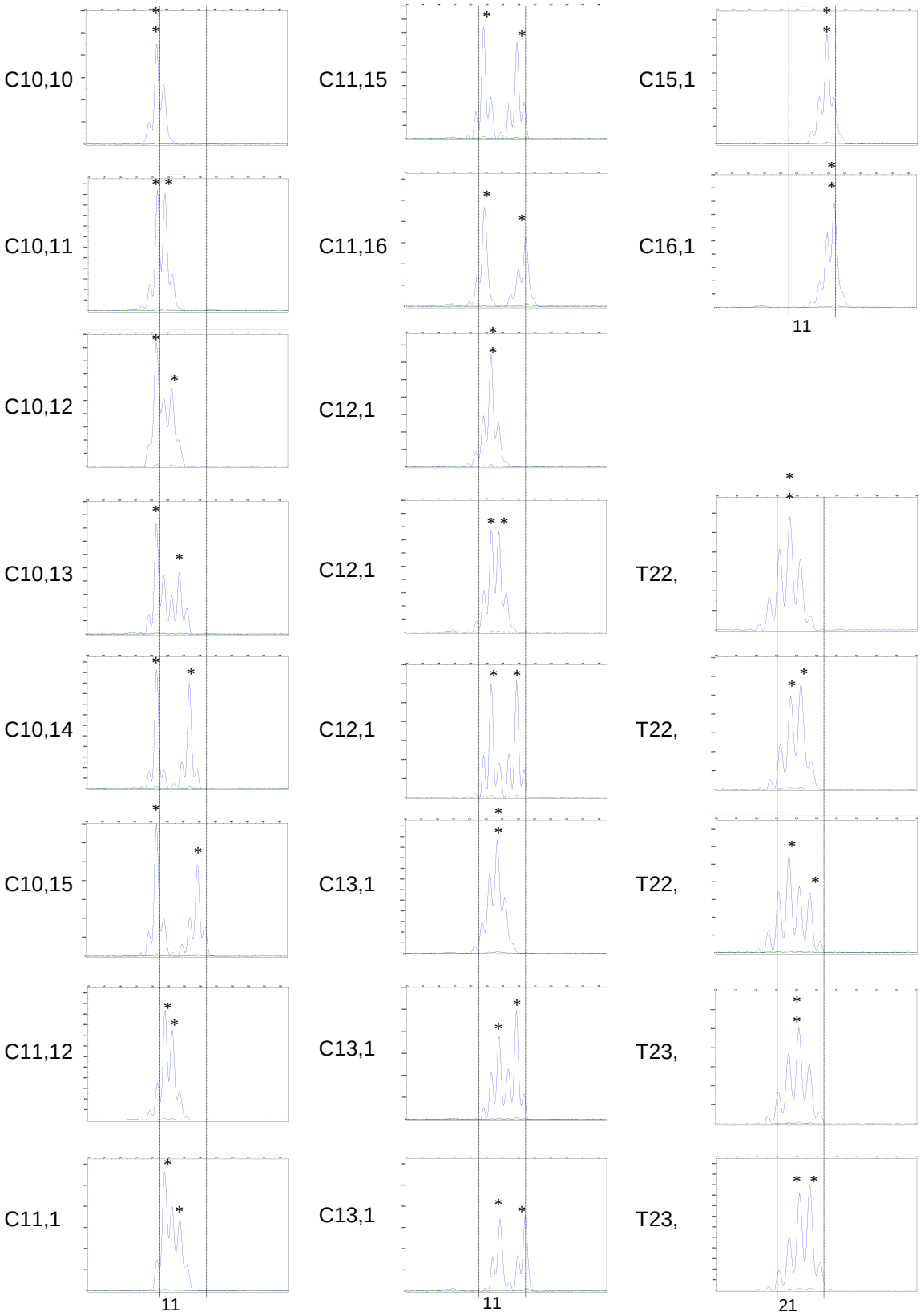


Figure 1. Examples of electropherograms showing various poly-C and poly-T alleles. The alleles are indicated by asterisks (*). The vertical lines align identical peaks among different electropherograms.

Table 1 Information on the study cohort of 54 unrelated individuals

Origin	Genotype t/t	t/M	c/t	c/M
Swiss	**9** (unrelated controls)	**5** (from unrelated EPP families)*	**5** (unrelated controls)	**10** (from unrelated EPP families)*
Israeli	**10** (unrelated controls) **4** (unrelated family members: individual 3 and 4 from *Family 1*; individual 3 and 7 from *Family 2*; see Fig.2)	**2** (unrelated family members: individual 2 from *Family 1*; individual 2 from *Family 2*)	**8** (unrelated controls) **1** (individual 1 from *Family* 1)	-
total	**23**	**7**	**14**	**10**

* The individuals are not related to each other.

Table 2 Distribution of poly-C alleles among EPP patients, carriers and control subjects (n= 54)

Length	Genotype t/t and t/M			Genotype c/t and c/M			calculated frequency (%)*	
of polyC	t/t	t/M	t/t & t/M	c/t	c/M	c/t & c/M	t	Diff c-t
C10	11	3	14	1	0	1	23.3	-21.3
C11	7	1	8	0	0	0	13.3	-13.3
C12	9	5	14	14	12	26	23.3	30.8
C13	7	2	9	6	6	12	15.0	10.0
C14	2	1	3	0	0	0	5.0	-5.0
C15	9	1	10	7	2	9	16.7	2.1
C16	1	1	2	0	0	0	3.3	-3.3
sum	**46**	**14**	**60**	**28**	**20**	**48**	**100**	**0**

* The frequency of t-associated alleles was based on the frequency observed in the combined M/t and t/t genotypes; the difference c-t was calculated from the observed frequency in the c/t and M/c genotype after subtraction of the frequency of the (t/t and M/t) genotype. For example, for the C10 allele, the calculated t allele frequency equals 14/60 x 100% = 23.3; Diff c-t = (1/48 x 100%) – 23.3 = -21.3. Positive numbers indicate over-representation and negative under representation of these alleles in the c/t genotype assuming an identical distribution of poly-C alleles in the t/t and c/t genotype.

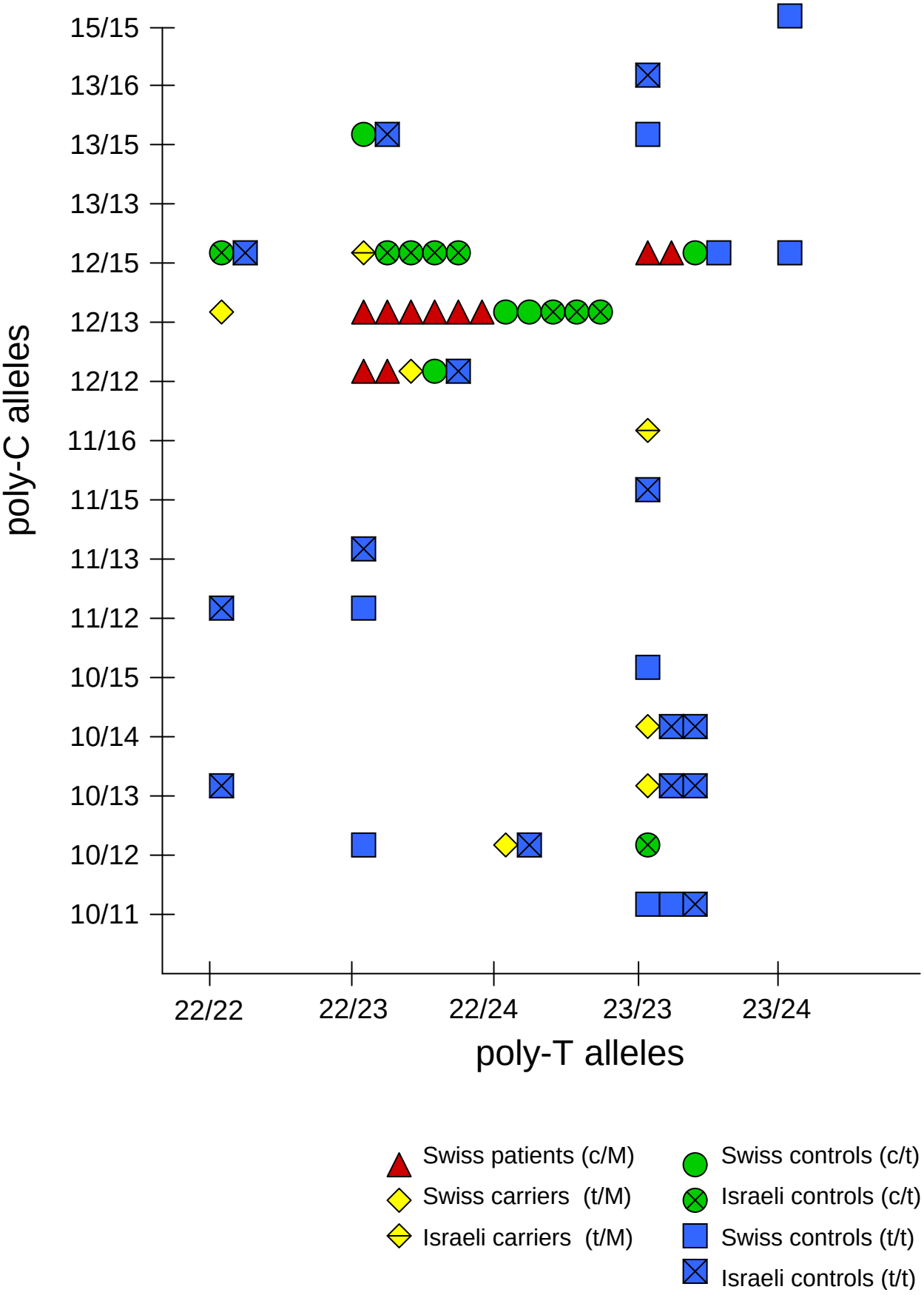


Figure 2. Genotypes of poly-C and poly-T tracts from 54 unrelated individuals of EPP patients, carriers and of control subjects.

among EPP patients, we observed variations in the length of the poly-C and poly-T tracts. In this study, we analyzed these variations among individuals with the four above-mentioned genotypes, t/t, c/t, t/M and c/M.

MATERIALS AND METHODS

Study cohort

A cohort of 54 unrelated individuals including EPP patients, carriers and control subjects was studied. The detailed information on the individuals included, of both Swiss and Israeli origins is given in Table 1. None of the individuals shared any alleles with another individual within this cohort. In addition, two previously published Israeli EPP families and 5 Swiss EPP families were analyzed (8,9, Fig.3, Table 4). All EPP patients and carriers had documented mutations in the *FECH* gene.

Analysis of poly-C and poly-T tracts

Genomic DNA from each individual was amplified by PCR using fluorescence labeled primers. For amplification of the poly-C tract, primers 5'-FAM-CCT TGC ACT CCC AGT TAT C-MGB-3' and 5'-TAC CTT CAC ATT TGT GTA ACG-3' were used. And for amplification of the poly-T tract, primers 5'-FAM-TCG TTA CAC AAA TGT GAA GGT-MGB-3' and 5'-GGA GGG ATG GCA TTA GGA-3' were used. The lengths of the PCR products were analyzed on the ABI Prism 310 genetic analyzer (Applied Biosystems, Foster City, California, USA). To determine the number of "C"s and "T"s in the poly-C and poly-T tracts, the electropherograms of probands were compared with that of "markers". These markers were generated by cloning of the non-fluorescent PCR products from a control subject into pBluescript. The number of "C"s or "T"s in individuals clones was determined by sequencing. Three poly-C clones, C10, C12 and C13, and three poly-T clones, T19, T21 and T22, were obtained. Purified plasmid DNA from the each of the six clones were amplified with fluorescence labeled primers before being analyzed on the ABI Prism 310 genetic analyzer.

Analysis of the IVS3-48c/t polymorphism was performed as previously described (9).

Statistical analysis

Non-identity of the distribution of the variable poly-C and poly-T tracts among the polymorphic IVS3-48 alleles was tested by Chi-square test including Monte Carlo simulation of the multinomial sampling distribution (http://faculty.vassar.edu/lowry/VassarStats.html).

RESULTS

Identification of different poly-C and poly-T genotypes

The analysis of the 54 unrelated individuals and members of the seven Israeli and Swiss EPP families, identified a total of 18 different poly-C and 5 different poly-T genotypes. As shown in Figure 1, the electropherogram of each genotype consists of one or two major peaks that are surrounded by a number of stutter peaks. Due to overlapping between major and stutter peaks, the assignment of individual alleles in each electropherogram was only possible with the help of "markers". As described in the Materials and Methods section, these "markers" were made from poly-C and poly-T clones with a known number of the respective nucleotides.

Distribution of poly-C and poly-T among genotypes t/t, M/t, c/t and c/M

The 54 unrelated individuals carried different genotypes with respect to IVS-48 i.e., 23 individuals with the t/t genotype, 7 with the t/M genotype, 14 with the c/t genotype and 10 with c/M genotype (Table 1). Among them, all individuals with the M/t and c/M genotypes were selected from different EPP families.

The poly-C and poly-T genotypes of the 54 unrelated individuals were assigned according to the electropherograms shown in Figure 1. Among the 108 alleles of this cohort, seven different poly-C alleles, C10 to C16, and three different poly-T alleles, T22, T23 and T24, were identified. The allelic distribution of poly-C and poly-T among the four IVS-48 genotypes are shown in Table 2 and Table 3, respectively.

To obtain an overview on the genotype distribution among the 54 individuals, the poly-C genotype (a combination of two poly-C alleles) from each individual was plotted against the poly-T genotype (a combination of two poly-T alleles) (Figure 2). Except for C10/10 and C16/16, the remaining 16 poly-C genotypes (see Figure 1) were found among the 54 individuals. All five poly-T genotypes (see Figure 1) were observed among the 54 individuals. Variations in poly-C among EPP patients (genotype c/M, n= 10) differ significantly from the control group (genotype t/t, n=23; $p<0.001$). However, no statistically significant difference was found in poly T ($p=0.23$).

Poly-C and poly-T in the low-expressed FECH allele

To study the relationship between low-expressed *FECH* allele (IVS3-48c) and poly-C/poly-T tracts, the results in Table 2 and 3 were analyzed by chi-square test separately. Since both genotype t/M and t/t are in fact homozygous t/t at IVS3-48, the poly-C/poly-T results from these two genotypes were combined (see column "t/t & t/M" in Table 2 and 3). Likewise, c/t individuals were combined with c/M individuals (see column "c/t & c/M" in Table 2 and 3). Chi-square test was then performed in these data sets.

Table 3 Distribution of poly-T alleles among EPP patients, carriers and control subjects (n=54)

Length	Genotype t/t and t/M			Genotype c/t and c/M			calculated frequency (%)*	
of polyT	t/t	t/M	t/t & t/M	c/t	c/M	c/t & c/M	t	Diff c-t
T22	12	5	17	13	8	21	28.3	15.4
T23	31	7	38	15	12	27	63.3	-7.1
T24	3	2	5	0	0	0	8.3	-8.3
sum	**46**	**14**	**60**	**28**	**20**	**48**	**100**	**0**

*see table 2.

Table 4 Haplotypes of the IVS3-48c allele in seven Swiss and Israeli EPP families

Haplo-types	EPP families							
	A	B	C	Israeli 2(1)	Israeli 1	D	E	Israeli 2(2)
poly-C	12	12	12	13	12	12	13	13
poly-T	22	22	22	22	23	23	23	23
IVS3-48	c	c	c	c	c	c	c	c

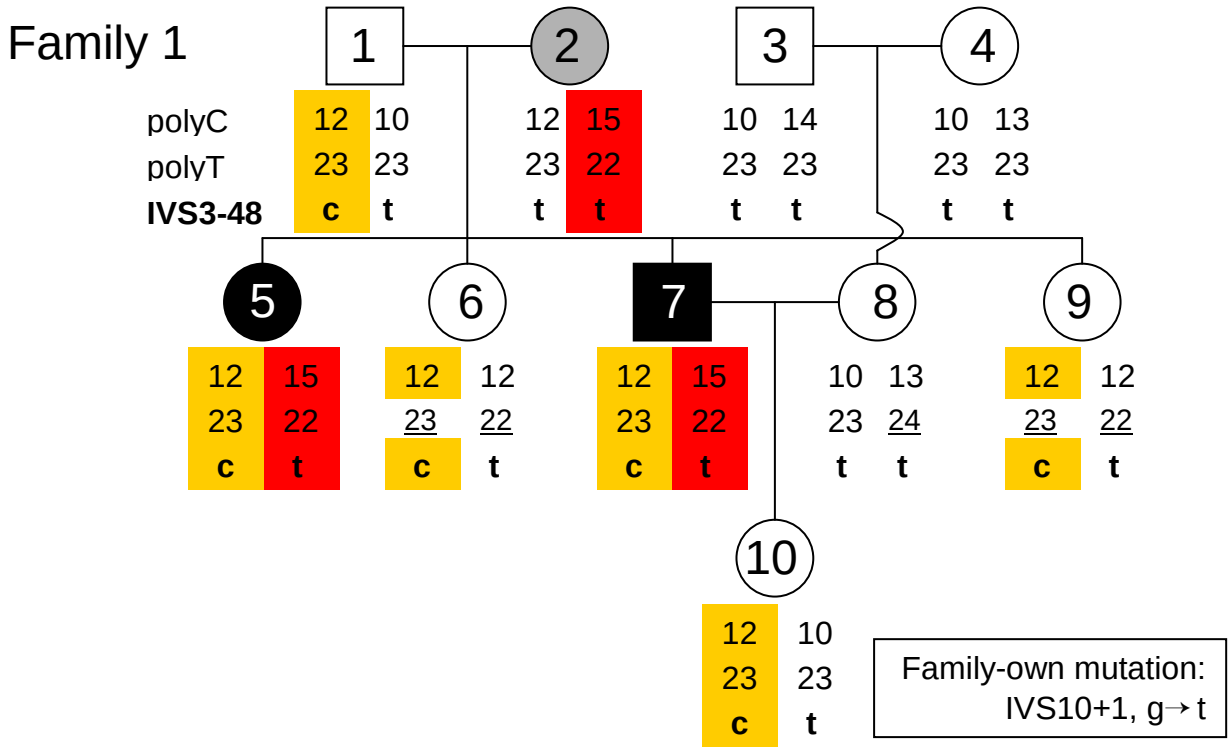
Swiss families: A, B, C, D and E (one member of families A and E with genotype t/M are included in Table 1; one member of families B, C and D with genotype c/M are included in Table 1);
Israeli 1: *Family 1* in Figure 2;
Israeli 2 (1): the first "c" allele in individual 1 of *Family 2* (see Figure 2);
Israeli 2 (2): the second "c" allele in individual 1 of *Family 2* (see Figure 2).

The results showed that the distribution of poly-C alleles was significantly different between the t/t genotype (including M/t) and the c/t genotype (including (M/c) (p<0.0001). The calculated frequency of different poly-C alleles based on the frequency observed in the t/t genotypes showed that in the c/t genotype, C12, C13 and C15 were more frequent and the other alleles were less frequent than expected (Table 2). With the same approach, T22 was found to be associated with the IVS3-48c allele (c) (p<0.0149; Table 3).

To verify the results of chi-square test, haplotyping analysis was performed among the seven Israeli and Swiss EPP families. Eight different "poly-C/poly-T/IVS3-48c" haplotypes were resolved including two from Israeli *Family 1* (Table 4, Figure 3). Among these eight haplotypes, C12 and C13 appeared 5 and 3 times, respectively. This result is consistent with that of chi-square test. On the other hand, T22 and T23 both appeared 4 times among the eight haplotypes, although the chi-square test only indicated an association between T22 and the low-expressed allele (c).

DISCUSSION

PCR amplification of mononucleotide repeat artificially introduces insertions and/or deletions within the repeat which give rise to a mixture of products in different lengths (2). As we experienced, sequencing of such a mixture of PCR products does not produce any informative results. To overcome this analytical difficulty, we amplified the poly-C and poly-T by PCR using fluorescently labeled primers and analyzed the products by capillary electrophoresis. However, the PCR products - a mixture as they remain, appear as complex patterns in the electropherograms. Each of the electropherogram contains one or two major peaks representing homo- or heteroallele in diploids such as human. The identification of heteroalleles (major peaks) could be hampered by the presence of stutter peaks (PCR artifacts). With respect to the poly-C tract, heteroalleles can be readily recognized if the length difference between the two alleles is $\geq$ 3 nucleotides (examples are C10,13, C10,14, C10,15, C11,15, C11,16, C12,15, C13,16 in Figure 1). With the help of "markers", we were



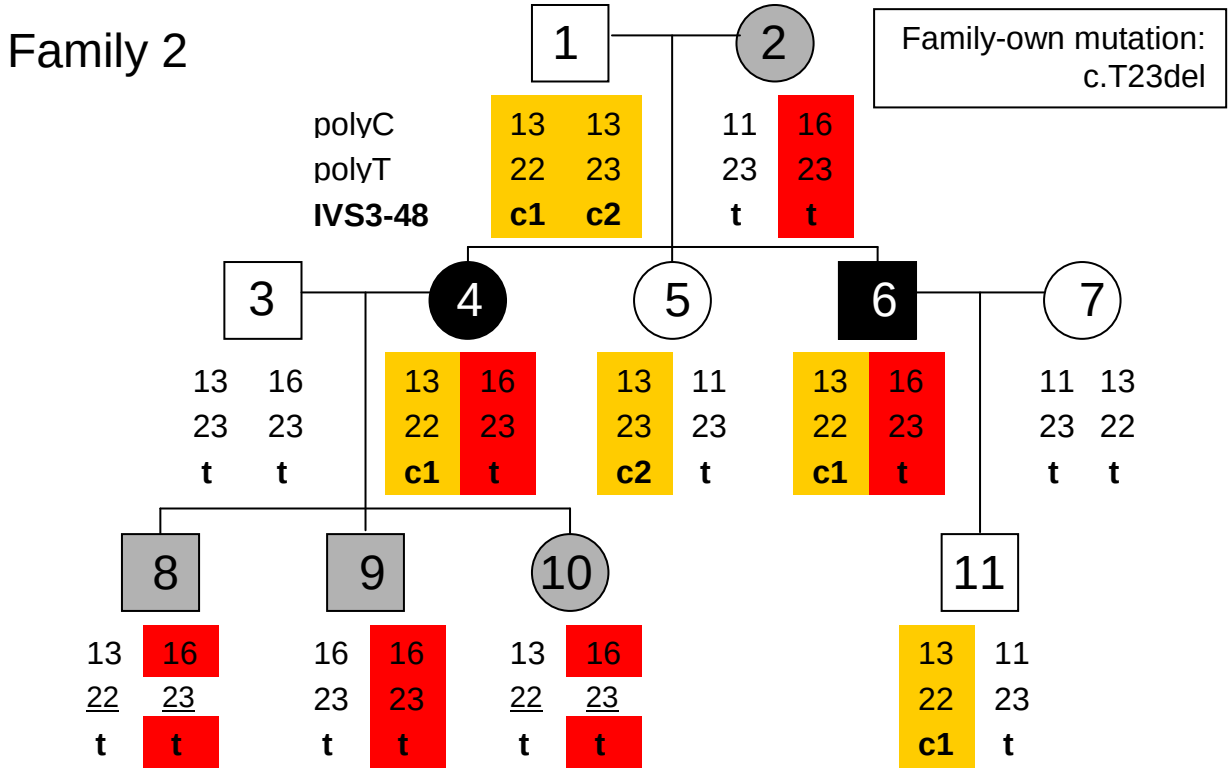


Figure 3. Segregation of poly-C, poly-T and IVS3-48c/t in two Israeli EPP families. The mutated and low-expressed FECH alleles are highlighted in *red* and *yellow*, respectively. Changes, either insertion or deletion, in the length of poly-T tract between two generations are underlined. In *Family 2*, two different IVS3-48 c alleles are represented by *c1* and *c2*, respectively.

able to assign the remaining homo- and heteroalleles.

In the poly-T tract, a difference of only one nucleotide between two alleles has been observed in the study subjects. The identification of different genotypes was made possible with the help of "markers" (Figure 1). Another observation on the poly-T tract is its instability between two generations. As shown in Figure 3, both deletion and insertion of a single T occurred in the two Israeli families (underlined numbers).

In a recent publication by Thompson and Salipante, the authors demonstrated the use of a software, named PeakSeeker, in genotyping of mononucleotide tracts (10). A computer algorithm such as PeakSeeker, could facilitate the interpretation of genotypes of mononucleotide repeats and thereby permits a more widely use of this type of highly informative markers in genetic research.

Based on both haplotypic and phylogenic analyses, Gouya et al came to a conclusion that the low-expressed IVS3-48c allele (c) arose from a single recent mutational event that occurred on the normal *FECH* allele IVS3-48t (t) some 40,000 to 60,000 years ago (4). The finding of this study supported the notion of IVS3-48c being a younger haplotype compared to IVS3-48t. As visualized in Figure 2, genotypes encompassing an IVS3-48c allele i.e., c/M and c/t shown in red and green symbols, respectively, are located closely to each other. In contrast, the blue and yellow symbols representing genotypes t/t and t/M, respectively, are more widely spread. The genotype distribution, as well as the result of the statistical analysis, suggest that the IVS3-48c haplotype is associated with fewer variations in the poly-C and poly-T tract compared to the IVS3-48t haplotype. The fewer variations were compatible with less meiotic or replicative modifications due to the younger age. The fewer variations could be further interpreted by the fact that the poly-C and poly-T tracts are located in a close vicinity to IVS3-48. According to a study of Gouya et al, the IVS3-48c variant is located within an extended haplotype [GGTA] spanning from the 5' noncoding region (-3670A/G) to intron 4 (IVS4-1197C/A) (5). The finding of the associated poly-C and poly-T alleles fit within this haplotype.

Deletions in the poly-T tract resulting in aberrant splicing, have been observed in cancer related genes such as *ATM* and *MRE 11* (3,7). The instability of poly-T tract has therefore been suggested to play a role in the development of cancer. In general, homopolymer tracts are significantly over-represented and more widely spread in the genome than simply by chance (6,12). However, any possible functions of the poly-C and poly-T tracts of the *FECH* gene in the low expression mechanism remain to be studied.

REFERENCES

1. Anderson K.E., Sassa S., Bishop D., Desnick R.J., Disorders of heme biosynthesis: X-linked sideroblastic anemia and the porphyrias. In Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The Metabolic and Molecular Basis of Inherited Disease* Ed 8, McGraw-Hill, New York, NY, 2991-3062, 2001.

2. Clarke L.A., Rebelo C.S., Gonçalves J., Boavida M.G., Jordan P., PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. *Mol. Pathol*. 2001, **54**: 351-353.

3. Ejima Y., Yang L., Sasaki M.S., Aberrant splicing of the ATM gene associated with shortening of the intronic mononucleotide tract in human colon tumor cell lines: a novel mutation target of microsatellite instability. *Int. J. Cancer* 2000, **86**: 262-268.

4. Gouya L., Martin-Schmitt C., Robreau A.M., Austerlitz F., Da Silva V., Brun P., Simonin S., Lyoumi S., Grandchamp B., Beaumont C., Puy H., Deybach J.C., Contribution of a common single-nucleotide polymorphism to the genetic predisposition for erythropoietic protoporphyria. *Am. J. Hum. Genet.* 2006, **78**: 2-14.

5. Gouya L., Puy H., Robreau A.M., Bourgeois M., Lamoril J., Da Silva V., Grandchamp B., Deybach J.C., The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype FECH. *Nat, Genet*. 2002, **30**: 27-28.

6. Kashi Y., King D.G., Simple sequence repeats as advantageous mutators in evolution. *Trends Genet*. 2006, **22**: 253-259.

7. Ottini L., Falchetti M., Saieva C., De Marco M., Masala G., Zanna I., Paglierani M., Giannini G., Gulino A., Nesi G., Mariani Costantini R., Palli D., MRE11 expression is impaired in gastric cancer with microsatellite instability. *Carcinogenesis*. 2004, **25**: 2337-2343.

8. Schneider-Yin X., Mamet R., Minder E.I., Schoenfeld N., Biochemical and molecular diagnosis of erythropoietic protoporphyria in an Ashkenazi Jewish family. *J. Inherit. Metab. Dis.* 2008 Aug 31. [Epub ahead of print]

9. Schoenfeld N., Mamet R., Minder E.I., Schneider-Yin X., A "null allele" mutation is responsible for erythropoietic protoporphyria in an Israeli patient who underwent liver transplantation: relationships among biochemical, clinical, and genetic parameters. *Blood Cells Mol. Dis.* 2003, **30**: 298-301.

10. Thompson J.M., Salipante S.J., PeakSeeker: a program for interpreting genotypes of mononucleotide repeats. *BMC Res. Notes*. 2009, Feb 3;2(1):17. [Epub ahead of print]

11. Whatley SD, Ducamp S, Gouya L, Grandchamp B, Beaumont C, Badminton MN, Elder GH, Holme SA, Anstey AV, Parker M, Corrigall AV, Meissner PN, Hift RJ, Marsden JT, Ma Y, Mieli-Vergani G, Deybach JC, Puy H, C-terminal deletions in the ALAS2 gene lead to gain of function and cause X-linked dominant protoporphyria without anemia or iron overload. *Am. J. Hum. Genet.* 2008, **83**: 408-14.

12. Zhou Y., Bizzaro J.W., Marx K.A., Homopolymer tract length dependent enrichments in functional regions of 27 eukaryotes and their novel dependence on the organism DNA (G+C)% composition. *BMC Genomics*. 2004, **5**: 95.