

## ORIGINAL ARTICLE

# Activation of cryptic splice sites in three patients with chronic granulomatous disease

Martin de Boer<sup>1</sup> | Karin van Leeuwen<sup>1</sup> | Mathias Hauri-Hohl<sup>2</sup> | Dirk Roos<sup>1</sup> 

<sup>1</sup>Sanquin Research and Landsteiner Laboratory, Amsterdam Medical Center, University of Amsterdam, Amsterdam, The Netherlands

<sup>2</sup>Department of Stem Cell Transplantation Research, University Children's Hospital Zürich, Zürich, Switzerland

**Correspondence**

Dirk Roos, Sanquin Research, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands.  
Email: d.roos@sanquin.nl

**Abstract**

**Background:** Chronic granulomatous disease (CGD) is a primary immune deficiency caused by mutations in the genes encoding the structural components of the phagocyte NADPH oxidase. As a result, the patients cannot generate sufficient amounts of reactive oxygen species required for killing pathogenic microorganisms.

**Methods:** We analyzed NADPH oxidase activity and component expression in neutrophils, performed genomic DNA and cDNA analysis, and used mRNA splicing prediction tools to evaluate the impact of mutations.

**Results:** In two patients with CGD, we had previously found mutations that cause aberrant pre-mRNA splicing. In one patient an exonic mutation in a cryptic donor splice site caused the deletion of the 3' part of exon 6 from the mRNA of *CYBB*. This patient suffers from X-linked CGD. The second patient, with autosomal CGD, has a mutation in the donor splice site of intron 1 of *CYBA* that activates a cryptic donor splice site downstream in intron 1, causing the insertion of intronic sequences in the mRNA. The third patient, recently analyzed, also with autosomal CGD, has a mutation in intron 4 of *CYBA*, 15 bp from the acceptor splice site. This mutation weakens a branch site and activates a cryptic acceptor splice site, causing the insertion of 14 intronic nucleotides into the mRNA.

**Conclusion:** We found three different mutations, one exonic, one in a donor splice site and one intronic, that all caused missplicing of pre-mRNA. We analyzed these mutations with four different splice prediction programs and found that predictions of splice site strength, splice enhancer and splice silencer protein binding and branch site strength are all essential for correct prediction of pre-mRNA splicing.

**KEYWORDS**

chronic granulomatous disease (CGD), cryptic splice site, *CYBA*, *CYBB*, gp91<sup>phox</sup> deficiency, p22<sup>phox</sup> deficiency

**Abbreviations:** (P)ESE, (putative) exonic splicing enhancer; (P)ESS, (putative) exonic splicing silencer; AR-CGD, autosomal recessive CGD; CGD, chronic granulomatous disease; *CYBA*, cytochrome *b* alpha-chain; *CYBB*, cytochrome *b* beta-chain; EJC, exon junction complex; hnRNP, heterogeneous nuclear ribonuclear protein; NMD, nonsense-mediated decay; NOX, NADPH oxidase; PABP, poly(A)-binding proteins; phox, phagocyte oxidase; snRNP, small nuclear ribonuclear protein; SR proteins, serine- and arginine-rich proteins; SRSF1, serine/arginine-rich splicing factor-1; Ter, protein synthesis termination codon; UTR, untranslated region; X-CGD, X-linked CGD.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 Sanquin Blood Supply Organization. *Molecular Genetics & Genomic Medicine* published by Wiley Periodicals, Inc.

## 1 | INTRODUCTION

Chronic granulomatous disease (CGD) is characterized by the failure of activated phagocytes to generate superoxide, which is needed for killing pathogens. As a result, CGD patients suffer from serious episodes of bacterial and fungal infections. Defects in at least five different genes lead to CGD. Patients with the X-linked form of CGD (OMIM #306400) have mutations in *CYBB* (OMIM \*300481) encoding the beta-subunit of cytochrome *b*<sub>558</sub> (gp91<sup>phox</sup>/NOX2). Patients with a rare autosomal recessive form of CGD (OMIM #233690) have mutations in *CYBA* (OMIM \*608508) encoding the alpha-subunit of this cytochrome (p22<sup>phox</sup>). Usually, mutations in either of these genes lead to the absence of cytochrome *b*<sub>558</sub> in the phagocytes (X91<sup>0</sup> and A22<sup>0</sup> CGD).

Mutations that cause missplicing of pre-mRNA are a frequent cause of disease, not only in CGD (De Boer et al., 1992), but also in other diseases. Most frequently, this is caused by a mutation within a donor or acceptor splice site, which then “weakens” this splice site and usually causes the adjacent exon to be skipped from the mRNA. However, sometimes a nearby cryptic splice site is strong enough to take over, resulting in skipping of part of the adjacent exon or inclusion of part of the adjacent intron. Other mechanisms are also possible. For instance, we (De Boer et al., 1992) and others (Caminsky, Mucaki, & Rogan, 2015, in a review) have found mutations that create or activate a “stronger” splice site than the wild-type (w.t.) splice site, resulting in an incorporation of incomplete or extended exons. Caminsky et al. (2015) report 28 cases with creation or activation of an exonic or intronic cryptic donor splice site and 11 with activation or creation of an exonic or intronic acceptor splice site. Of these 28 cases, 18 concerned an exonic missense mutation, five an exonic synonymous variation, and five an intronic variation. Exonic nonsense mutations were not reported. Of the 11 cases with creation or activation of a cryptic acceptor splice site, eight were caused by an intronic variation, two by an exonic missense mutation and one by an exonic synonymous variation. Also deep intronic mutations can change the splicing process, usually by facilitating the inclusion of a pseudoexon in the mRNA (Dhir & Buratti, 2010; Greer et al., 2015; Homolova et al., 2010; Yamaguchi et al., 2010). Indeed, with the present-day use of whole genome sequencing (WGS), many of these deep intronic variants have been found.

We here describe the activation of a cryptic donor splice site caused by an exonic single base-pair (bp) substitution in *CYBB* (patient 1). We and others found this mutation in *CYBB* already several years ago (De Boer et al., 1992), but at that time no information was available about serine- and arginine-rich (SR) proteins that act as splicing enhancers of pre-mRNA. These proteins bind to distinct exonic splicing

enhancer (ESE) sites to facilitate inclusion of the exon in the mature mRNA transcript. Similarly, heterogeneous nuclear ribonuclear proteins (hnRNPs) can bind to exonic splicing silencer (ESS) sequences to prevent inclusion of the exon in the mature mRNA transcript. We have now evaluated with the widely used Alamut<sup>®</sup> software the effects of the presence of these binding sites on the activation of this exonic cryptic splice site in *CYBB*.

In patient 2, we have described the activation of a cryptic donor splice site 79 bp downstream in intron 1 of *CYBA*, due to a 4-bp deletion in the donor splice site of this intron (De Boer, Hartl, Wintergerst, Belohradsky, & Roos, 2005). This is a more commonly found cause of cryptic acceptor splice site activation, but less frequently as a cause of cryptic donor splice site activation. Caminsky et al. (2015) reported 42 cases of cryptic acceptor splice site activation but only three cases of cryptic donor splice site activation, of which one due to an intronic variation, one to an exonic missense mutation, and one to an exonic synonymous variation.

In patient 3, we recently found a mutation on one allele in intron 4 of *CYBA*, 15 nucleotides upstream of the intron 4/exon 5 border. This intronic mutation created an alternative splice site that was preferred over the original acceptor splice site, thus causing the insertion of 14 intronic nucleotides into exon 5.

## 2 | MATERIALS AND METHODS

### 2.1 | Editorial policies and ethical considerations

This study was performed according to national regulations with respect to the use of human materials from healthy, anonymized volunteers and patients with informed consent, and all experiments were approved by the Medical Ethical Committee of the Academic Medical Center in Amsterdam according to the Declaration of Helsinki principles (version Seoul 2008).

The clinical histories of patients 1 and 2 have been described previously (De Boer et al., 1992, 2005).

### 2.2 | Clinical history patient 3

Patient 3 is the second of two girls, her sister died at age of 2.5 years due to septicemia, post-mortem primary immunodeficiency suspected. The patient was born after an uneventful pregnancy and birth to non-consanguineous Caucasian parents; at the age of 1 month she developed bilobar pneumonia, hepatitis, enterocolitis and vesiculopustulous skin lesions with uneventful recovery. The causative pathogen was not identified but was possibly viral. At the age of 3 months, she developed left-sided lobar

pneumonia, with granuloma formation in the lung on CT scan and positive galactomannan testing in bronchioalveolar lavage. A pathological dihydrorhodamine-1,2,3 (DHR) test was found. Besides antimicrobial treatment, anti-inflammatory treatment with kineret (anakinra) was started to treat granulomatous inflammation in the lungs, resulting in resolution of pulmonary lesions. At the age of 14 months, onset of inflammatory changes in the colon with proctitis was noted, with formation of anorectal fistula and intermittent bloody diarrhea with severe pancolitis, partially responsive to systemic steroids, azathioprine and gut decontamination, yet remained steroid-dependent. Due to prolonged colitis and steroid treatment, the patient demonstrated failure to thrive. At the age of 2 years and 4 months, the patient received transplantation of bone marrow graft from a 10/10-human leukocyte antigen (HLA)-identical donor after reduced-intensity conditioning with pharmacokinetic-adjusted busulfan to submyeloablative levels, high-dose fludarabine, and low-dose alemtuzumab. Posttransplant development of mild acute graft-versus-host disease was noted, with complete resolution. No other complications were observed so far (currently more than 1 year post-transplant).

### 2.3 | Neutrophil function tests

NADPH oxidase activity by purified neutrophils was measured with the Amplex Red assay as described previously (De Boer et al., 2014). Expression of gp91<sup>phox</sup> was measured by FACS analysis of mAb 7D5 binding to purified neutrophils (De Boer et al., 2014).

### 2.4 | Mutation analysis of gDNA and cDNA

Genomic DNA was purified from human blood and analyzed by Sanger sequencing as described before (De Boer et al., 2014). mRNA was purified from human blood, converted into cDNA and analyzed by Sanger sequencing as described before (De Boer et al., 2014).

### 2.5 | Splicing prediction tools

With the prediction tools from The Edmond J. Safra Center for Bioinformatics from the university of Tel Aviv, Israel: <http://astlab.tau.ac.il/index.php> (Cartegni, Wang, Zhu, Zhang, & Krainer, 2003; Fairbrother, Yeh, Sharp, & Burge, 2002; Goren et al., 2006; Kol, Lev-Maor, & Ast, 2005; Schwartz et al., 2008; Voelker & Berlund, 2007; Wang et al., 2004; Yeo & Burge, 2004; Yeo, Nostrand, & Liang, 2007; Zhang & Chasin, 2004), we analyzed the scores of w.t. and mutant splice sites. In the Analyser Splice Tool, the 3' (acceptor) splice site TTTTTTTTTTTCAG/G will give a score of 100, and the 5' (donor) splice site CAG/GTAAGT will

also give a score of 100. We used Sroogle from the same website to analyze the number of putative binding sites for ESE and ESS proteins. Also the Alamut<sup>®</sup> software ([www.interactive-biosoftware.com/doc/alamut-visual/2.6/splicing.html](http://www.interactive-biosoftware.com/doc/alamut-visual/2.6/splicing.html)) was used to analyze possible splice sites. This tool includes SpliceSiteFinder-like, MaxEntScan, NNSPLICE, and GeneSplicer.

Hexplorer score for flanking splicing regulatory elements was analyzed with HEXplorer algorithm at [www2.hhu.de/rna/html/hexplorer\\_score.php](http://www2.hhu.de/rna/html/hexplorer_score.php).

## 3 | RESULTS

### 3.1 | Patient 1

Patient 1 had died before the start of our investigations. In the DNA from his mother we found—as described before—a heterozygous mutation in exon 6 of the *CYBB*, viz. c.621C>A (ref. seq. NM\_000397.3), p.Tyr207Ter (family 5 in De Boer et al., 1992). Besides the fact that this mutation introduces a premature termination codon, cDNA analysis revealed skipping of the 3' part of exon 6 from *CYBB* (Figure 1A). In this female carrier, neither mutations in the exon 5 / intron 5 boundary sequence nor any other mutations were identified in *CYBB*. Because part of the mRNA of this individual lacked 57 bp of the 3' part of exon 6, starting at nucleotide c.618 (De Boer et al., 1992), it appears that this mutation has activated a cryptic donor splice site at nucleotide c.615\_623 (TTG/GTAAAC).

The Analyser Splice Tool indicated a score for the w.t. exon 6 donor splice site of 70.4 (Table 1), which is low compared to the average score of 80.0 found for normal donor splice sites (Itoh, Washio, & Tomita, 2004). Also the other prediction programs indicated low values for the original donor splice site, but nevertheless higher than the new splice site at c.615\_623 used in the patient. This explains why the original site was used in the w.t. situation. In the patient, the new splice site contained the mutation and therefore showed slightly higher values than in the w.t. situation. However, the strength of the new site appeared lower than the original site (Table 1), so it remained to be explained why in the patient the new site was preferred over the original one. Nevertheless, the SPANR tool (Xiong et al., 2015) predicted 29% mRNA correctly spliced in the patient, in contrast to 83% correctly spliced in the w.t. situation.

We then used Sroogle to determine the number and position of putative exonic splicing enhancer (PESE) and silencer (PESS) motifs in the complete exon 6, the 5' part (c.484\_617) and the 3' part (c.618\_684) of this exon (Figure S1). This showed that the 5' part (upstream of the cryptic donor splice site) had a much higher PESE:PESS ratio than the 3' part (Table 1), which thus favored skipping of the 3' part from the mature mRNA. The ESEfinder web tool showed that due

**TABLE 1** Splice site strengths and PESE/PSS scores

	wild-type	patient
<b>CGD patient 1</b>		
Original splice site strength		
SpliceSiteFinder-like	70.4	70.4
MaxEntScan	6.5	6.5
NNSPLICE	0.7	0.7
GeneSplicer	b.t.	b.t.
New splice site strength		
SpliceSiteFinder-like	b.t.	b.t. <sup>a</sup>
MaxEntScan	0.5	3.2
NNSPLICE	b.t.	0.7
GeneSplicer	b.t.	b.t.
Branch points strengths	unaffected	unaffected
<b>complete exon 6</b>	<b>5' part exon 6</b>	<b>3' part exon 6</b>
PESE score <sup>b</sup>	PESE score	PESE score
14/1.91 = 7.3%	10/1.34 = 7.5%	4/0.57 = 7.0%
PSS score <sup>b</sup>	PSS score	PSS score
2/1.91 = 1.1%	1/1.34 = 0.8%	1/0.57 = 1.8%
Ratio PESE:PSS	Ratio PESE:PSS	Ratio PESE:PSS
7.0	9.9	4.0
<b>CGD patient 2</b>		
Original splice site strength		
SpliceSiteFinder-like	70.4	b.t.
MaxEntScan	6.5	b.t.
NNSPLICE	0.7	b.t.
GeneSplicer	7.0	b.t.
New splice site strength		
SpliceSiteFinder-like	71.3	71.3
MaxEntScan	2.2	2.2
NNSPLICE	b.t.	b.t.
GeneSplicer	1.0	1.0
Branch points strengths	unaffected	
<b>w.t. exon 1</b>	<b>mutated exon 1</b>	<b>3' part exon 1 (mutated)</b>
PESE score	PESE score	PESE score
6/1.97 = 3.1%	6/2.76 = 2.2%	0/0.79 = 0%
PSS score	PSS score	PSS score
9/1.97 = 4.6%	14/2.76 = 5.2%	5/0.79 = 6.3%
Ratio PESE:PSS	Ratio PESE:PSS	Ratio PESE:PSS
0.7	0.4	0

(Continues)

**TABLE 1** (Continued)

	wild-type	patient
<b>CGD patient 3</b>		
Original splice site strength		
SpliceSiteFinder-like	81.5	81.5
MaxEntScan	8.2	5.9
NNSPLICE	0.9	0.9
GeneSplicer	11.4	7.3
New splice site strength		
SpliceSiteFinder-like	b.t.	79.1
MaxEntScan	b.t.	7.7
NNSPLICE	b.t.	0.5
GeneSplicer	b.t.	9.8
Branch point strength at c.288-16	94.8 <sup>c</sup> (3.0 <sup>d</sup> )	84.7 <sup>c</sup> (b.t. <sup>d</sup> )
Branch point strength at c.288-42	89.5 <sup>c</sup> (3.45 <sup>d</sup> )	89.5 <sup>c</sup> (3.45 <sup>d</sup> )
<b>w.t. exon 5</b>	<b>new exon 5</b>	<b>5' part new exon 5</b>
PESE score	PESE score	PESE score
1/0.82 = 1.2%	1/0.96 = 1.0%	0/0.14 = 0%
PSS score	PSS score	PSS score
2/0.82 = 2.4%	1/0.96 = 1.0%	0/0.14 = 0%
Ratio PESE:PSS	Ratio PESE:PSS	Ratio PESE:PSS
0.5	0.5	N.A.

Note: In silico prediction scores as reported in the Alamut<sup>®</sup> web tool for patient 1 (upper part), patient 2 (middle part) and patient 3 (lower part). The splice sites scores for the original and the new splice sites are given in the top sections, the branch point strengths in the middle sections, and the putative exonic splicing enhancer (PESE) and silencer (PSS) motif scores (number of motifs per 100 bp × 100%) and the ratio of enhancer and silencer binding motifs in the bottom sections.

Abbreviations: b.t., below threshold; N.A., not applicable

<sup>a</sup>After lowering the threshold, the patient value was about 12% higher than the w.t. value.

<sup>b</sup>PESE or PSS score: number of PESE or PSS per 100 nucleotides × 100%.

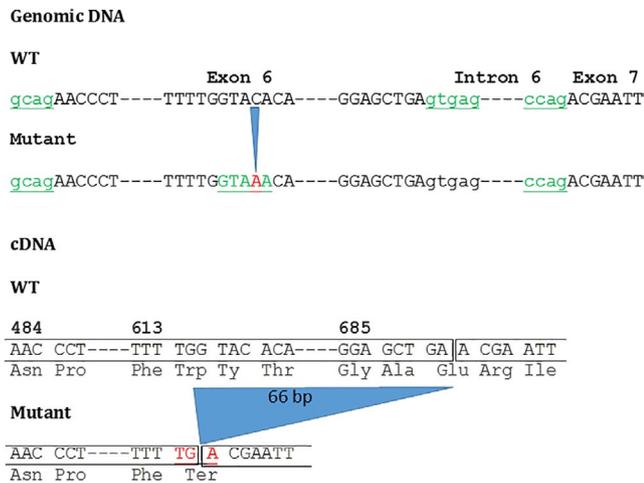
<sup>c</sup>According to Alamut<sup>®</sup>.

<sup>d</sup>According to ESE Finder.

to this mutation, a cryptic donor splice site appeared with a score of 2.5 and an SRSF1 binding site disappeared in the mutation area (Figures S2 and S3).

### 3.2 | Patient 2

Analysis of genomic DNA in patient 2, with an autosomal recessive form of CGD (AR-CGD), had revealed a 4-bp deletion at the exon-1/intron-1 boundary in *CYBA* (c.58+4\_7delAGTG, ref. seq. NM\_000101.3), predicting



**FIGURE 1A** Schematic representation of the consequences of the mutations in the patients compared to the wild-type situation. (A) Patient 1. The upper part represents the situation on the genomic level of the *CYBB* w.t. and the mutant sequence. Intronic sequences in lower case lettering, exonic sequences in boxed capitals. The blue arrow points to the mutation indicated in red. Indicated in green are the locations of the acceptor, donor and cryptic donor splice sites used for splicing of the pre-mRNAs. The lower part represents the situation on the cDNA and protein level. The w.t. and mutant sequence are shown. The blue triangle indicates the 66-bp region of the mRNA that is removed from the transcript after splicing of the pre-mRNA in the patient, due to the use of the cryptic donor splice site in exon 6 in the *CYBB* gene. In red is indicated the premature termination codon in the mutant sequence, which is located on the boundary of exons 6 and 7. Exon-exon boundaries are indicated by |

p.Ile20SerfsTer77 (De Boer et al., 2005; Ishibashi, Nunoi, Endo, Matsuda, & Kanegasaki, 2000), (minimal allele frequency <0.0001%). In the patient's cDNA, we found a low expression of an abnormal product, containing exon 1 extended by 79 nucleotides from intron 1, joined to exon 2. This extension was apparently caused by the activation of a cryptic donor splice site with a GT sequence at position 84–85 in intron 1 (c.58+84\_85; Figure 1B). Both parents of the patient had the same mutation in their genomic DNA, in heterozygous form, but their cDNA contained exclusively the w.t. p22<sup>phox</sup> cDNA sequence, indicating that the mutant mRNA might be degraded (De Boer et al., 2005).

The Analyser Splice Tool indicated a score for the w.t. *CYBA* exon 1 donor splice site of 70.4, but this was decreased by the mutation to below threshold (Table 1). Apparently for this reason, the cryptic donor splice site in intron 1, with a score of 71.3, was preferred. The other prediction programs also indicated a decrease in the score for the mutated splice site and similar values for the activated cryptic splice site and the original one in the w.t. situation (Table 1). Sroogle showed that the original exon 1 had a very low PESE:PESS score, which was even further decreased by inclusion of the 79 nucleotides of intron 1 in this exon (Table 1).

### 3.3 | Patient 3

Patient 3 was identified as a CGD patient by the near absence of NADPH oxidase activity and absence of gp91<sup>phox</sup> expression in her neutrophils, as shown in Figure 2. Genetically, patient 3 was a compound heterozygote for two different mutations in *CYBA*. On one allele she had a one-nucleotide deletion (c.246delC, Figure 3A), inducing a shift in the reading frame that predicts termination of protein synthesis 108 codons downstream (p.Phe83LeufsTer108), see Figure 1C. On the other allele, her *CYBA* contained a variation in intron 4, 15 nucleotides upstream of the start of exon 5 (c.288-15C>G, Figure 3B). Apparently, this variation created a cryptic splice site that was preferred over the original acceptor splice site, because cDNA analysis showed inclusion of 14 nucleotides from intron 4 (c.288-14\_-1) into the mRNA. This mutation predicts p.Leu97ArgfsTer94 (Figure 1D).

The new acceptor splice site in patient 3 was also analyzed with splice prediction programs. As Table 1 shows, MaxEntScan and Genesplicer gave higher scores for the new acceptor splice site than for the original one, but SpliceSiteFinder-like and NNSPLICE did not. Hexplorer predicted about 60% enhanced ESE binding in the most 5' region of the new exon (Figure 4). Most importantly, the c.288-15C>G mutation affected the branch site presumably used by the w.t. pre-mRNA (Table 1).

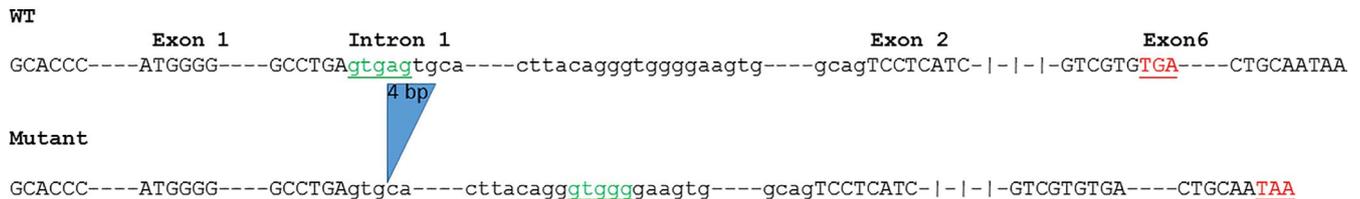
Analysis of the DNA of the parents of this patients indicated that her mother carried the c.246delC mutation whereas her father was a carrier of the c.288-15C>G mutation. Comparison of the w.t. and the mutant sequence in the parents' cDNA with those in the patient indicated that the stability of both mutant mRNAs was only slightly less than that of the w.t. mRNA (Figures 3A and 3B).

## 4 | DISCUSSION

### 4.1 | Activation of an exonic splice site and stabilization of the mutant mRNA

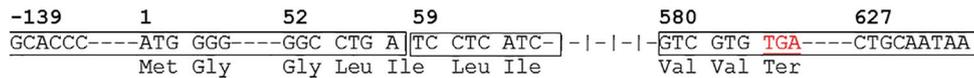
In patient 1 the low scores for the w.t. as well as for the activated cryptic donor splice site indicate that a weak splicing signal is present for this exon. Therefore, the binding of SR proteins to ESE binding sites will probably play an important role for correct splicing of this exon. Table 1 shows that the ratios of the PESE and PESS scores are 7.0, 9.9, and 4.0 for the whole exon 6 of *CYBB*, the 5' half (upstream of the mutation), and the 3' half (downstream of the mutation), respectively. These scores are relatively high compared with the average scores for constitutive exons (about 2), probably to compensate for the weak w.t. and cryptic donor splice site signals. Although the scores for the w.t. and activated cryptic donor splice site do not differ, the competition in the patient between the two sites for members of the splicing machinery is in favor

## Genomic DNA



## cDNA

## WT



## Mutant



**FIGURE 1B** Schematic representation of the consequences of the mutations in the patients compared to the wild-type situation. (B) Patient 2. The upper part represents the situation on the genomic level of the *CYBA* w.t. and the mutant sequence. The blue arrow points to the position of the 4-bp (gag) deletion in the donor splice site of intron 1 of the *CYBA* gene. Indicated in green are the locations of the donor and cryptic donor splice sites. Indicated in red are the w.t. termination codon and the downstream mutant termination codon, both in exon 6. The exon–exon boundaries are indicated by |–|–|. The lower part represents the situation on the cDNA and protein level. The w.t. and mutant sequence are shown. The blue triangle indicates the 79-bp insertion into the transcript of the patient due to the activation of the cryptic donor splice site downstream in intron 1. In red are the w.t. and the downstream mutant termination codon

of the cryptic donor splice site. The sizes of the flanking introns are 2,140 bp for intron 5 and 2,813 bp for intron 6. The larger the flanking intron size the more important the ESEs are for correct pre-mRNA splicing (Wu & Hurst, 2015).

The premature termination codon is located about half-way the coding region of this transcript. The curtailed mRNA seems to be quite stable, so nonsense-mediated decay (NMD) has not been activated in this case. Its termination codon lies across the border between the new exons 6 and 7, and therefore underneath the exon junction complex (EJC). Indeed, premature termination codons that lie within 50 bp distance from a splice junction escape from NMD (Neu-Yilik et al., 2011). However, there are other, more downstream EJCs present in this transcript that are also not removed by the first round of translation. These EJCs are thought to initiate NMD, but apparently this does not happen in this case, since the expression of the mutated transcript is at about the same level as the w.t. allele.

According to the ESE Finder website, this mutation is indeed creating a donor splice site, but the score for this splice site is below the default threshold and also lower than the w.t. donor splice site score (Table 1). This mutation is also damaging a binding motif for SRSF1 (serine/arginine-rich splicing factor 1) (Figure S3), one of the ESE binding proteins in human neutrophils (GeneCards® (www.genecards.org) and our own observations).

Missplicing of pre-mRNA can also be caused by other intra-exonic mutations. In another patient with X-CGD, we found a mutation in exon 5 of *CYBB* in a putative ESE site for SRSF1. As a result, 95% of the transcripts of this gene lacked exon 5, leading to a shift in the reading frame and premature termination codon (De Boer et al., 2017). This proves the importance of ESE binding for inclusion of the relevant exon in the mature mRNA.

## 4.2 | Activation of a cryptic intronic splice site and degradation of the mutant mRNA

In the genomic DNA of patient 2, we have found a 4-bp deletion in the donor splice site of intron 1 of *CYBA*, c.58+4\_7delAGTG. Due to this mutation, the score for the original donor splice site is lowered from 70.4 to below threshold. The score for the cryptic splice site is 71.3 (Table 1). By using the ESE Finder website, we found that the score for the w.t. donor splice is just below the default threshold value, but this donor splice site disappears due to the mutation (threshold set to 1). The value of the cryptic donor splice site is above threshold value (Figure S4). Because a terminal exon is involved in this case, a different splicing mechanism plays a role.

Terminal exon definition is not controlled by exonic splicing enhancers and silencers. The first exon is defined by the

## Genomic DNA

## WT

cccagGGGAC---GGCCCTTTA---CTCCTgtgag---cccagGCTCT---TACTGgtgag---ccgcagGCGGC---CGGTGA---GTTGA

## Mutant allele 1

cccagGGGAC---GGCCCTTTA---CTCCTgtgag---cccagGCTCT---TACTGgtgag---ccgcaGCGGC---CGGTGA

## cDNA

## WT

199	242	283	365	566	583
CGC TG G GGA C---GGG CCC TTT---CTC CT G CTC T---TA CTG GCG GC---CG GTG A---GTG <u>TGA</u>					
Arg Trp Gly	Gly Pro Phe	Leu Leu Leu	Leu Ala	Val	Val Ter

## Mutant allele 1

199	242	282	364	565
CGC TG G GGA C---GGG CCT TTA---C TCC T GC TCT---TAC TG G CCG C---CGG <u>TGA</u>				
Arg Trp Gly	Gly Pro Leu	Ser Cys Ser	Tyr Trp Arg	Arg Ter

**FIGURE 1C** Schematic representation of the consequences of the mutations in the patients compared to the wild-type situation. (C) Patient 3, allele 1. The upper part represents the situation on the genomic level of the *CYBA* w.t. and the mutant sequence. The blue arrow points to the position of the c.246G deletion, causing a shift in the reading frame. The lower part represents the situation on the cDNA and protein level. The w.t. and mutant sequence are shown. In red are the w.t. and the mutant termination codon

## Genomic DNA

## WT

CTCCTgtgag---gtgtccctgcctctcaccgctgtccccagGCTCT---TACTGgtgag---ccgcagGCGGC---CGGTGA---GTTGA

## Mutant allele 2

CTCCTgtgag---gtgtccctgcctctcagccgctgtccccagGCTCT---TACTGgtgag---ccgcagGCGGC---CGGTGA

## cDNA

## WT

283	365	566	583
CTC CT G CTC T---TA CTG GCG GC---CG GTG A---GTG <u>TGA</u>			
Leu Leu	Leu	Leu Ala	Val Val Ter

## Mutant a

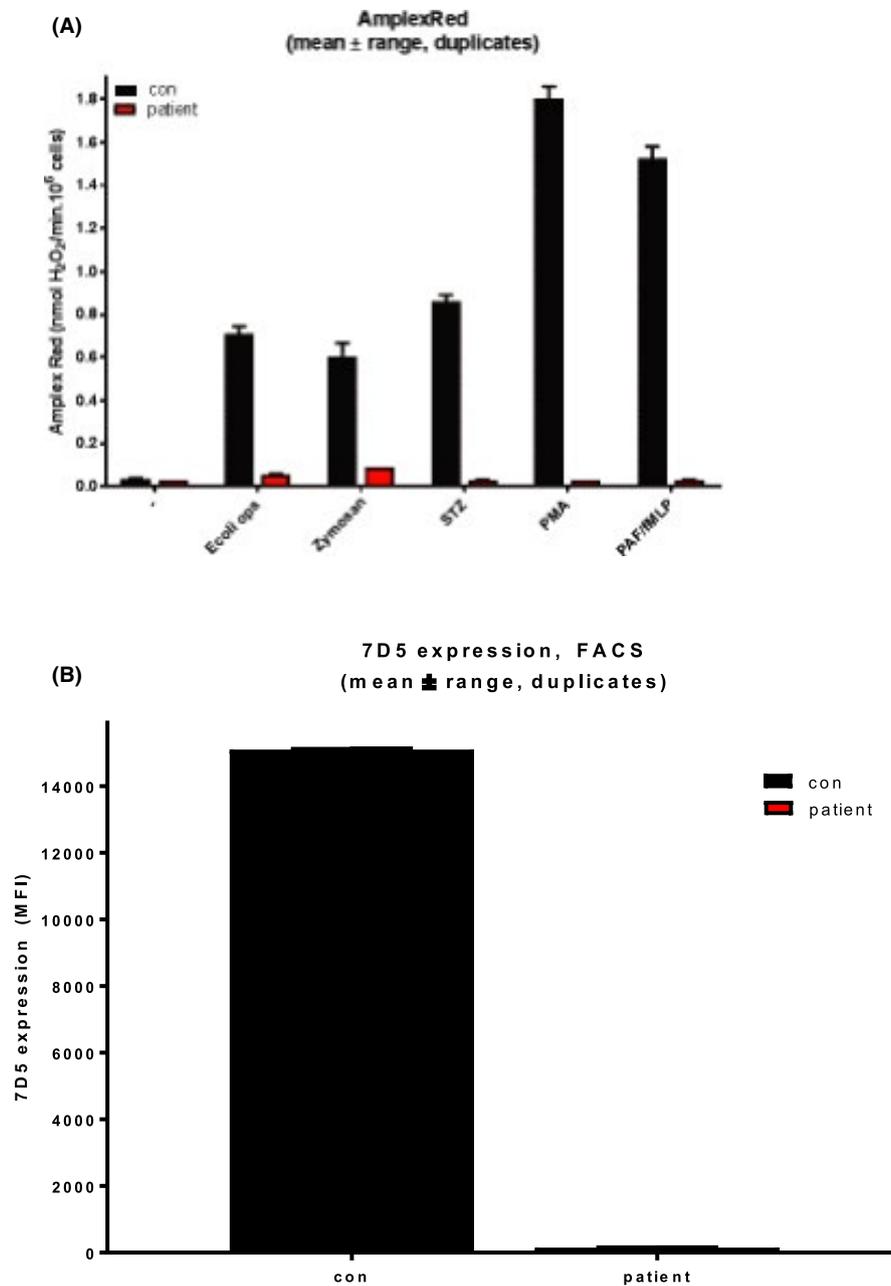
283	14 bp	379	580
CTC CT C CGC TGT CCC CCA GGC TCT---TAC TG G CGG C---CGG <u>TGA</u>			
Leu Leu	Arg Cys Pro Pro Gly Ser	Tyr Trp Arg	Arg Ter

**FIGURE 1D** Schematic representation of the consequences of the mutations in the patients compared to the wild-type situation. (D) Patient 3, allele 2. The upper part represents the situation on the genomic level of the *CYBA* w.t. and the mutant sequence. The blue arrow points to the position of the c.288-15C>G mutation in intron 4, causing insertion of 14 intronic nucleotides into exon 5. The lower part represents the situation on the cDNA and protein level. The w.t. and mutant sequence are shown. Highlighted in red are the w.t. and the mutant termination codon

interaction between the capsite-binding complex and the U1 snRNP (small nuclear ribonuclear protein) complex on the 5' donor splice site at the start of the first intron (Berget, 1995). The scores for splicing enhancer and silencer binding sites (Table 1) are very low, indicating that these values are not relevant for outsplicing of the first intron. Only the scores for the donor splice site of this intron are relevant in case of patient 2. The average internal exon size in vertebrates is

137 bp; internal exons longer than 400 bp make up less than 1%. Terminal exons are usually larger in size, especially the last exon containing the 3' untranslated region (UTR; Berget, 1995). In patient 2, the size of exon 1 is increased by 79 bp, raising the total length of this mutated exon 1 from 197 bp to 255 bp.

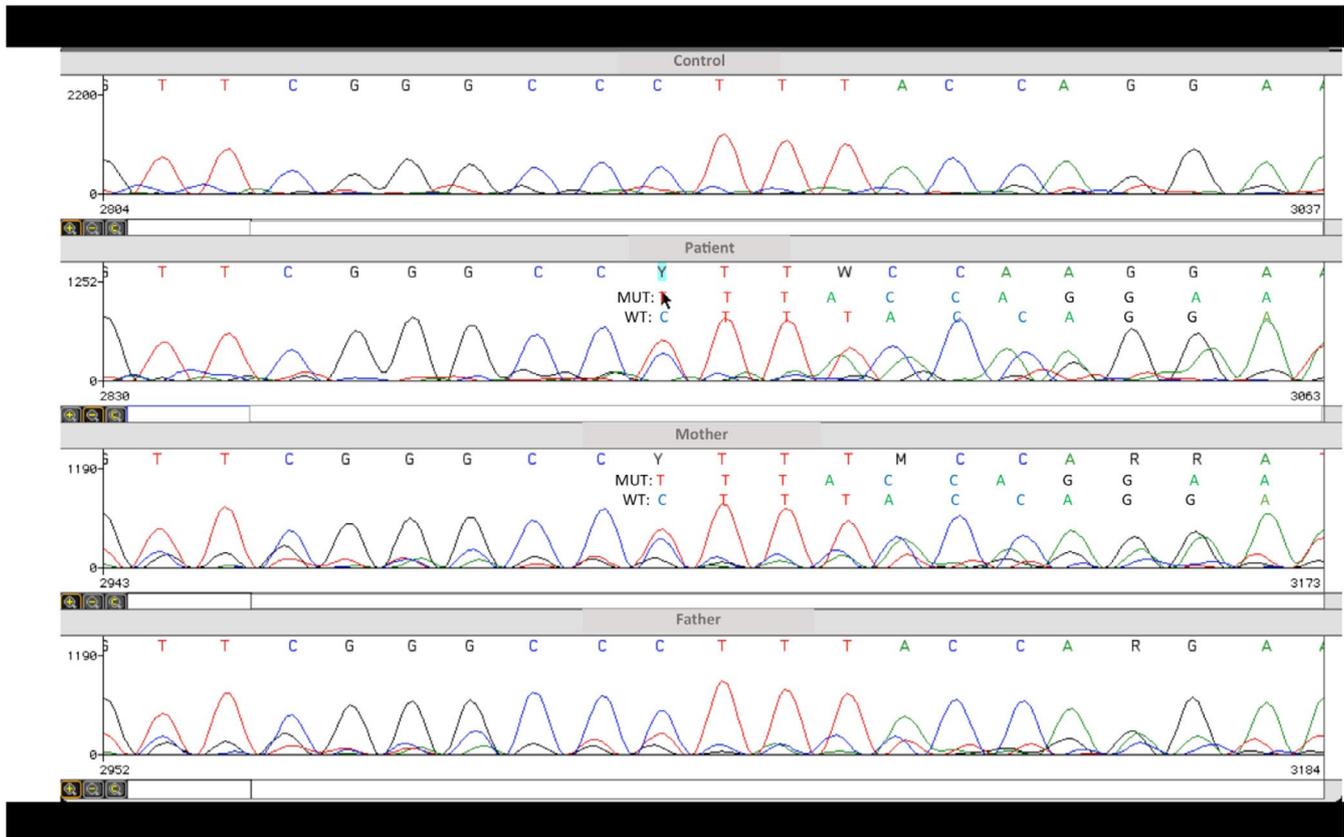
The translation termination codon in the mutated *CYBA* mRNA of patient 2 lies within the poly-adenylation signal in



**FIGURE 2** NADPH oxidase activity and gp91<sup>phox</sup>/p22<sup>phox</sup> expression by neutrophils from patient 3. (A) Results of Amplex Red assay. Neutrophils from the patient and from a control individual (con) were activated with serum-treated *E. coli*, zymosan, serum-treated zymosan (STZ), phorbol-myristate-acetate (PMA), or a combination of platelet-activating factor (PAF) and formyl-methionine-leucine-phenylalanine (fMLP). Results in mean and range of duplicate experiments with one cell sample. (B) Results of mAb 7D5 binding to neutrophils, analyzed by FACS. Means and range of duplicate assays with one cell sample are shown

the 3' UTR, rendering the protein longer than the w.t. protein, not only because of the 79 bp insertion from intron 1 but also by 47 extra bp from the 3' UTR. Together, this adds 42 amino acids to the protein, increasing it from 195 to 237 amino acids. In the cDNA of the parents of this patient, only w.t. sequence was found, while on the genomic level both parents were found to be carriers for the mutation. This indicates that the mutated transcript is labile. A consequence of the late

termination codon is a very short 3' UTR of just 23 bp between the termination codon and the poly(A) tail. According to Tanguay and Gallie (1996), this might explain the instability of this transcript, although the w.t. transcript is just 47 bp longer. However, when the poly(A) tail is located close to the stop codon of the mRNA, these investigators found in their model a more pronounced inhibitory effect on the translational efficiency in vivo, rather than decreasing the mRNA



**FIGURE 3A** Sequence analysis of the cDNA from patient 3 and her parents. *CYBA* c.246T until c.255G region in control, patient 3, mother, and father. The patient and her mother are heterozygotes for the c.246del mutation (arrow), and therefore show combinations of the w.t. and the mutated sequence, as indicated above the observed sequence

half-life in comparison with the other poly(A)<sup>+</sup> constructs with longer 3' noncoding regions. According to Gebauer and Hentze (2004) mRNA has a circular structure, in which the poly(A)-binding proteins (PABP) interact with proteins present in the cap complex. It is thought that this circular structure stabilizes the mRNA, and that the presence of these proteins on both sides of the mRNA molecule promotes the translation process. We speculate that in patient 2 steric hindrance may occur between the large ribosomal subunit and PABPs because of the short 3' UTR, thereby displacing some of the PABPs from the poly(A) tail. When less PABPs are bound to the poly(A) tail, protein synthesis will be downregulated, but also exonucleases might have a chance to bind to this “naked” poly(A) tail, initiate decapping of the mRNA molecule, and thereby initiate the mRNA degradation process.

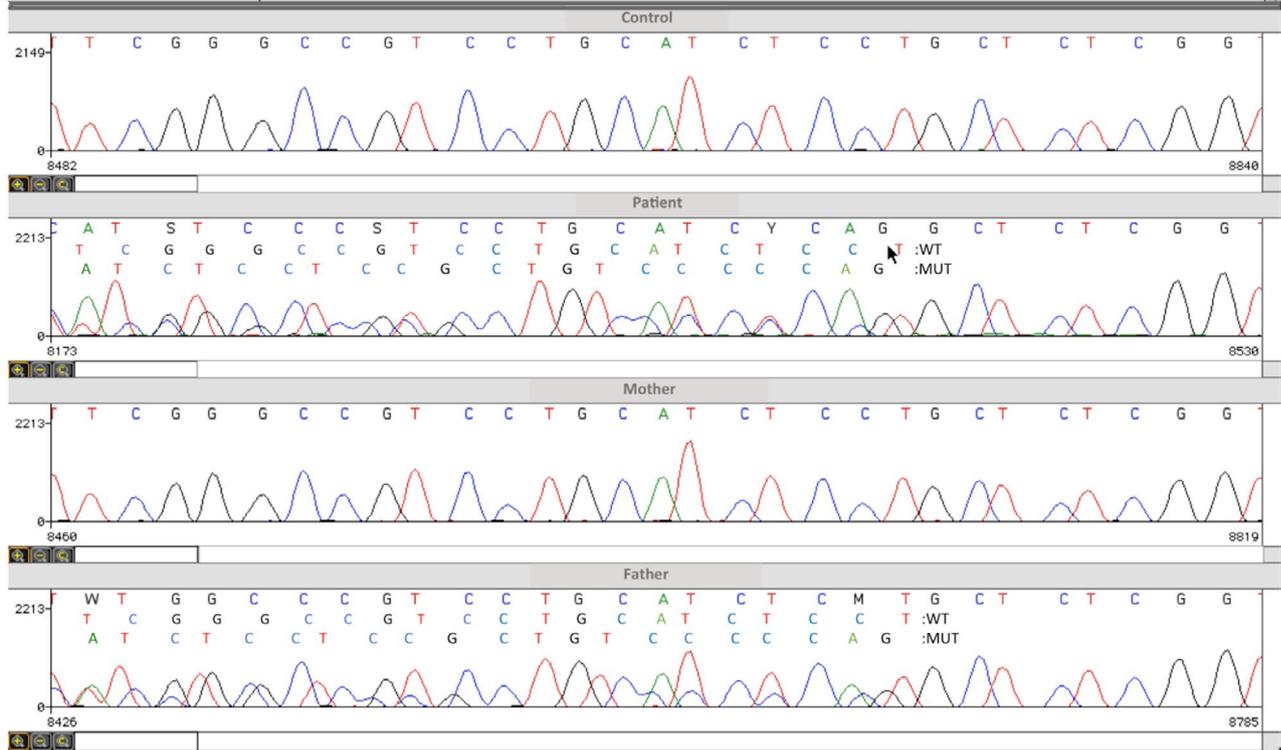
### 4.3 | Creation of an intronic acceptor splice site and stabilization of the mutant mRNA

Although the gp91<sup>phox</sup> expression in the neutrophils from patient 3 was absent (Figure 2), the causative mutation was found in *CYBA*, encoding p22<sup>phox</sup>. This is a consequence of the mutual stabilization of expression of these two proteins (Porter

et al., 1994). Since direct detection of p22<sup>phox</sup> in intact or permeabilized neutrophils is difficult, 7D5 binding is still a preferred assay, also in the autosomal forms of CGD. In case of negative 7D5 binding, mutations in *CYBB* have then to be excluded, which was done in this patient. In the genomic DNA of patient 3, we found a variation in *CYBA* in intron 4, close to the boundary with exon 5. This variation changed the sequence CACCCGCTG at position c.288-17\_-9 into CAG/CCGCTG, which was then apparently used as a new acceptor splice site, although the new splice site score was not more than 30% higher than that of the w.t. score according to some programs (Table 1). However, the mutation weakened or even destroyed a branch site (Table 1) that was probably used by the w.t. mRNA, and thus favored the missplicing. Branch site mutations are known to cause missplicing (Caminsky et al., 2015; Maslen, Babcock, Raghunath, & Steinmann, 1997). In this case, only Hexplorer indicated increased ESE scores in the new exon (Figure 4).

### 4.4 | Comparison of different pre-mRNA splicing prediction programs

Whole exome sequencing is now rapidly replacing conventional Sanger sequencing. This leads to neglect of



**FIGURE 3B** Sequence analysis of the cDNA from patient 3 and her parents. *CYBA* c.267T until c.294G region in control, patient 3, mother and father. The patient and her father are heterozygotes for the c.288-15C>G mutation, causing the inclusion of CCGCTGTCCCCCAG from intron 4 into the cDNA, as indicated above the observed sequence

possible intronic variations, which can nevertheless be important for identifying variations that may cause pre-mRNA missplicing (e.g., in patient 3). Therefore, it may be better to use WGS instead.

In general, it is difficult to predict whether or not a certain nucleotide variation in a gene will induce pre-mRNA missplicing. Proof for such a consequence can only be obtained from mRNA analysis, but RNA from relevant tissue is often difficult or even impossible to obtain. As such, blood cell diseases are easy to study, but still, RNA analysis is seldom performed. Thus, there is a growing need to predict the effect of a certain nucleotide variation on pre-mRNA splicing. A number of prediction programs for that purpose are available, but their reliability is under discussion (Ohno, Takeda, & Masuda, 2018; Soukarieh et al., 2016). Often, conclusions in this respect are based on results obtained with minigene assays, thus neglecting the effects of tissue-specific trans-acting factors such as splicing enhancing and splicing silencing proteins. We have now compared the reliability of a number of splicing predicting programs on three clinical cases with proven mRNA missplicing in the relevant tissue. The conclusion is that knowledge about the “strength” of a splice site as

such is insufficient to predict its use: the number and nature of cis elements to which trans-acting factors can bind must also be analyzed. However, even then this will not always lead to the right conclusion, as intactness of, for example, a suitable branch point is also important. The only general conclusion can be that all of these elements must be taken into account, but even then, prediction programs can only give an indication of the in vivo situation. Proof can still only be obtained by analysis of the mRNA from the relevant tissue.

## CONFLICT OF INTEREST

None declared.

## ACKNOWLEDGMENTS

We thank Dr. A.T.J. Tool (Sanquin Research) for performance of the experiments shown in Figure 2.

## ORCID

Dirk Roos  <https://orcid.org/0000-0001-5532-0202>



## REFERENCES

- Berget, S. M. (1995). Exon recognition in vertebrate splicing. *Journal of Biological Chemistry*, *270*, 2411–2414. <https://doi.org/10.1074/jbc.270.6.2411>
- Caminsky, N. G., Mucaki, E. J., & Rogan, P. K. (2015). Interpretation of mRNA splicing mutations in genetic disease: Review of the literature and guidelines for information-theoretical analysis. *F1000Research*, *3*, 282. <https://doi.org/10.12688/f1000research.5654.1>
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., & Krainer, A. R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Research*, *31*, 3568–3571. <https://doi.org/10.1093/nar/gkg616>
- De Boer, M., Bolscher, B. G., Dinauer, M. C., Orkin, S. H., Smith, E. C., Ahlin, A., ... Roos, D. (1992). Splice site mutations are a common cause of X-linked chronic granulomatous disease. *Blood*, *80*, 1553–1558.
- De Boer, M., Hartl, D., Wintergerst, U., Belohradsky, B. H., & Roos, D. (2005). A donor splice site mutation in intron 1 of CYBA, leading to chronic granulomatous disease. *Blood Cells Molecules and Diseases*, *35*, 365–369. <https://doi.org/10.1016/j.bcmd.2005.08.002>
- De Boer, M., van Leeuwen, K., Geissler, J., Belohradsky, B. H., Kuijpers, T. W., & Roos, D. (2017). Mutation in an exonic splicing enhancer site causing chronic granulomatous disease. *Blood Cells Molecules and Diseases*, *66*, 50–57. <https://doi.org/10.1016/j.bcmd.2017.08.012>
- De Boer, M., van Leeuwen, K., Geissler, J., Weemaes, C. M., van den Berg, T. K., Kuijpers, T. W., ... Roos, D. (2014). Primary immunodeficiency caused by an exonized retroposed gene copy inserted in the CYBB gene. *Human Mutation*, *35*, 486–496. <https://doi.org/10.1002/humu.22519>
- Dhir, A., & Buratti, E. (2010). Alternative splicing: Role of pseudoexons in human disease and potential therapeutic strategies. *FEBS Journal*, *277*, 841–855. <https://doi.org/10.1111/j.1742-4658.2009.07520.x>
- Fairbrother, W. G., Yeh, R. F., Sharp, P. A., & Burge, C. B. (2002). Predictive identification of exonic splicing enhancers in human genes. *Science*, *297*, 1007–1013. <https://doi.org/10.1126/science.1073774>
- Gebauer, F., & Hentze, M. W. (2004). Molecular mechanisms of translational control. *Nature Reviews in Molecular Cell Biology*, *5*, 827–835. <https://doi.org/10.1038/nrm1488>
- Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., ... Ast, G. (2006). Comparative analysis identifies exonic splicing regulatory sequences—The complex definition of enhancers and silencers. *Molecular Cell*, *22*, 769–781. <https://doi.org/10.1016/j.molcel.2006.05.008>
- Greer, K., Mizzi, K., Rice, E., Kuster, L., Barrero, R. A., Bellgard, M. I., ... Fletcher, S. (2015). Pseudoexon activation increases phenotype severity in a Becker muscular dystrophy patient. *Molecular Genetics and Genomic Medicine*, *3*, 320–326. <https://doi.org/10.1002/mgg3.144>
- Homolova, K., Zavadakova, P., Doktor, T. K., Schroeder, L. D., Kozich, V., & Andresen, B. S. (2010). The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to pseudoexon activation and causes the cblE type of homocystinuria. *Human Mutation*, *31*, 437–444. <https://doi.org/10.1002/humu.21206>
- Ishibashi, F., Nunoi, H., Endo, F., Matsuda, I., & Kanegasaki, S. (2000). Statistical and mutational analysis of chronic granulomatous disease in Japan with special reference to gp91-phox and p22-phox deficiency. *Human Genetics*, *106*, 473–481. <https://doi.org/10.1007/s004390000288>
- Itoh, H., Washio, T., & Tomita, M. (2004). Computational comparative analyses of alternative splicing regulation using full-length cDNA of various eukaryotes. *RNA*, *1*, 1005–1018. <https://doi.org/10.1261/rna.5221604>
- Kol, G., Lev-Maor, G., & Ast, G. (2005). Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Human Molecular Genetics*, *14*, 1559–1568. <https://doi.org/10.1093/hmg/ddi164>
- Maslen, C., Babcock, D., Raghunath, M., & Steinmann, B. (1997). A rare branch-point mutation is associated with missplicing of fibrillin-2 in a large family with congenital contractural arachnodactyly. *American Journal of Human Genetics*, *60*, 1389–1398. <https://doi.org/10.1086/515472>
- Neu-Yilik, G., Amthor, B., Gehring, N. H., Bahri, S., Paidassi, H., Hentze, M. W., & Kulozik, A. E. (2011). Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. *RNA*, *17*, 843–854. <https://doi.org/10.1261/rna.2401811>
- Ohno, K., Takeda, J., & Masuda, A. (2018). Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdisciplinary Reviews RNA*, *9*, e1451. <https://doi.org/10.1002/wrna.1451>
- Porter, C. D., Parkar, M. H., Verhoeven, A. J., Levinsky, R. J., Collins, M. K., & Kinnon, C. (1994). P22 phox deficient chronic granulomatous disease: Reconstitution by retrovirus-Mediated expression and identification of a biosynthetic intermediate of gp91-phox. *Blood*, *84*, 2767–2775.
- Schwartz, S. H., Silva, J., Burstein, D., Pupko, T., Eyraes, E., & Ast, G. (2008). Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Research*, *18*, 88–103. <https://doi.org/10.1101/gr.6818908>
- Soukarieh, O., Gaildrat, P., Hamieh, M., Drouet, A., Baert-Desurmont, S., Frébourg, T., ... Martins, A. (2016). Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using *in silico* tools. *PLoS Genetics*, *12*, e1005756. <https://doi.org/10.1371/journal.pgen.1005756>
- Tanguay, R. L., & Gallie, D. R. (1996). Translational efficiency is regulated by the length of the 3' untranslated region. *Molecular Cell Biology*, *1*, 146–156. <https://doi.org/10.1128/MCB.16.1.146>
- Voelker, R. B., & Berglund, J. A. (2007). A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Research*, *17*, 1023–1033. <https://doi.org/10.1101/gr.6017807>
- Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M., & Burge, C. B. (2004). Systematic identification and analysis of exonic splicing silencers. *Cell*, *119*, 831–845. <https://doi.org/10.1016/j.cell.2004.11.010>
- Wu, X., & Hurst, L. D. (2015). Why selection might be stronger when populations are small: Intron size and density predict within and between-species usage of exonic splice associated cis-motifs. *Molecular Biology and Evolution*, *32*, 18–29. <https://doi.org/10.1093/molbev/msv069>
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K., ... Frey, B. J. (2015). RNA splicing: The human splicing code reveals new insights into the genetic determinants of disease. *Science*, *347*, 1254806. <https://doi.org/10.1126/science.1254806>
- Yamaguchi, H., Fujimoto, T., Nakamura, S., Ohmura, K., Mimori, T., Matsuda, F., & Nagata, S. (2010). Aberrant splicing of the milk fat globule-EGF factor 8 (MFG-E8) gene in human systemic lupus

- erythematosus. *European Journal of Immunology*, 40, 1778–1785. <https://doi.org/10.1002/eji.200940096>
- Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Comparative Biology*, 11, 377–394. <https://doi.org/10.1089/1066527041410418>
- Yeo, G. W., Van Nostrand, E. L., & Liang, T. Y. (2007). Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genetics*, 3, e85. <https://doi.org/10.1371/journal.pgen.0030085>
- Zhang, X. H., & Chasin, L. A. (2004). Computational definition of sequence motifs governing constitutive exon splicing. *Genes and Development*, 18, 1241–1250. <https://doi.org/10.1101/gad.1195304>

## SUPPORTING INFORMATION

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

**How to cite this article:** de Boer M, van Leeuwen K, Hauri-Hohl M, Roos D. Activation of cryptic splice sites in three patients with chronic granulomatous disease. *Mol Genet Genomic Med*. 2019;7:e854. <https://doi.org/10.1002/mgg3.854>