ARTICLE | Genetics inMedicine

Targeted resequencing of *FECH* locus reveals that a novel deep intronic pathogenic variant and eQTLs may cause erythropoietic protoporphyria (EPP) through a methylation-dependent mechanism

Matteo Chiara, PhD[1], Ilaria Primon, BSc[2], Letizia Tarantini, BSc[3], Luca Agnelli, PhD[4], Valentina Brancaleoni, MSc[2], Francesca Granata, MSc[2], Valentina Bollati, PhD[3] and Elena Di Pierro, PhD[2]

Purpose: Existing data do not explain the reason why some individuals homozygous for the hypomorphic *FECH* allele develop erythropoietic protoporphyria (EPP) while the majority are completely asymptomatic. This study aims to identify novel possible genetic variants contributing to this variable phenotype.

Methods: High-throughput resequencing of the *FECH* gene, qualitative analysis of RNA, and quantitative DNA methylation examination were performed on a cohort of 72 subjects.

Results: A novel deep intronic variant was found in four homozygous carriers developing a clinically overt disease. We demonstrate that this genetic variant leads to the insertion of a pseudo-exon containing a stop codon in the mature *FECH* transcript by the abolition of an exonic splicing silencer site and the concurrent institution of a new methylated CpG dinucleotide. Moreover, we show that the hypomorphic *FECH* allele is linked to a single haplotype of about 20 kb in size that encompasses three noncoding variants that were previously associated with expression quantitative trait loci (eQTLs).

Conclusion: This study confirms that intronic variants could explain the variability in the clinical manifestations of EPP. Moreover, it supports the hypothesis that the control of the *FECH* gene expression can be mediated through a methylation-dependent modulation of the precursor messenger RNA (pre-mRNA) splicing pattern.

Genetics in Medicine (2020) 22:35–43; https://doi.org/10.1038/s41436-019-0584-0

Keywords: erythropoietic protoporphyria; deep intronic pathogenic variant; eQTLs; CpG sites; pre-mRNA splicing pattern

INTRODUCTION

Erythropoietic protoporphyria (EPP, MIM 177000) is a heritable metabolic disorder resulting from a reduction, to less than 35% of normal levels, of ferrochelatase (FECH, EC 4.99.1.1) activity.[1] FECH is the last enzyme of the heme biosynthetic pathway and reduced activity leads to significantly elevated metal-free protoporphyrin (PPIX) levels mainly in erythrocytes and subsequently in skin and liver, causing clinical manifestations of the disease.[2] EPP patients experience severe cutaneous phototoxic reactions in sun-exposed areas beginning in early childhood. Besides this, the accumulation of PPIX in the liver may lead to mild hepatic injury and, in approximately 2% of cases, to severe cholestatic damage progressing to liver failure and requiring consequent liver transplantation.[3]

Clinical manifestation of EPP is typically associated with compound heterozygous variants at the *FECH* locus, mapped to chromosome 18q21.3. Currently about 200 pathogenic variants are reported in the Human Gene Mutation Database (HGMD Professional v2018.3; http://www.hgmd.cf.ac.uk/ac/index.php). Although a limited number of patients carrying two null *FECH* alleles have been reported,[4] in the majority of the affected individuals, a rare null *FECH* allele is coinherited in *trans* with a common hypomorphic allele that is associated with decreased levels of *FECH* gene expression.[5,6] This hypomorphic *FECH* allele is characterized by the presence of a common variant in intron 3 (c.315–48 T>**C**, **C** variant, rs2272783), which modulates the usage of a cryptic alternative acceptor splice site, 63 bp upstream of the constitutive site

[1]Dipartimento di Bioscienze, Università degli Studi di Milano, Milan, Italy; [2]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, UOC Medicina Generale, Milan, Italy; [3]EPIGET–Epidemiology, Epigenetics and Toxicology Lab Department of Clinical Sciences and Community Health, Università degli Studi di Milano, Milan, Italy; [4]Department of Oncology and Hemato-oncology, University of Milan; and Hematology 1, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. Correspondence: Elena Di Pierro (elena.dipierro@policlinico.mi.it, elena.dipierro@unimi.it)
These authors contributed equally: Matteo Chiara and Elena Di Pierro

Submitted 14 January 2019; accepted: 5 June 2019
Published online: 5 July 2019

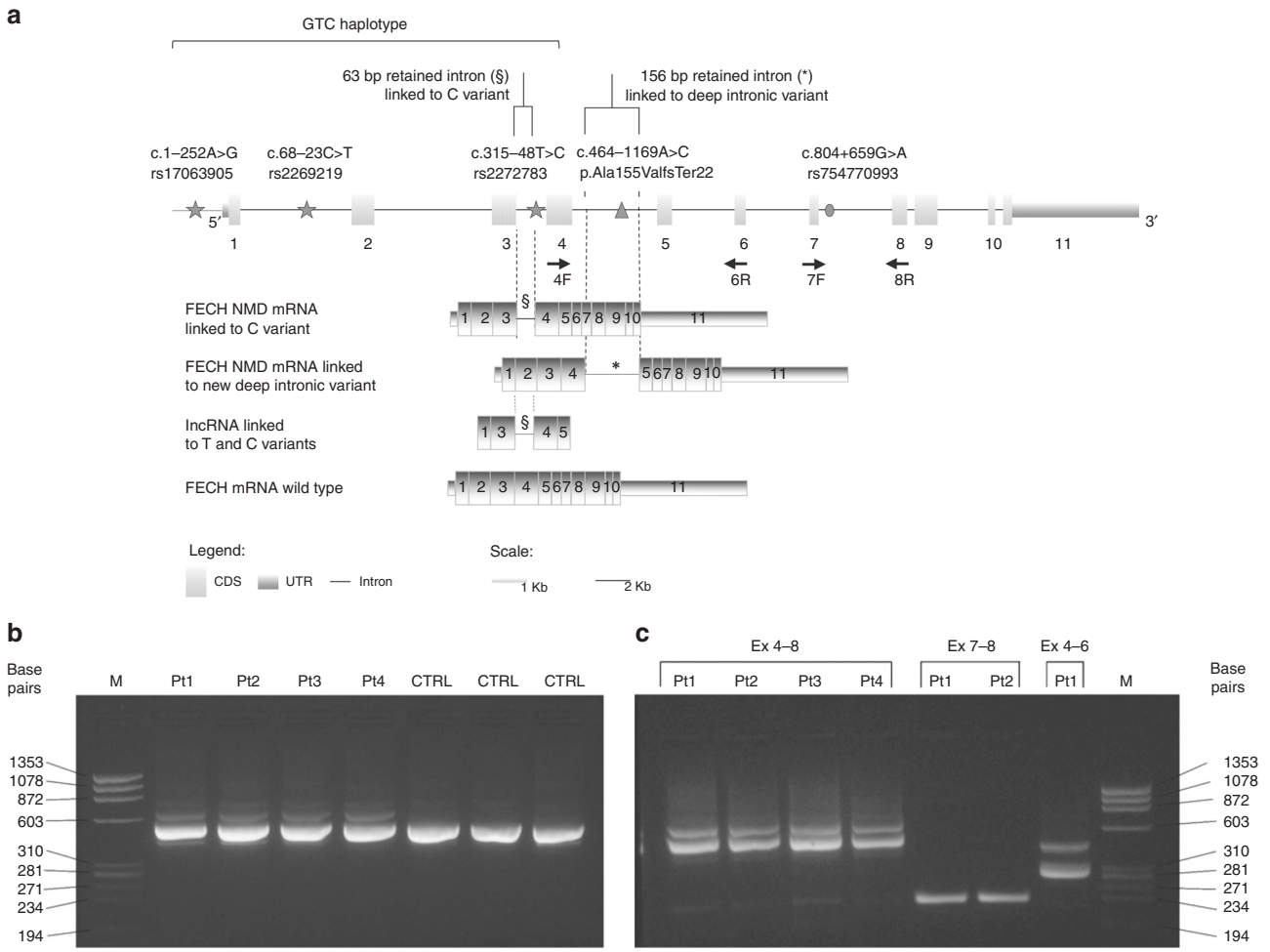

Fig. 1 Scheme of the transcriptional isoforms of the *FECH* gene analyzed in this study and reverse transcription polymerase chain reaction (RT-PCR) experiments. In panel (**a**), gray stars are used to indicate the variants that form the GTC haplotype; the black triangle indicates the deep intronic variant in intron 4, identified in this study; and the gray point indicates the common variant in intron 7. Primers used for the RT-PCR and nested RT-PCR are also displayed. Panel (**b**) shows the PCR product spanning exons 4 to 8. Lane M is the molecular weight marker. The GTC homozygous symptomatic patients carrying the deep intronic pathogenic variant are indicated as Pt1 to Pt4. The following individuals were used as controls: an asymptomatic GTC homozygous carrier, an EPP patient with the common GTC haplotype trans to a pathogenic variant and an healthy subject. Panel (**c**) shows the reamplification of the PCR fragment with primers encompassing exons 4–8, 7–8, and 4–6, respectively. *CDS* coding DNA sequence, *lncRNA* long noncoding RNA, *mRNA* messenger RNA, *NMD* nonsense-mediated decay, *UTR* untranslated region.

(Fig. 1a). It is widely accepted that the aberrant alternatively spliced messenger RNA (mRNA) is degraded by a nonsense-mediated decay (NMD) mechanism, leading, in the context of the null allele, to additional deficiency of ferrochelatase, which is necessary for protoporphyrin accumulation and clinical symptoms.[7,8] Recent independent studies have reported a few cases with a mild EPP phenotype in the presence of the **C** variant in homozygosis[9,10] while others have described a late onset of the EPP phenotype, secondary to a myelodysplastic syndrome.[11] The **C** variant is normally present in healthy human populations, with frequencies ranging from 1–5% in Africa and Europe to 32–37% in East Asia and in South America. Increased frequencies of the **C** variant are not associated with an increase in the prevalence of EPP, even in populations where a high percentage of homozygous subjects (19–22%) is observed.[12]

Several independent studies have reported that two other common variants, c.1–252 A>**G** in the promoter (**G** variant, rs17063905) and c.68–23 C>**T** in intron 1 (**T** variant, rs2269219) are consistently found in association with the c.315–48 T>**C** variant in EPP patients, forming a so-called **GTC** haplotype.[13,14] Interestingly all three variants of the **GTC** haplotype are associated with expression quantitative trait loci (eQTL) that reduce the expression of the *FECH* gene according to the GTEX study (GTEx Analysis Release V7 [dbGaP Accession phs000424.v7.p2]), suggesting a possible role for the **GTC** haplotype as a whole in the pathogenesis of EPP. However, at present it is still unclear whether a homozygous **GTC** state in isolation is sufficient to provoke EPP.[15] In this study, we have performed a high-throughput targeted resequencing of the *FECH* locus in a cohort of 72 individuals belonging to 24 unrelated Italian EPP families. Notably, 5 of

72 subjects were homozygous for the hypomorphic **GTC** allele and showed a variable phenotype where one was completely asymptomatic, while others developed a clinically overt disease from childhood. By comparing hypomorphic **GTC** alleles between patients and unaffected carriers, the study aimed to identify possible functional variants responsible for the variable outcome of EPP.

MATERIALS AND METHODS

Study subjects

All patients were diagnosed based on the clinical history of photosensitivity in the presence of plasma porphyrin peak at 635 nm and high levels of protoporphyrin in the erythrocytes and feces. The sequence of the *FECH* gene was also determined by Sanger sequencing. Both parents when available and healthy relatives were recruited in order to facilitate the reconstruction of the hypomorphic **GTC** haplotype. The study was approved by the Fondazione IRCCS Ca' Granda ethics committee (number 2952, 18 December 2015) and all the subjects signed informed consent prior to their inclusion in the study (Table S1).

Targeted *FECH* resequencing

A custom enrichment panel was designed using Agilent SureDesign™ software to capture 90.2 kb of genomic DNA, from 40 kb upstream to 10 kb downstream of the *FECH* gene, including all exons and introns (chr18 55202704–55292856 on the hg19 human genome assembly). Genomic DNA was extracted from peripheral blood using the Maxwell®16 Automated System (Promega Corporation, Madison, WI) and spectrofluorometrically quantified using QuantiFluor® One dsDNA kit on GloMax Discover® instrument (Promega) according to manufacturer's instructions. Standard Haloplex Target Enrichment system procedure (Agilent Technologies, Santa Clara, CA) was applied for library preparation and 150-bp paired-end reads were generated using a MiSeq sequencer (Illumina, San Diego, CA). Several coverage metrics were recorded to define the accuracy and possible limitations of the enrichment panel.

Variant calling, phasing, and association tests

Variant calling was performed by the CoVaCS pipeline as described by Chiara et al.[16] The ANNOVAR software was employed for variant annotation.[17] The following annotation resources were considered for the estimation of allele frequencies: ExAC (version 1.0 updated 27 February 2017),[18] 1000 Genomes (phase 3),[19] gnomAD (version 2.1, updated 10 December 2018),[18] dbSNP (build 151),[20] Kaviar (version 160204-Public)[21] and TopMed (freeze5, accessed on 28 February 2019, nhlbiwgs.org). RefSeq (release 106)[22] was considered for genes and transcript annotations, ClinVar (version 1.55, updated 26 December 2018)[23] and HGMD-Pro 2018.3[24] for the annotation of disease-causing variants, and the dbNSPF (v4.0b1, updated 30 December 2018)[25] database for the evaluation of nonsynonymous substitutions effect. Nucleotides are numbered based on the FECH NM_000140.3–Human RefSeq transcript, with the A of the ATG initiation codon as "+1."

Adapted functions of the R alleHap package (https://cran.r-project.org/web/packages/alleHap/index.html) were used to obtain the most likely genotype combination and haplotype phasing for trios. Single marker association statistics and visualization of local linkage disequilibrium (LD) were obtained using Haploview software (https://www.broadinstitute.org/haploview/haploview).[26]

FECH gene expression analysis using public data

Vcf files containing the genetic profiles of the individuals included in the GTEx[27] and 1000 Genome studies were retrieved from the dbGaP database[28] (dbGaP study accession: phs000424.v7.p2) and the 1000 Genome data repository (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/) respectively. The expression profiles of the *FECH* gene were obtained directly from the GTEx portal (2016-01-15_v7_RNASeQCv1.1.8_gene_tpm.gct.gz). A custom Perl script was used to extract haplotypes for the *FECH* gene and cross-reference genetic with expression data. A total of 136 individuals, for which both genotypic and gene expression data were available, were considered.

Qualitative RNA analysis by reverse transcription polymerase chain reaction (RT-PCR)

Total RNA was isolated from peripheral blood of patients, using the LEV simplyRNA Blood Kit for Maxwell®16 (Promega), according to the protocol described in Fiorentino et al.[29] First, 400 ng of total RNA were reverse transcribed using the Superscript IV VILO Master Mix (Thermo Fisher, San Francisco, CA) following the protocol supplied with the kit. Then, 50 ng of complementary DNA (cDNA) was amplified using 10 pmol/μL of each primer, in the presence of 1× buffer, 1.5 mM Mg2+, 0.2 mM dNTPs, and 1.25 U of Taq polymerase. The reaction was performed under the following conditions: an initial step at 94 °C for 5 minutes followed by 35 cycles of denaturation at 94 °C for 30 seconds, amplification at 60 °C for 30 seconds and 72 °C for 1 minute, and a final extension at 72 °C for 10 minutes. The region spanning from exons 1 to 8 of the *FECH* gene was amplified using different sets of primer pairs and directly sequenced (Table S2). To increase the signal of abnormal bands during sequencing, cDNA products were reamplified with original and nested primers. Splicing motifs analysis was carried out using Human Splicing Finder v.3 software (http://www.umd.be/HSF3/index.html).[30]

DNA methylation analysis

For DNA methylation analysis 500 ng of DNA was treated with bisulfite using the EZ–96 DNA methylation-Gold kit (Zymo Research, Irvine, CA) in a final elution volume of 200 μL according to manufacturer's instructions. Bisulfite-treated DNA was amplified with PCR for each region of interest: a PCR reaction in 50 μL volume was carried out with 25 μL of Hot Start GoTaq Green Master mix (Promega), 1 pmol of forward primer, 1 pmol of reverse primer, and 25 ng of

bisulfite-treated genomic DNA. Biotin-labeled primers (forward or reverse, depending on the assay) were used to purify the PCR products with Sepharose beads. PCR products were bound to a Streptavidin Sepharose HP (Amersham Biosciences, Uppsala, Sweden), purified, washed, denatured with 0.2 M NaOH, and washed again with the Pyrosequencing Vacuum Prep Tool (Pyrosequencing, Inc., Westborough, MA). Then, pyrosequencing primer (0.3 µm) was annealed to the purified single-stranded PCR product, and methylation analysis was performed by PyroMark MD Q96 (Pyrosequencing, Inc. Westborough, MA). PCR cycling conditions and primer sequences are shown in Table S3.

RESULTS

Targeted resequencing of the *FECH* gene

Observed coverage levels (reported in Table S4) were well in line with the recommendations for the usage of next-generation sequencing (NGS) based resequencing assays in diagnostics: 92.5% of the target regions were covered at 10× or more and 87.8% at 20× or more. However, a consistent proportion of the targeted regions, corresponding to 14.5% of total target, was covered by 10× or less reads in more than 20 samples, suggesting systematic biases in the coverage profiles. Accordingly, these regions were excluded from subsequent analyses. Interestingly we noticed that 94.27% of the low coverage regions correspond with RepeatMasker[31] annotated repeats in the hg19 human genome assembly, reflecting either a reduced rate of mapping of short Illumina reads in highly repetitive regions of the genome or the possible presence of large structural rearrangements or repeat copy-number alterations. A markedly increased coverage (no single region below 20× in all the samples) was observed for the exonic and nonrepetitive regions (Table S4).

Genotyping and haplotype phasing

A total of 510 variable sequence positions at the target *FECH* region were identified in the 72 individuals included in this study (Table S5). Of these 109 were completely novel and had never been reported in any of the publicly available repositories of human genetic variations considered in this study (i.e., 1000 Genomes, gnomAD, ExAC, TopMed, and Kaviar). Importantly, all the EPP-causing variants, previously identified by the Sanger sequencing, were also recovered by the NGS-based assay. In fact, resequencing confirmed that among the 24 patients, 18 coinherited a null allele in *trans* to the hypomorphyic allele, 2 carried only one hypomorphic allele and no pathogenic variants in *trans*, while 4 carried two hypomorphic alleles. Notably the asymptomatic mother of a patient with classical genotype also carried two hypomorphic alleles. The most likely genotypes were obtained by haplotype phasing for all the 72 sequenced subjects to compare single alleles.

Two novel deep intronic variants are associated with EPP

A novel deep intronic variant (NG_008175.1[FECH_v002]: c.464–1169 A>C) and an annotated low-frequency genetic variant (c.804+659 G>A; rs754770993, frequency in 1000 Genomes and ExAC 0; Kaviar and gnomAD 0.0001; TopMed 0.0002), were found in four patients carrying the hypomorphic allele in homozygosis and displaying a clinically overt disease. These variants, which were located, respectively, in introns 4 and 7 of the *FECH* gene, were inherited in *cis* with one of the two hypomorphic alleles from one of the two parents. More importantly, they were not observed in the homozygous hypomorphic subject without symptoms of EPP and in all the other 20 patients carrying the hypomorphic allele.

c.464–1169 A>C variant activates a pseudo-exon in intron 4

RT-PCR was performed to investigate possible effects of the c.464–1169 A>C deep intronic variant on the *FECH* splicing pattern. For amplification covering exons 4–8, an additional longer band was amplified in the four symptomatic **GTC** homozygous patients (Fig. **1b**). Reamplification of the exon 4–8 cDNA fragments with the original and nested primers revealed that the longer band is caused by an insertion between exons 4 and 6 and not between exons 7 and 8 (Fig. **1c**). Direct sequencing showed an insertion in the transcript of a pseudo-exon of 156 bp containing a stop codon sequence (Fig. **2**). This exon, located in intron 4, also encompassed the c.464–1169 A>C deep intronic variant and presented common consensus sequences of splicing.

The c.464–1169 A>C variant is associated with the disruption of an exonic splicing silencer (ESS)

Functional annotation of the deep intronic variants using Human Splicing Finder (HSF) was performed to identify possible functional effects on the splicing pattern. Three different algorithms predicted the disruption of an exonic splicing silencer site (ESS) in intron 4, by the c.464–1169 A>C substitution. The ESS wild-type sequence ([GT]T**A**GGAG) was also recognized as a core binding site for the hnRNP A1 protein; in the presence of the ESS mutated sequence ([GT]T**C**GGAG) this binding site was disrupted. No relevant alterations were reported for the c.804+659 G>A variant.

A single GTC haplotype is linked to the hypomorphic allele

Using simple Mendelian inheritance rules, 446 variants were assigned to distinct haplotypes of *FECH* gene. Of these, 176 were found to be consistently shared among all the individuals carrying the hypomorphic *FECH* allele. Segregation analysis was carried out to establish the parental origin of each variant and the most likely genotype combination in phasing with the c.315–48 **C** was obtained for each trio. The comparisons of the selected hypomorphic alleles showed that 23 of 24 unrelated patients shared an identical haplotype of *FECH* gene of about 20 kb in size. The same presumed haplotype was also identified in both copies of the *FECH* locus in the five individuals homozygous for hypomorphic allele. The haplotype spans from 3.7 kb upstream to the transcription start site (rs75861770) to 1.7 kb in the intron 4 (rs11874117) and contains 47 annotated single-nucleotide

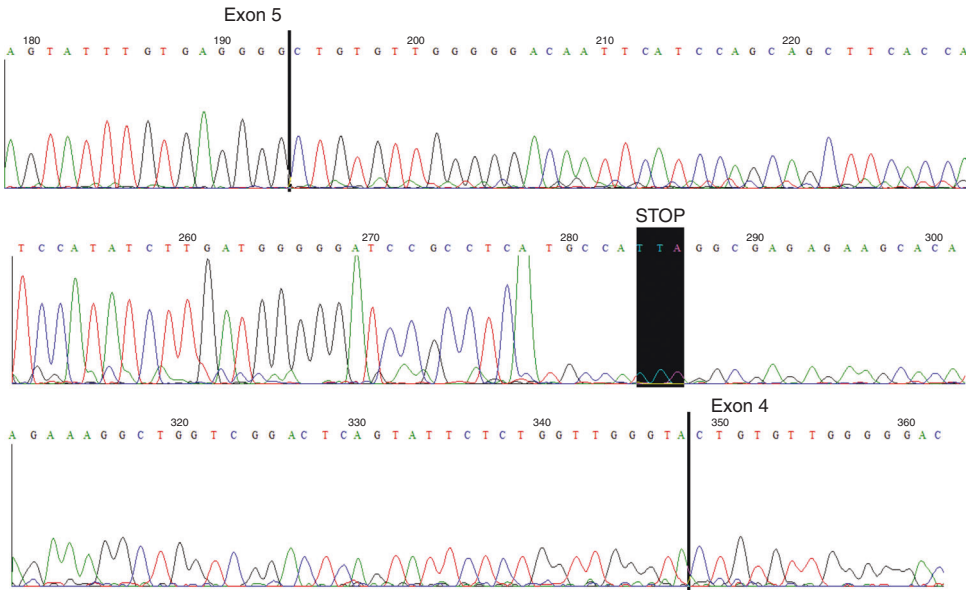

Fig. 2 Sanger sequencing of the polymerase chain reaction (PCR) product encompassing exons 4–6. The sequence shows the insertion of a fragment corresponding to a portion of intron 4 into the *FECH* transcript. The upper panel shows the beginning of the inserted sequence, the middle panel highlights the presence of a stop codon, and the bottom panel shows the junction with exon 4.

polymorphisms (SNPs) including the c.1–252 **G** and the c.68–23 **T** variants (Fig. 3). Interestingly only a single recombination event of the proposed haplotype is observed, in one patient, where only the portion from intron 1 (rs32166686) to intron 4 is retained.

The GTT and ATC haplotypes are associated with reduced *FECH* mRNA levels

Among the 47 SNPs included in the observed haplotype only the known c.315–48 **C** variant was never observed in *trans* to a mutated *FECH* allele in asymptomatic carriers. Notably only two asymptomatic parents presenting very light accumulation of protoporphyrins (6.3 and 5.2 mcg/gHb respectively, n.v. <3), inherited in *trans* to the mutated allele the other two known variants which are part of the GTC haplotype (c.1–252 A>**G** and c.68–23 C>**T**). A chi-squared test, as implemented by the Haploview software, was used to evaluate the level of association for single markers. The analysis confirmed that the c.315–48 **C** variant has the strongest association with the disease (*p* value: 3.29 E–7) while a reduced but still significant association is observed for the c.68–23 **T** (*p* value: 7.03E–6) and c.1–252 **G** (*p* value: 2.45E–5) variants (Table S6). An extensive analysis of the publicly available genotypic and gene expression data (GTEx) of 136 individuals included in the GTEx study evidenced a significant decrease in the expression level of the *FECH* gene both in the carriers of GTT and GTC haplotypes. Consistent with previous observations this decrease is more pronounced in the individuals carrying the GTC haplotype (Fig. 4**a**).

Notably, while the GTC haplotype is overrepresented in our cohort of patients with respect to a population of healthy individuals, we observe no, or very weak, evidence of recombination within the GTC haplotype in the healthy population (Fig. 4**b**). Interestingly, among the patients included in the present study, only one individual with a clinically overt disease carried a "modified" A**TC** haplotype. At the same time, no EPP patient carried the AC**C** haplotype, which is as frequent as the A**TC** (Fig. 4**b**). Of note, a logarithm of the odds (LOD) score of 12.65 was found between the **T** and **C** variants where LOD >2 indicates significant linkage disequilibrium (LD) (Figure S1). Moreover, the HSF analysis of the sequences around the c.68–23 A>**T** and c.315–48 T>**C** eQTLs indicated formation of new ESS sites ([TT]TCATG**T**[GAG] and [G] G**C**TG[CTAA] respectively).

Differential methylation around *FECH* variants are associated with altered splicing patterns

Considering the emerging role of intragenic methylation in the regulation of the alternative splicing, different CpG sites along the gene were analyzed by bisulfite pyrosequencing. As expected for an expressed gene, the methylation in the promoter region was minimal and no measurable difference was noted between patients. Surprisingly, the patients carrying the c.464–1169 A>C deep intronic pathogenic variant presented a new methylated CpG site that was not observed in other patients with a classical GTC haplotype or in unaffected subjects. Conversely, the c.68–23 **T** variant in intron 1 and the variant in intron 7 abolished commonly methylated CpG sites. No alteration was detected in the region encompassing the c.315–48 **C** variant of intron 3 (Table 1). To establish whether the c.68–23 **T** variant also affects the modulation of the *FECH* splicing, a forward primer encompassing exons 1 and 3 and a reverse primer located in the region of alternative splicing of intron 3 were used for RT-PCR (Table S2). The direct sequencing of the PCR product confirmed the presence of an isoform of splicing showing a

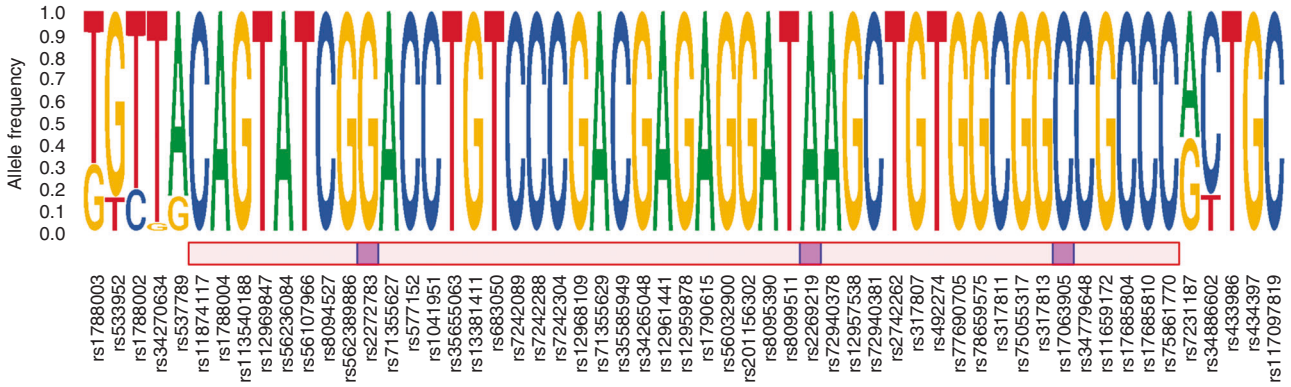

Fig. 3 Sequence logo of the conserved haplotype associated with the hypomorphic allele. Common single-nucleotide polymorphisms (SNPs) that form the haplotype are designated by their respective dbSNP rs code (*x*-axis). A red rectangle is used to illustrate the polymorphic positions that are conserved between all the carriers of the haplotype. The G (rs17063095), T (rs2269219), and C (rs2272783) variants are marked by a purple square.

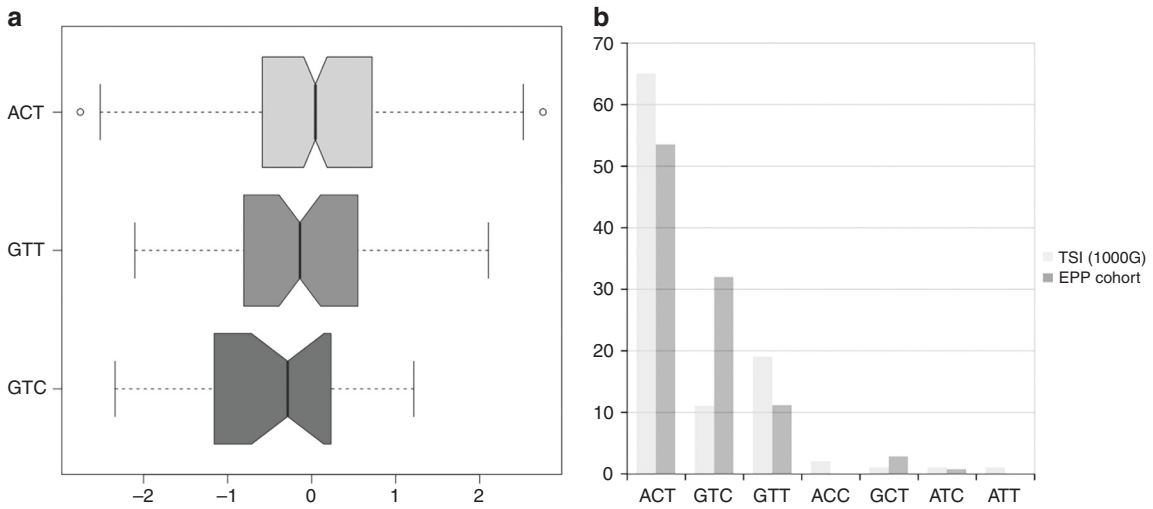

Fig. 4 Comparison of *FECH* gene expression in publicly available data. (**a**) Boxplots of fold change of expression levels of *FECH* in 136 individuals from the GTEx study carrying the GTC, GTT, and ATC haplotypes. Median expression of *FECH* across all the 48 tissues considered in GTEx is used as the baseline for the calculation of the fold change. Fold changes are expressed using a base 2 logarithmic scale. Values lower than 0 indicate downregulation. Values higher than 0 indicate upregulation. (**b**) Barplot of haplotype frequencies in a healthy population (TSI, Tuscans from Italy) and in our cohort of erythropoietic protoporphyria patients (EPP cohort). Haplotypes are indicated based on the G (rs17063095), T (rs2269219), and C (rs2272783) variants.

complete skipping of the constitutive exon 2 and the insertion of the 63 bps of intron 3 (Figure S2). The identified sequence is reported in Ensembl as a noncoding processed transcript (ENST00000585699.1).

DISCUSSION

In this study, by using deep targeted resequencing of the *FECH* gene we report the first evidence of a deep intronic variant causing erythropoietic protoporphyria (EPP). We demonstrate that the c.464–1169 A>C intronic substitution (p.Ala155ValfsTer22) disrupts, likely through the institution of new methylated CpG site, an exonic splicing silencer (ESS) site causing the insertion of a "cryptic exon" containing a stop codon, in the mature *FECH* transcript. It is now clear that constitutive and alternative splicing events in higher eukaryotes are finely regulated through the concerted recognition of multiple well-defined and weak *cis*-acting elements by *trans*-acting factors. Depending on the effect they exert, these weak *cis*-acting elements are generally referred to as either enhancers or silencers.[32] Several lines of evidence suggest that silencers have a fundamental role in preventing pseudo-exon inclusion in mature transcripts and in defining constitutive exons by suppressing nearby decoy splice sites.[33] Additionally, DNA methylation is emerging as an important factor in exon selection by the splicing machinery and also in the regulation of alternative splicing.[34] In particular, the increase of DNA methylation has been reported to promote the inclusion of alternative exons.[35] All these considerations are highly consistent with the proposed significance of the c.464–1169 A>C variant.

Our data also provide independent confirmation that clinically overt EPP is strongly associated with inheritance

Table 1 DNA methylation analysis

Analyzed region	CpG dinucleotides	Analyzed patient Genotype	CTRL1 ACT/ ACT	CTRL2 ACT/ ATT	CTRL3 ATT/ ATT	1561 ACT/ ACT	1553 ACT/ ACT	1574 ACT/ ACT	1611 ACT/ ACT	1655 ACT/ ACT	1704 ACT/ GTT	717 ACT/ GTT	1535 ACT/ GTT	1185 ACT/ GTC	713 ACT/ GTC	1525 ACT/ ATC	1526 ACT/ ATC	1169 GTC/GTC c.464–1169A>C c.804+659G>A	1262 GTC/GTC c.464–1169A>C c.804+659G>A	1382 GTC/GTC c.464–1169A>C c.804+659G>A	1683 GTC/GTC c.464–1169A>C c.804+659G>A	1448 GTC/ GTC
Promoter	CG1		0.6	1.8	2.0	2.3	0.5	0.0	0.5	0.0	nd	0.0	0.7	8.8	1.0	0.5	3.1	0.9	3.0	0.5	1.2	0.6
Sequencing A	CG2		0.0	0.6	1.1	1.1	0.0	0.0	1.7	1.4	nd	0.0	0.0	1.2	0.4	0.0	8.3	1.9	1.1	1.2	0.0	1.0
	CG3		4.0	2.6	3.8	4.0	3.7	5.3	3.6	4.6	nd	4.6	5.1	6.5	2.8	1.9	4.5	3.2	3.1	3.3	6.0	2.2
	CG4		0.0	0.0	1.1	0.7	0.0	3.4	0.7	0.0	nd	0.0	0.0	1.7	0.0	0.0	0.0	0.0	0.7	0.7	0.0	0.8
	CG5		0.7	0.6	0.0	1.1	2.8	0.0	1.2	1.7	nd	0.0	0.0	0.5	1.1	0.0	0.0	0.0	0.6	0.5	0.0	0.0
	CG6		0.0	0.0	0.8	0.6	0.0	0.0	0.0	0.0	nd	0.0	0.0	0.5	0.0	1.1	4.0	0.0	0.0	0.6	2.8	0.0
	CG7		0.0	0.8	1.9	0.9	0.0	0.0	2.4	1.0	nd	0.0	6.3	1.9	2.1	2.7	3.8	0.0	1.3	1.8	0.0	1.7
	CG8		2.9	2.7	2.5	4.0	5.0	4.3	3.5	4.4	nd	11.7	5.0	3.3	4.0	5.2	5.2	3.5	3.2	3.3	6.7	3.3
	CG9		1.6	1.9	1.5	1.9	2.3	2.8	4.8	2.6	nd	4.6	3.5	2.7	2.0	2.2	2.3	2.2	1.7	1.5	2.8	1.7
	CG10		3.2	3.1	3.0	3.4	5.1	6.2	4.7	6.7	nd	10.0	6.4	3.7	3.0	4.3	6.8	4.2	3.4	3.2	7.1	3.0
	CG11		1.6	1.8	1.5	2.0	2.1	2.2	2.3	2.9	nd	1.9	3.5	2.2	1.6	2.5	3.2	1.5	1.6	1.8	4.0	1.9
Promoter	CG1		0.0	0.0	1.3	0.0	1.4	1.3	0.0	0.0	0.0	0.0	0.0	0.7	1.6	0.0	0.0	0.0	1.3	0.0	0.0	0.0
Sequencing B	CG2		0.9	2.2	1.0	0.0	1.7	0.9	1.0	0.0	0.0	1.5	0.0	1.9	2.0	1.5	0.0	0.0	1.9	2.0	1.8	2.4
	CG3		9.8	9.9	9.5	10.9	9.0	8.5	9.8	9.5	16.0	13.7	10.6	6.7	14.2	13.8	14.8	8.3	9.9	8.6	16.8	11.6
	CG4		1.0	2.2	1.0	0.0	1.2	1.9	0.0	0.0	0.0	0.0	0.0	1.3	1.6	0.0	2.0	0.0	0.0	0.0	1.6	0.0
	CG5		1.7	2.9	2.4	3.9	2.5	2.5	3.5	3.6	0.0	0.0	1.9	2.5	4.4	0.0	2.6	2.3	3.3	3.0	1.7	3.8
	CG6		0.0	0.0	0.0	0.0	2.2	0.0	1.1	0.0	0.0	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.9	0.0	0.0	1.0
	CG7		0.0	1.1	0.0	1.2	0.0	2.6	1.2	0.0	0.0	1.5	0.0	0.8	0.0	0.0	0.0	0.0	1.1	2.3	0.0	0.0
Intron 1	CG1		92.7	93.3	93.6	95.1	95.0	92.2	96.3	94.5	96.4	93.3	94.2	97.1	96.7	93.8	93.6	94.3	97.6	95.3	95.9	94.7
	CG2		94.5	96.6	96.4	96.0	96.2	96.7	97.1	95.7	96.6	96.4	96.2	97.1	98.4	96.6	96.6	97.0	96.6	95.6	96.8	96.8
	CG3 pol	c.68–23C>**T**	**72.0**	***38.4***	3.6	**73.8**	**66.9**	**71.3**	**75.6**	**74.9**	***38.2***	***39.4***	***32.0***	***41.6***	***42.7***	***38.9***	***34.6***	2.2	2.2	3.8	2.5	3.2
	CG4		82.8	74.3	75.4	80.7	72.8	82.6	82.9	80.1	84.5	82.2	73.9	84.5	80.6	77.8	75.6	69.0	84.3	71.1	83.5	79.9
	CG5		73.4	66.5	67.8	72.9	64.9	70.0	74.5	72.2	78.6	73.4	66.3	76.2	65.1	72.1	70.0	60.7	78.7	63.4	78.1	72.7
	CG6		83.2	76.1	78.2	79.5	68.1	76.9	78.2	79.7	82.6	77.6	68.7	83.6	77.7	79.6	76.5	73.3	84.4	65.9	81.9	77.1
Intron 3	CG1		95.6	92.2	93.3	88.6	91.4	91.1	91.3	89.9	92.1	92.5	90.5	91.0	85.3	92.9	92.7	90.5	93.5	92.3	92.0	88.7
	CG2		85.1	81.0	81.0	86.0	83.3	80.3	80.9	79.6	82.6	84.5	82.8	84.3	83.5	84.7	82.3	83.9	80.7	81.4	80.5	85.1
	CG3		80.1	89.6	93.3	82.0	89.2	93.8	88.8	87.5	93.0	85.9	89.5	82.1	67.2	88.8	87.2	82.2	87.7	92.2	94.3	81.2
	CG4		99.1	95.4	95.0	95.0	98.0	92.1	95.4	93.0	95.7	94.6	97.9	98.6	96.3	94.1	95.1	93.3	97.9	96.6	96.6	96.8
	CG5		98.0	96.6	94.6	96.9	95.0	94.2	93.0	94.9	94.6	95.4	94.3	95.5	94.3	94.5	95.1	95.8	95.7	96.0	94.7	95.7
Intron 4	CG1		95.4	99.6	99.7	nd	nd	nd	nd	nd	nd	nd	nd	98.7	99.3	96.2	98.3	98.6	99.9	100.0	98.7	98.9
	CG2 mut	c.464–1169A>**C**	8.1	6.2	5.6	nd	nd	nd	nd	nd	nd	nd	nd	6.9	7.5	11.9	8.7	***54.3***	***53.3***	***52.2***	***52.6***	7.3
Intron 7	CG1 pol	c.804+659G>**A**	**97.3**	**96.6**	**97.0**	nd	nd	nd	nd	nd	nd	nd	nd	**95.4**	**96.8**	**96.2**	**97.2**	***55.7***	***56.0***	***57.5***	***57.1***	**95.2**

Methylation levels are reported in the form of dinucleotides percentage of methylation. Relevant differences between the genotype are underlined. Bold values are used to indicate low, medium, and high levels of methylation.

of the c.315–48 **C** variant and that clinical expression of disease typically occurs when this hypomorphic allele exists in *trans* to a null *FECH* allele. These results also confirm that the c.315–48 **C** variant in isolation is necessary but not sufficient to cause an overt disease even when inherited in homozygosis. It was recently shown that abnormal splicing events are twice as frequent in the presence of the c.314–48 **C** variant in heterozygosis. At the same time, this figure does not increase further in homozygous **C** EPP patients, who show *FECH* mRNA levels comparable with those of EPP patients with a classical genotype.[15] Moreover, FECH activity in Japanese healthy controls, homozygous for the **C** variant, was reported to be <50% of that reported for noncarriers, but was still increased by 40% with respect to that of EPP patients.[36] Therefore the presence of the deep intronic pathogenic variant identified in this study, in *cis* with the c.315–48 **C** variant in one of the two hypomorphic alleles, could explain, at least in part, the variable outcome of EPP in homozygous **C** individuals worldwide.

According to our findings in the majority of our unrelated patients (23 of 24), the c.315–48 **C** variant is linked to a single haplotype encompassing the first 20 kb of the *FECH* gene. Among all the variants included in this haplotype, however, the Haploview association analysis recovered a significant association with EPP only for two other known variants: c.1–252 **G** and c.68–23 **T**. Both these single variants were functionally evaluated by in vitro analyses. The c.1–252 A>**G** substitution in the promoter region has been reported to result in a slight decrease in *FECH* transcriptional activity.[37] While, the c.68–23 C>**T** substitution in intron 1, was found to alter the pre-mRNA structure leading to exon 2 skipping during the splicing process.[38] Notably in our cohort, both these single variants were inherited in *trans* to a null allele in two subjects presenting very mild accumulation of protoporphyrins without any apparent clinical symptoms. This evidence suggests that both variants in *trans* to a null *FECH* allele can result in a slight decrease in the FECH activity and in a mild PPIX accumulation but are not sufficient to consistently cause clinical expression of the disease. Consistent with this hypothesis, extensive publicly available gene expression data from the GTEx study provide evidence for a decrease in the expression level of the *FECH* transcript also in individuals carrying the **GT**T haplotype. Taken together these observations are consistent with the idea that the hypomorphic allele is prevalently inherited in the form of the **GTC** haplotype. Considerations on the relative frequency of the **GT**T and **GTC** haplotypes (**GT**T approximately 22% and **GTC** approximately 11%) in the healthy population suggest that **GTC** is a derived form of the **GT**T haplotype and that it is associated with a more marked decrease in the expression levels of the *FECH* gene.

The observation that in our cohort one symptomatic patient was not carrying the c.1–252 **G** variant in the promoter suggests that this might not be required to induce an overt disease. On the contrary, all patients showed extended levels of linkage disequilibrium between the c.68–23 **T** and c.315–48 **C** variants, located in introns 1 and 3 respectively, supporting the hypothesis that both are necessary for the lower steady state level of *FECH* mRNA resulting in protoporphyrin overproduction and photosensitivity. Importantly both these variants are associated with the creation of new exonic splicing silencer sites (ESS) according to the HSF tool. Moreover, we demonstrate that the c.68–23 **T** variant alters the DNA methylation pattern by abolishing a methylated CpG site, with an opposite effect with respect to the c.464–1169 A>C deep intronic pathogenic variant. Additionally, the identification of a noncoding splicing isoform, with a complete skipping of the constitutive exon 2 and the insertion of the 63 bps of intron 3, strongly supports the conclusion that both variants are required for clinically relevant downregulation of the *FECH* gene. Considerations regarding the relatively low frequency of the A**TC** and AC**C** with regard to the **GTC** haplotype suggest that the functional characterization of the **T** variants warrants further investigation and probably requires the study of a larger cohort of EPP patients worldwide.

In conclusion, our findings suggest that although the majority of EPP-causing variants has been shown to have a "radical" effect on the coding sequence of the *FECH* gene, the presence of noncoding variants in the pathogenic process should consistently be evaluated especially in EPP patients carrying only one hypomorphic allele. Moreover, we believe that this study supports the recent findings that methylation-dependent modulation of the pre-mRNA splicing patterns may function directly to control gene expression levels through the incorporation of "poison exons" leading to NMD.[39,40] All in all the findings of this study confirm the validity of the hypothesis that "hidden" sources of variability that are not normally considered in clinical genetic screenings might explain at least in part the variability in the clinical manifestations of diseases with incomplete penetrance.

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-019-0584-0) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

The authors are grateful to all patients and their families who donated samples for this study. We thank Pasquale Missineo for biochemical analysis, Valeria Fiorentino for support in the panel design, Luca Turchet for graphic assistance, and David Horner for English editing and comments that greatly improved the manuscript. We also thank M.D. Cappellini for her constant and continuous support in the research activity. This research was supported by grants from the Italian Ministry of Health (GR–2011–02347129 to E.D.P.) and in part from Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (RC2019).

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Lecha M, Puy H, Deybach JC. Erythropoietic protoporphyria. Orphanet J Rare Dis. 2009;4:19.
2. Balwani M, Naik H, Anderson KE, Bissell DM, Bloomer J, Bonkovsky HL, et al. Clinical, biochemical, and genetic characterization of North American patients with erythropoietic protoporphyria and X-linked protoporphyria. JAMA Dermatol. 2017;153:789–796.
3. Wahlin S, Stal P, Adam R, Karam V, Porte R, Seehofer D, et al. Liver transplantation for erythropoietic protoporphyria in Europe. Liver Transpl. 2011;17:1021–1026.
4. Whatley SD, Mason NG, Khan M, Zamiri M, Badminton MN, Missaoui WN, et al. Autosomal recessive erythropoietic protoporphyria in the United Kingdom: prevalence and relationship to liver disease. J Med Genet. 2004;41:e105.
5. Whatley SD, Mason NG, Holme SA, Anstey AV, Elder GH, Badminton MN. Molecular epidemiology of erythropoietic protoporphyria in the U.K. Br J Dermatol. 2010;162:642–646.
6. Balwani M, Doheny D, Bishop DF, Nazarenko I, Yasuda M, Dailey HA, et al. Loss-of-function ferrochelatase and gain-of-function erythroid-specific 5-aminolevulinate synthase mutations causing erythropoietic protoporphyria and x-linked protoporphyria in North American patients reveal novel mutations and a high prevalence of X-linked protoporphyria. Mol Med. 2013;19:26–35.
7. Gouya L, Puy H, Robreau AM, Bourgeois M, Lamoril J, Da S,V, Grandchamp B, et al. The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype FECH. Nat Genet. 2002;30:27–28.
8. Barman-Aksozen J, Beguin C, Dogar AM, Schneider-Yin X, Minder EI. Iron availability modulates aberrant splicing of ferrochelatase through the iron- and 2-oxoglutarate dependent dioxygenase Jmjd6 and U2AF(65.). Blood Cells Mol Dis. 2013;51:151–161.
9. Schneider-Yin X, Mamet R, Minder EI, Schoenfeld N. Biochemical and molecular diagnosis of erythropoietic protoporphyria in an Ashkenazi Jewish family. J Inherit Metab Dis. 2008;31 suppl 2:S363–S367.
10. Mizawa M, Makino T, Nakano H, Sawamura D, Shimizu T. Incomplete erythropoietic protoporphyria caused by a splice site modulator homozygous IVS3-48C polymorphism in the ferrochelatase gene. Br J Dermatol. 2016;174:172–175.
11. Suzuki H, Kikuchi K, Fukuhara N, Nakano H, Aiba S. Case of late-onset erythropoietic protoporphyria with myelodysplastic syndrome who has homozygous IVS3-48C polymorphism in the ferrochelatase gene. J Dermatol. 2017;44:651–655.
12. Nakano H, Nakano A, Toyomaki Y, Ohashi S, Harada K, Moritsugu R, et al. Novel ferrochelatase mutations in Japanese patients with erythropoietic protoporphyria: high frequency of the splice site modulator IVS3-48C polymorphism in the Japanese population. J Invest Dermatol. 2006;126:2717–2719.
13. Li C, Di Pierro E, Brancaleoni V, Cappellini MD, Steensma DP. A novel large deletion and three polymorphisms in the FECH gene associated with erythropoietic protoporphyria. Clin Chem Lab Med. 2009;47:44–46.
14. Colombo FP, Rossetti MV, Mendez M, Martinez JE, Enriquez de SR, del CBA, et al. Functional associations of genetic variants involved in the clinical manifestation of erythropoietic protoporphyria in the Argentinean population. J Eur Acad Dermatol Venereol. 2013;27:754–762.
15. Brancaleoni V, Granata F, Missineo P, Fustinoni S, Graziadei G, Di Pierro E. Digital PCR (dPCR) analysis reveals that the homozygous c.315-48T>C variant in the FECH gene might cause erythropoietic protoporphyria (EPP). Mol Genet Metab. 2018;124:287–296.
16. Chiara M, Gioiosa S, Chillemi G, D'Antonio M, Flati T, Picardi E, et al. CoVaCS: a consensus variant calling system. BMC Genomics. 2018;19:120.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
18. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–291.
19. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
20. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29:308–311.
21. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. Bioinformatics. 2011;27:3216–3217.
22. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733–D745.
23. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):D980–D985.
24. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet. 2017;136:665–677.
25. Liu X, Wu C, Li C, Boerwinkle E. dbNSFPv3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum Mutat. 2016;37:235–241.
26. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–265.
27. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648–660.
28. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res. 2014;42(Database issue):D975–D979.
29. Fiorentino V, Brancaleoni V, Granata F, Graziadei G, Di Pierro E. The assessment of noncoding variant of PPOX gene in variegate porphyria reveals post-transcriptional role of the 5′ untranslated exon 1. Blood Cells Mol Dis. 2016;61:48–53.
30. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37:e67.
31. Smith AF, Hubley R, Green P. RepeatMasker Open-3.0. 2010. http://www.repeatmasker.org.
32. Chasin LA. Searching for splicing motifs. Adv Exp Med Biol. 2007;623:85–106.
33. Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, et al. Silencer elements as possible inhibitors of pseudoexon splicing. Nucleic Acids Res. 2004;32:1783–1791.
34. Lev MG, Yearim A, Ast G. The alternative role of DNA methylation in splicing regulation. Trends Genet. 2015;31:274–280.
35. Yearim A, Gelfman S, Shayevitch R, Melcer S, Glaich O, Mallm JP, et al. HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. Cell Rep. 2015;10:1122–1134.
36. Tahara T, Yamamoto M, Akagi R, Harigae H, Taketani S. The low expression allele (IVS3–48C) of the ferrochelatase gene leads to low enzyme activity associated with erythropoietic protoporphyria. Int J Hematol. 2010;92:769–771.
37. Di Pierro E, Cappellini MD, Mazzucchelli R, Moriondo V, Mologni D, Zanone PB, et al. A point mutation affecting an SP1 binding site in the promoter of the ferrochelatase gene impairs gene transcription and causes erythropoietic protoporphyria. Exp Hematol. 2005;33:584–591.
38. Nakahashi Y, Fujita H, Taketani S, Ishida N, Kappas A, Sassa S. The molecular defect of ferrochelatase in a patient with erythropoietic protoporphyria. Proc Natl Acad Sci USA. 1992;89:281–285.
39. Carvill GL, Engel KL, Ramamurthy A, Cochran JN, Roovers J, Stamberger H, et al. Aberrant inclusion of a poison exon causes Dravet syndrome and related SCN1A-associated genetic epilepsies. Am J Hum Genet. 2018;103:1022–1029.
40. Yan Q, Weyn-Vanhentenryck SM, Wu J, Sloan SA, Zhang Y, Chen K, et al. Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. Proc Natl Acad Sci USA. 2015;112:3445–3450.