GENOTYPE-PHENOTYPE CORRELATION

IN

JUNCTIONAL EPIDERMOLYSIS BULLOSA

by

David Wen

A thesis submitted to the University of Birmingham for the degree of
MASTER OF RESEARCH

Supervisors: Professor Iain L. Chapple and Professor Adrian H. M. Heagerty

Institute of Clinical Sciences
College of Medical and Dental Sciences
September 2022

ABSTRACT

Junctional epidermolysis bullosa (JEB) is a rare autosomal recessive genodermatosis characterised by mucocutaneous cleavage within the lamina lucida of the basement membrane zone following mild mechanical trauma. Genes implicated in JEB include *LAMB3*, *LAMA3*, *LAMC2* and *COL17A1*. A broad spectrum of JEB phenotypes exist, with severity ranging from death in infancy, to mild and localised blistering. Current genotype-phenotype paradigms are insufficient to accurately predict JEB subtype and characteristics from genotype, which has implications for clinicians, patients and their families. This study evaluated genetic and clinical findings from a cohort of individuals with JEB and systematically investigated genotype-phenotype correlations through bioinformatic analyses and comparison with mutations already reported in the literature.

Mutations were identified through Sanger sequencing, and were annotated and mapped to reference genes and proteins. Splice site mutations underwent analysis with the in-silico tools MMSplice and SpliceAI in order to predict resultant transcripts. A dedicated JEB deep phenotyping tool was developed which was used to systematically examine the clinical features of JEB individuals.

Eighteen unique mutations in *LAMB3*, *LAMA3*, *LAMC2* or *COL17A1* were identified from seventeen individuals (thirteen homozygotes and four compound heterozygotes). There were seven cases of severe JEB, nine intermediate JEB and one laryngo-onycho-cutaneous syndrome. Seven mutations were novel, and the majority of mutations identified in the cohort were found in *LAMB3* and were spread throughout the gene. *LAMB3* mutations included five splice site mutations. Functional effects predicted by the in-silico tools included exon skipping and activation of cryptic splice sites, which provided potential explanations for disease severity. RT-PCR and cDNA sequencing of one case confirmed the presence of a correctly predicted transcript, and also revealed an additional transcript generated through alternate splicing. Deep phenotyping was completed for all intermediate JEB cases and demonstrated substantial variation between individuals.

Reviewing the cases together, and with other mutations reported in the literature, this study highlights the importance of investigation at the RNA level for accurate phenotype prediction, and the need for a centralised database of pathogenic variants associated with JEB and corresponding phenotypes which would be a vital resource for genetic counselling and pre-natal diagnosis.

DEDICATION

Ling and An

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

INTRODUCTION

METHODOLOGY

RESULTS

JEB MUTATION LANDSCAPE AND GENOTYPE-PHENOTYPE CORRELATION

DISCUSSION

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AF   Anchoring Fibril

ASO   Anti-sense oligonucleotide

BEBS   Birmingham epidermolysis bullosa score

COL17   Type XVII collagen

DEB   Dystrophic epidermolysis bullosa

EB   Epidermolysis bullosa

EBS   Epidermolysis bullosa simplex

HGMD   Human Genome Mutation Database

IFM   Immunofluorescence microscopy

JEB   Junctional epidermolysis bullosa

LOC   Laryngo-onycho-cutaneous

NMD   Nonsense-mediated decay

PTC   Premature termination codon

RNA-seq   RNA sequencing

RT-PCR   Reverse transcriptase polymerase chain reaction

TEM   Transmission electron microscopy

INTRODUCTION

1.1 Epidermolysis Bullosa

Epidermolysis Bullosa (EB) encompasses a group of heterogeneous genetic mucocutaneous disorders characterised by skin and mucosal membrane fragility where blistering occurs following mild mechanical stress (Has et al., 2020a). Mutations in structural proteins result in compromised function and loss of adhesion between skin planes with ensuing blister formation within the cutaneous basement membrane zone, reflecting the ultrastructural location of the abnormal protein. Skin fragility often presents from birth or early childhood, and the term 'butterfly skin' has been coined to reflect how fragile EB patients' skin can be. Although rare, with an estimated incidence of approximately 1 in 50,000 live births (Fine, 2016), EB has a substantial negative impact on patients' quality of life, and is associated with significant healthcare costs and economic burden (Angelis et al., 2016).

Common complications secondary to extensive skin and mucosal blistering include pain, itch, infection, chronic ulceration, scarring, and risk of neoplasia, most commonly squamous cell carcinoma (SCC). Expression of mutated structural proteins in EB are not restricted to skin, and hence disease manifestations can involve other tissue types and organs. These include mucosal blistering and scarring of various tissue types which can lead to oesophageal strictures, ocular complications, stridor and life-threatening airway obstruction. Connective tissue defects can result in nail dystrophy, tooth enamel defects and alopecia (Fine and Mellerio, 2009a). Chronic skin damage, inflammation and regeneration often adversely affects other organ systems; resulting

manifestations include anaemia, osteoporosis, renal impairment, sepsis, irreversible contractures, pseudosyndactyly and resorption of digits (Fine and Mellerio, 2009b), with particular features being associated with specific subtypes.

There is currently no cure for EB; management requires a multi-disciplinary approach that is largely supportive and focused on wound management and symptomatic relief. Treatments including bone marrow transplantation, cell therapies and gene therapies are presently being explored (Uitto et al., 2018).

1.2 Classification of Epidermolysis Bullosa

Currently, EB has been classified into three main types according to the ultrastructural anatomical location of skin cleavage, and the localisation of the protein coded for by the mutated gene (Has et al., 2020a). These are Epidermolysis Bullosa Simplex (EBS), Junctional Epidermolysis Bullosa (JEB) and Dystrophic Epidermolysis Bullosa (DEB). Kindler Epidermolysis Bullosa (KEB) is an exception where the level of cleavage is variable, as this is caused by an abnormal ubiquitous focal adhesion protein. The characteristics of the main EB types are summarised in Table 1 (Has et al., 2020a). Further sub-classification of each type is based on molecular defects, mode of inheritance, clinical features and severity.

Rarer syndromic EB subtypes within the four main types have also been identified, where mutations in genes primarily result in extracutaneous organ involvement in addition to skin fragility. Manifestations include cardiomyopathy, muscular dystrophy, pyloric atresia, lung fibrosis and renal impairment (Fine and Mellerio, 2009b).

Level of skin cleavage	EB type	Inheritance	Mutated gene(s)	Targeted protein(s)
Intraepidermal	EB simplex	Autosomal dominant	*KRT5, KRT14*	Keratin 5, keratin 14
			PLEC	Plectin
			KLHL24	Kelch-like member 24
		Autosomal recessive	*KRT5, KRT14*	Keratin 5, keratin 14
			DST	Bullous pemphigoid antigen 230 (BP230) (syn. BPAG1e, dystonin)
			EXPH5 (syn. *SLAC2B*)	Exophilin-5 (syn. synaptotagmin-like protein homolog lacking C2 domains b, Slac2-b)
			PLEC	Plectin
			CD151 (syn. *TSPAN24*)	CD151 antigen (syn. tetraspanin 24)
Junctional	Junctional EB	Autosomal recessive	*LAMA3, LAMB3, LAMC2*	Laminin-332
			COL17A1	Type XVII collagen
			ITGA6, ITGB4	Integrin α6β4
			ITGA3	Integrin α3 subunit
Dermal	Dystrophic EB	Autosomal dominant	*COL7A1*	Type VII collagen
		Autosomal recessive	*COL7A1*	Type VII collagen
Mixed	Kindler EB	Autosomal recessive	*FERMT1* (syn. *KIND1*)	Fermitin family homolog 1 (syn. kindlin-1)

Table 1: Epidermolysis Bullosa Classification. Adapted from 'Consensus reclassification of inherited epidermolysis bullosa and other disorders with skin fragility' by C Has, JW Bauer , C Bodemer, MC Bolling, L Bruckner-Tuderman, A Diem, et al. 2020. Br J Dermatol.

1.3 Laboratory Diagnosis of Epidermolysis Bullosa

In addition to evaluation of clinical features, genetic testing may be used to detect pathological variants in implicated genes. This not only confirms the diagnosis, but can also aid in subclassification of EB and help predict likely disease course based on existing genotype-phenotype correlation paradigms (Has et al., 2020b). Methods include Sanger sequencing of suspected genes based on clinical suspicion, and with next generation sequencing becoming more accessible, targeted EB gene panels, whole exome sequencing (WES) and occasionally whole genome sequencing (WGS) are now used too (Has et al., 2020b, Wen et al., 2022). Genetic testing of other family

members can also uncover whether mutations were inherited or de novo, with implications for genetic counselling.

Diagnosis also commonly involves microscopy of biopsied skin samples. Immunofluorescence mapping (IFM), also known as antigen mapping, utilises monoclonal antibodies complementary to constitutive skin proteins for immunostaining, allowing visualisation of the exact plane of blister formation. Direct immunofluorescence to target protein domains can also reveal relative abundance of specific proteins (Has and He, 2016). Transmission electron microscopy (TEM) yields useful information regarding plane of skin cleavage and allows visualisation of the presence, abnormality or absence of cellular ultrastructural features such as desmosomes, hemidesmosomes and anchoring fibrils (AFs).

1.4 Junctional Epidermolysis Bullosa

Junctional epidermolysis bullosa (JEB) is one of the four main types of EB, and is defined by skin cleavage within the lamina lucida of the basement membrane zone when examined by TEM or IFM of biopsied skin samples (Fine et al., 2008). It is also characterised by the presence of mutations in one of seven JEB genes outlined in the following section (Has et al., 2020a).

Two main subtypes of JEB exist with varying clinical courses. Children born with severe JEB (previously known as Herlitz-JEB) often do not survive past the first few years of life, due to the severity of mucocutaneous blistering and extracutaneous manifestations. Mortality is often secondary to sepsis, dehydration, cardiac failure, and

metabolic disturbances, and management focuses on palliative care (Varki et al., 2006). Individuals with intermediate JEB (previous names include JEB generalised non-Herlitz, generalised atrophic benign EB, and JEB generalised other) follow a different clinical course and survive into adulthood, yet clinical manifestations are heterogenous, with varying severity and tissue types affected. In addition to cutaneous blistering, clinical features in JEB include dysmorphic or absent nails, chronic wounds, over-granulation and atrophic scarring, alopecia, upper airway obstruction, conjunctival blistering, corneal scarring, blistering of the oral mucosa, enamel hypoplasia, caries and anaemia (Fine et al., 2014, Fine et al., 2008). The incidence of JEB has been estimated at 2.68 per million live births. Prevalence is lower at 0.49 per million, likely reflecting the short lifespan of individuals with severe JEB (Fine, 2016).

1.5 Laryngo-onycho-cutaneous syndrome

The 151insG mutation in exon 39 of *LAMA3* has been identified as a founder mutation in patients with JEB laryngo-onycho-cutaneous (LOC) syndrome, a rare subtype of JEB. This mutation occurs in an exon specific to the laminin α3a isoform (*LAMA3A*), and through utilisation of an alternative downstream start codon, results in an N-terminally truncated polypeptide (McLean et al., 2003).

LOC syndrome has its own characteristic collection of features. These include nail dystrophy, tooth enamel defects and chronic production of granulation tissue, most commonly in the ocular region, nailbed and larynx. Complications from over-granulation at these sites include blindness and life-threatening airway obstruction requiring tracheostomy (McLean et al., 2003, Shabbir et al., 1986).

1.6 Key Genes and Proteins in Junctional Epidermolysis Bullosa

Transmission in JEB is autosomal recessive and currently, mutations in seven genes have been identified to result in JEB (Table 1). These are *LAMA3*, *LAMB3* and *LAMC2*, which encode the α3, β3 and γ2 chains of laminin-332, *COL17A1*, which encodes Type XVII collagen (COL17), *ITGA6* and *ITGB4*, which encode integrin α6β4, and *ITGA3*, which encodes the integrin α3 subunit.

The *LAMB3* gene, located on 1q32, is composed of 23 exons and produces a 4,014 nucleotide transcript (NM_000228.3) which is translated into the laminin β3 chain, a 1,172 amino acid long polypeptide. *LAMA3*, on 18q11.2, consists of 76 exons and generates multiple transcripts. The *LAMA3A* transcript (NM_000227.6) consists of exons 39–76 (5,618 nucleotides) which encode for the laminin α3a isoform (1,724 amino acids). The *LAMA3B* transcript (NM_198129.4) comprises exons 1–38 and 40–76 (10,651 nucleotides in total) which encode the longer laminin α3b isoform (3,333 amino acids) (McLean et al., 2003). *LAMC2* (1q25-q31), consists of 23 exons which produce a 5,398-long nucleotide transcript encoding a 1,193-long amino acid polypeptide, the laminin γ2 chain.

Laminin-332 is a heterotrimer composed of α3, β3, and γ2 chains and is secreted by basal keratinocytes to form anchoring filaments which bridge the lamina lucida of the basement membrane zone (Figure 1). It interacts with multiple cell surface receptors and extracellular structural proteins, including integrins α3β1 and α6β4, syndecans 1 and 4, type VII and XVII collagen, and nidogen 1 (Bardhan et al., 2020). It plays an

important structural role by linking epidermal proteins to the lamina densa and dermis, maintaining skin integrity (Has et al., 2018). In addition to its structural role, laminin-332 has also been shown to have other functions which include regulation of cell migration, adhesion, proliferation and wound healing (O'Toole et al., 1997, Nystrom et al., 2013).
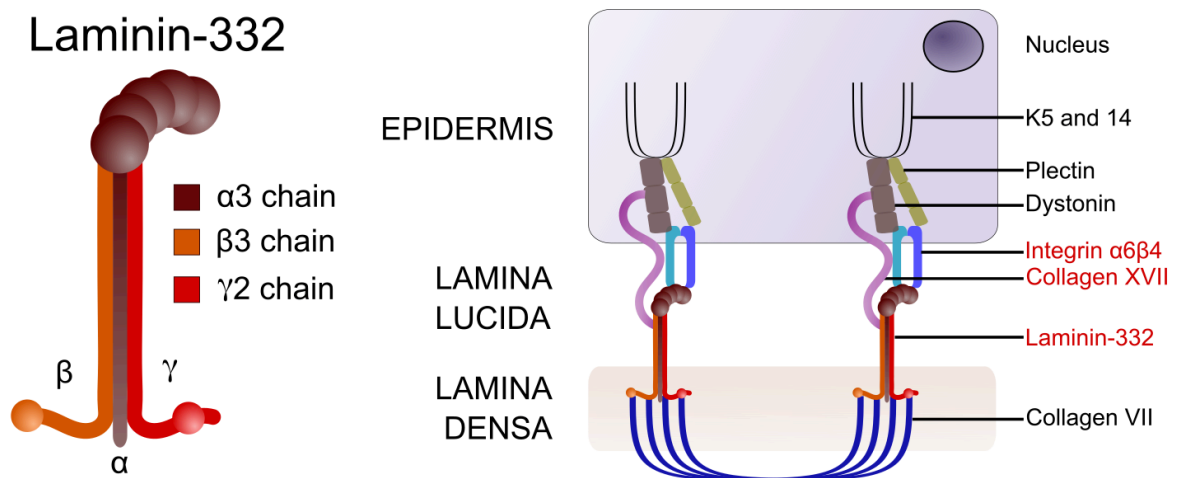
Figure 1: Schematic illustrations of laminin-332 and the basement membrane zone. K5 and 14 = keratins 5 and 14. Structural proteins implicated in JEB are highlighted in red.

The *COL17A1* gene, located on 10q25.1, produces a 5,612 nucleotide transcript (NM_000494.4) from 65 exons which is translated into the collagen type XVII α chain, a 1,497 amino acid polypeptide. Homotrimerisation of three α chains forms type XVII collagen protein (COL17), also known as BP180 and BPAG2. This transmembrane protein is a key component of the hemidesmosome and has a structural role in linking the epidermis to the basement membrane and dermis; its intracellular N-terminal endodomain interacts with plectin, dystonin and the integrin β4 subunit (Koster et al., 2003), whilst its C-terminal ectodomain interacts with laminin-332 and the integrin α6 subunit (Has et al., 2018). Further roles include regulation of epidermal proliferation

(Watanabe et al., 2017), cell migration (Jackow et al., 2016), and hair follicle development (Matsumura et al., 2016).

Reviewing mutations associated with JEB that have been reported in the literature (HGMD Professional 2021.1), the greatest number are associated with *LAMB3* (123 mutations), followed by *COL17A1* (114 mutations), *LAMA3* (54 mutations) and *LAMC2* (48 mutations), shown in Appendix 1 (p2–4). Common mutation types included small deletions (93 mutations), insertions (37 mutations), nonsense (90 mutations), missense (38 mutations) and splice site mutations (65 mutations).

In addition to *LAMA3*, *LAMB3*, *LAMC2* and *COL17A1*, mutations in *ITGA6*, *ITGB4* and *ITGA3* have also been found to result in rarer subtypes of JEB, such as JEB with pyloric atresia and JEB with interstitial lung disease and nephrotic syndrome (Has et al., 2020a). Exploration of mutational spectrum and phenotypic consequences of variants of these integrin genes is beyond the scope of this thesis.

1.7 Genotype-phenotype correlation in JEB

Currently, it is believed that the severity of functional defect arising from genetic mutations results in the two divergent presentations of severe JEB and intermediate JEB, which have different clinical courses and prognoses (Has et al., 2020a). In the majority of cases, severe JEB results from biallelic PTC-introducing loss-of-function mutations in *LAMA3*, *LAMB3* or *LAMC2*, which result in nonsense-mediated decay (NMD) of mRNA transcripts, or production of truncated non-functional polypeptides (Nakano et al., 2002).

PTCs can be introduced from nonsense mutations, frameshift indel mutations, or splice site mutations resulting in production of out-of-frame transcripts. The absence of any of the three polypeptide chains of laminin-332 disrupts its assembly and secretion from basal keratinocytes (Has et al., 2018, Hammersen et al., 2016), and these patients are generally found to have absent staining for laminin-332 on immunofluorescence with monoclonal antibodies such as GB3 (Heagerty et al., 1986, Schofield et al., 1990).

In contrast, most intermediate JEB cases with mutations in the laminin-332 genes result from compound heterozygosity for a PTC-introducing mutation and a non-loss-of-function mutation, or biallelic non-loss of function mutations, such as missense mutations, in-frame deletions, or splice site mutations generating in-frame transcripts (Nakano et al., 2000, Has and Bruckner-Tuderman, 2014). Indeed, it has been demonstrated that only a small amount of functional or partially functional laminin polypeptide (5–10% of residual protein) can considerably ameliorate disease severity (Has et al., 2020a).

Biallelic PTC-introducing mutations in *COL17A1* have usually been found to result in more severe forms of intermediate JEB with generalised blistering and nail dystrophy (Kiritsi et al., 2011, Has et al., 2018). Other mutations, such as missense, in-frame insertions or deletions (indels), or splice site mutations which preserve the reading frame have generally been associated with milder intermediate JEB phenotypes where disease is localised or where skin and extracutaneous involvement was only mild (Condrat et al., 2018, Kiritsi et al., 2011, Herisse et al., 2021).

However, these paradigms are not generalisable to all JEB cases. A small minority of JEB patients have been reported to have biallelic PTC-introducing mutations with a corresponding intermediate JEB phenotype rather than the expected severe JEB phenotype. Sequencing of one individual with intermediate JEB revealed compound heterozygosity for a nonsense (p.Q834X) and frameshift mutation (c.29insC, resulting in a downstream PTC) in both *LAMB3* alleles. This PTC/PTC combination would be expected to result in the absence of laminin-332, severe JEB and early mortality. However, the α3 and γ2 chains of laminin-332 were found to be present (although reduced) on immunofluorescence (McGrath et al., 1999). RT-PCR of transcripts demonstrated that in-frame exon skipping spliced out exon 17 which contained the nonsense mutation (p.Q834X), resulting in expression of a partially functional protein, resulting in a less severe phenotype for the patient and a longer than expected lifespan (McGrath et al., 1999).

In-frame skipping of the mutation-bearing exon in an individual with a homozygous R245X mutation in *LAMC2* similarly resulted in an intermediate JEB phenotype, rather than the expected severe JEB phenotype (Nakano et al., 2002). Examples of in-frame skipping of mutation harbouring exons have also been reported in JEB cases with biallelic PTC mutations in *COL17A1*, where the individuals were reported to have relatively mild JEB intermediate phenotypes (Herisse et al., 2021, Ruzzi et al., 2001).

Spontaneous readthrough of PTCs has been shown to be another mechanism where functional protein can be produced from a nonsense mutation to result in a milder than expected phenotype. The surrounding sequence flanking a specific TGA PTC in

LAMA3 allowed PTC readthrough to occur and generate a full length *LAMA3* transcript, which resulted in a very mild JEB phenotype in an individual with biallelic nonsense mutations (R943X/R1159X) in *LAMA3* (Pacho et al., 2011).

These uncommon but important cases highlight that more complex mechanisms may influence gene expression, and that more detailed molecular characterisation is required for accurate genotype-phenotype correlation.

1.8 Nonsense-mediated decay

Many disease-causing mutations in JEB introduce PTCs. The NMD pathway is a surveillance mechanism which degrades mRNA transcripts containing PTCs, preventing translation of aberrant and potentially harmful truncated proteins (Lindeboom et al., 2016). The canonical pathway of NMD is mediated through exon junction complexes (EJCs) which are multi-subunit complexes deposited at exon-exon junctions during mRNA splicing by the splicing machinery (Hug et al., 2016). In the absence of upstream PTCs, EJCs are usually displaced from mRNA transcripts by the translating ribosome. However, the presence of upstream PTCs, for example through mutation, can prevent EJC displacement by the translating ribosome (Hug et al., 2016). EJCs remaining bound to the transcript interact with cellular degradation factors to mediate NMD. Further non-canonical mechanisms of NMD which are independent of the EJC have also been suggested, and are currently under exploration (Hug et al., 2016). Factors affecting whether NMD occurs are currently under investigation; suggested influencing characteristics include proximity of the PTC to the translation

start site, proximity of the PTC to the last exon junction, exon size and mRNA half-life (Lindeboom et al., 2016).

The extent of which NMD occurs in PTC-containing transcripts is important to establish as this influences the balance of mRNA transcription and degradation (Kivirikko et al., 1996), which in turn affects whether a biologically significant amount of truncated protein is produced.

1.9 Pre-mRNA splicing

Many of the reported mutations in *LAMA3*, *LAMB3*, *LAMC2* and *COL17A1* are splice site mutations which affect mRNA splicing. This process involves the removal of pre-mRNA introns to form mature mRNA in a reaction facilitated by the spliceosome, a multi subunit RNA-protein complex (Montes et al., 2019). Splicing of different combinations of introns and exons (alternate splicing) allows production of multiple different protein isoforms from the same gene, allowing increased proteomic complexity.

Figure 2 outlines the canonical mechanism of splicing. Briefly, the 2' OH group of an adenosine within the intron (the branch point) acts as a nucleophile to attack the guanine at the 5' border of the intron to form a lariat (outlined in Figure 2 as step 1). In step 2, the exposed 3' OH of the upstream exon attacks the 3' intron border to ligate the exons and release the lariat, which is subsequently degraded (Montes et al., 2019). Intronic border regions are key to defining splice sites and are commonly marked by a GT dinucleotide at the 5' splice site, and an AG dinucleotide at the 3' splice site

(Chambon's rule), which act as binding regions for key spliceosomal snRNAs (Montes et al., 2019, Breathnach et al., 1978).
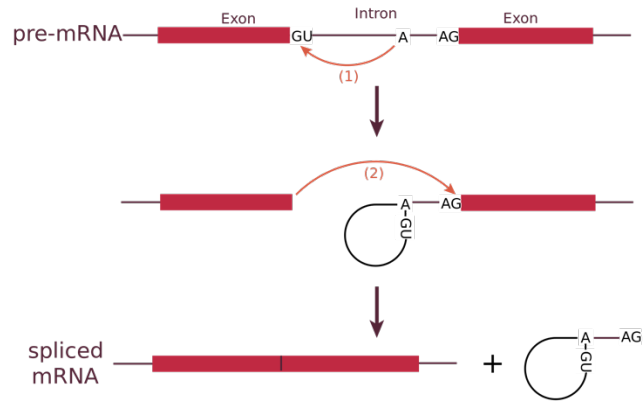
Figure 2: Mechanism of splicing.
Image from: https://en.wikipedia.org/wiki/RNA_splicing (CC BY-SA 4.0)

Regulation of splicing is essential, and in addition to the conserved splice site dinucleotide motifs, splice site enhancers and silencers have been found in both coding (exonic splicing enhancers and exonic splicing silencers) and non-coding regions of genes (intronic splicing enhancers and intronic splicing silencers) (Wang and Burge, 2008). These sequences act as binding sites for proteins involved in splicing regulation and affect splicing likelihood and efficiency (Zhu et al., 2001).

Mutations in conserved splicing sequences or enhancer and silencer motifs can result in disruption of normal splicing. Consequences include exon skipping, or activation of cryptic splice sites which can lead to inclusion or exclusion of introns or exons (Figure 3). The resulting consequence depends on the exact location of the gene or resulting protein affected, and whether an in-frame or out-of-frame transcript is produced.
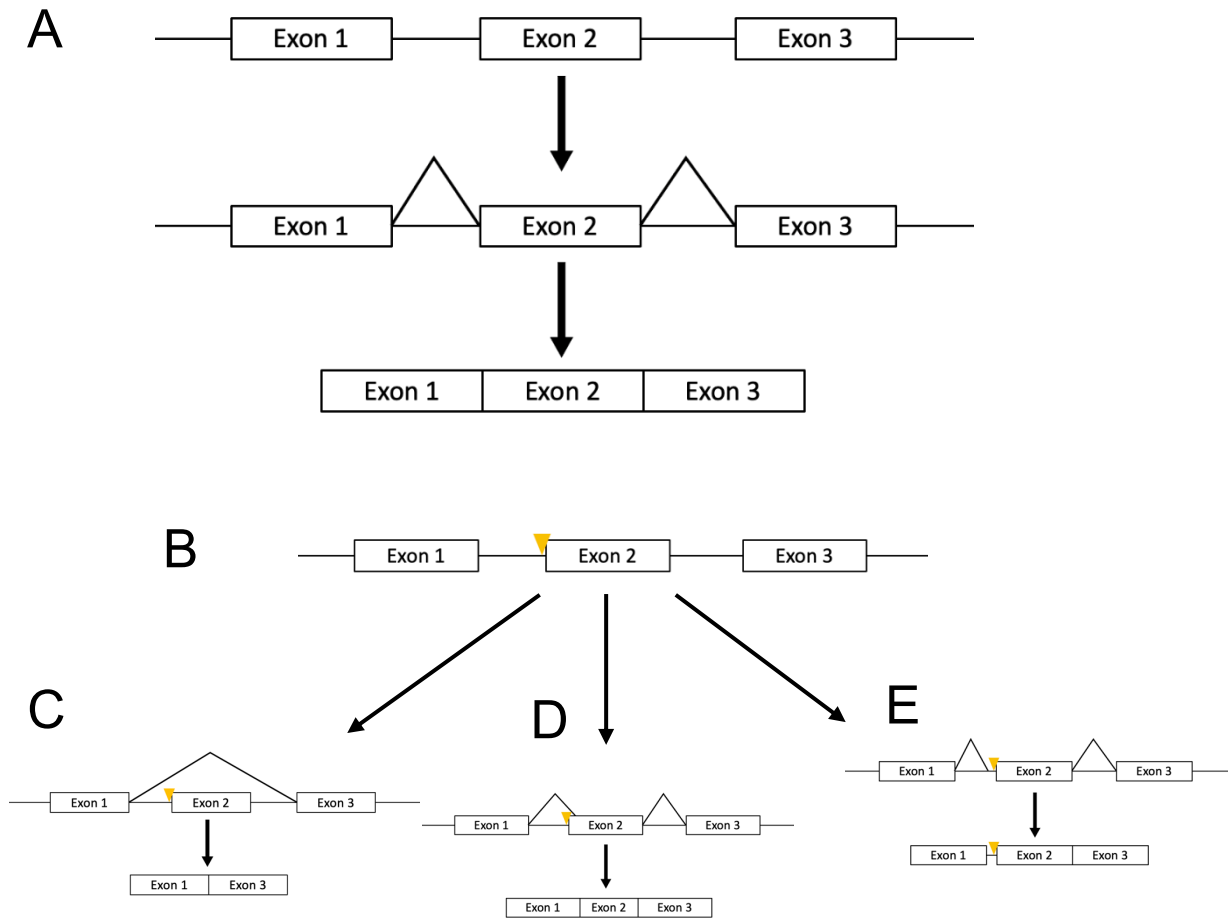
Figure 3: Potential splice site mutation outcomes. A) Normal splicing. B) A splice site mutation (marked with a yellow triangle) can result in: C) exon skipping. D) cryptic splice site activation within an exon which leads to loss of nucleotides from exon 2. E) cryptic splice site activation within an intron which leads to inclusion of nucleotides from intron 1.

In JEB, examples of in-frame exon skipping in *LAMB3* have been demonstrated to produce partially functional polypeptide and a resultant intermediate JEB phenotype (Kiritsi et al., 2015). In contrast, homozygous splice site mutations resulting in out-of-frame exon skipping and premature stop codons have been shown to result in the lethal phenotype of the disease (Vailly et al., 1995). This highlights the importance of sequence alignment and prediction tools for genotype-phenotype prediction following splice site mutations, where severity of splice site mutation outcomes has direct clinical relevance.

1.9 Clinical relevance of accurate genotype-phenotype correlation in JEB

Clinical presentation between intermediate JEB, severe JEB and other EB types is often similar at birth, and therefore genetic diagnosis with prediction of phenotype from laboratory results are critical for providing accurate prognostication to patients and their families (Has and Bruckner-Tuderman, 2014). Furthermore, phenotype prediction from genetic results of prenatal testing (e.g., via chorionic villus sampling or amniocentesis) may also inform parental decision making. Accurate prediction of corresponding phenotypes will allow patients, their families, clinicians and services to appropriately plan for the future.

As described, splice site mutations can result in severe or intermediate JEB depending on the resulting transcripts. Confirmation of splice site mutation consequences requires cDNA analysis through RT-PCR, which is currently not a routine investigation in UK EB genetic services. Furthermore, rare cases carrying homozygous or compound heterozygous PTC-causing mutations in the laminin-332 genes are occasionally associated with a milder than expected intermediate JEB phenotype and survival to adulthood. This is in contrast to an expected severe JEB phenotype and infant mortality. Prediction of these cases is not well established currently. Additionally, there is wide variation in phenotypic severity and affected tissue types within intermediate JEB which warrants further exploration and correlation with genotype.

Based on the above, it has become apparent that phenotype prediction is more complex than suggested by classical paradigms, and that further investigation and

more accurate approaches are required for the prediction of clinical course and features from genetic information provided through genetic tests.

In-silico tools can be used to predict the molecular consequences of mutations, such as pre-mRNA splicing, and can be used to map mutations and predict protein domains from primary sequence. This can lead to a greater understanding of disease mechanisms, and potentially improve prediction of clinical course and outcome.

1.10 Study aims

This study aims to evaluate genotype-phenotype correlations in a regional JEB patient cohort, and to provide explanations for the relationships. This will be achieved by:

1. Retrospectively collecting genetic information from a cohort of JEB patients from paediatric and adult EB specialist services.
2. Systematically annotating genetic variants.
3. Visualisation of data through mapping variants onto reference genes, transcripts and proteins.
4. Systematically analysing potential consequences of mutations on mRNA transcripts using in-silico tools.
5. Developing a pilot deep phenotyping tool which can be used to systematically characterise the clinical features of JEB patients.
6. Deep phenotyping patients to characterise their clinical features and disease severity.

7. Correlating genetic findings and bioinformatic predictions with phenotype, and comparing these with cases reported in the literature.

Overall, this project will help in establishing links between genetic defects and clinical characteristics, which is essential for accurate prognostication, genetic counselling and prenatal diagnosis.

METHODOLOGY

2.1. Setting and Subjects

Participants for this study were recruited from the adult EB unit at Solihull Hospital (University Hospitals Birmingham NHS Foundation Trust) and the paediatric EB unit at Birmingham Children's Hospital (Birmingham Women's and Children's NHS Foundation Trust).

Participants were eligible for inclusion in the study if they had a diagnosis of JEB or LOC syndrome, confirmed through genetic testing with a mutation present in *LAMA3*, *LAMB3*, *LAMC2* or *COL17A1*. Current patients, or those who used the EB service at Solihull hospital or Birmingham Children's Hospital within the last 5 years were eligible for inclusion into the study. This included patients who are living and also those who passed away in the last 5 years, for example from complications of severe JEB. Patients who were unable to give informed consent, or did not have anyone available who could consent on their behalf were excluded from the study. Additionally, JEB patients with mutations in *ITG3*, *ITGA6* and *ITGB4* were excluded.

All living adult patients gave informed consent to be enrolled into the study after reading specially designed participant information sheets (Appendix p5–16). Parents of children included in the study provided consent on their behalf. Assent was sought from children where possible if they had sufficient understanding of the study. Consent was not sought from the next of kin of deceased patients as data collection was retrospective and routine clinical data was available. The study was approved by the

Health Research Authority with IRAS reference number 290183 and was sponsored by University Hospitals Birmingham NHS Foundation Trust.

2.2 Genotyping

Candidate genes for sequencing were selected based on results from immunofluorescence or clinical suspicion. For example, splits at the DEJ with absence of staining for laminin-332 prompted sequencing of *LAMB3*, *LAMC2* and *LAMA3*, whereas absence of staining for collagen 17 prompted sequencing of *COL17A1*. Blood or saliva samples from patients underwent Sanger sequencing by the National Diagnostic EB (NDEB) laboratory at St Thomas' Hospital (Guy's and St Thomas' Hospital NHS Foundation Trust). The Sanger sequencing protocol is outlined in the Appendix (p17). Unless targeted sequencing could be performed (e.g. if the proband had a sibling who was already affected), all exons of the gene were sequenced, along with 20 bases of flanking intron in order to detect any potential splice site mutations. WES was used for one individual, as Sanger sequencing did not detect mutations in *LAMA3*, *LAMB3*, *LAMC2* or *COL17A1* initially. Patients' medical records and investigations were reviewed for genetic reports to identify pathogenic variants. Transcript accession numbers for identified variants were obtained from the NDEB laboratory.

2.3 Variant annotation

Identified pathogenic variants were mapped to the reference genome (GRCh 38) manually using the Ensembl genome browser (Hunt et al., 2018). Relevant details such as genomic coordinates, the affected exon or intron, and amino acid changes (if

applicable) were determined. Existing general databases of genetic variants (ClinVar and HGMD Professional 2021.1) were searched to determine whether the mutation had been reported previously (Landrum et al., 2018, Stenson et al., 2020).

2.4 Data visualisation

Schematics of genes and transcripts were constructed using genomic coordinates of the Homo sapiens genome assembly GRCh38 (hg38) imported from the Ensembl database (Howe et al., 2021). Mutations identified in each gene were mapped onto gene, transcript and protein schematics with Inkscape Version 1.0.2 (Inkscape Project, Brooklyn, New York, USA) and an open-source R script (Turner, 2015), in order to visualise and compare their locations. As both *LAMA3* transcripts (NM_000227.6 and NM_198129.4) are expressed in skin (Kiritsi et al., 2013), mutations were mapped onto both of these transcripts, and the isoforms they encode. For protein schematics, domains were predicted using Interpro with e-values cut off at <1E^{-3} (Blum et al., 2021).

2.5 Prediction of nonsense-mediated decay

The **MA**ke **S**ense **O**f **NMD** (MasonMD) is a tool which acts as a decision tree and utilises three rules to predict whether mutated transcripts containing PTCs undergo NMD (Hu et al., 2017). The MasonMD algorithm maps PTCs to transcripts and predicts NMD escape if any of the three rules are satisfied:

1. The gene contains one exon only and no exon-exon junctions
2. The PTC is located within 200bp downstream of the transcription start site / start codon.

3. The PTC is located in the last exon of the transcript, or within 50bp of the last exon-exon junction.

Nonsense and frameshift mutations identified in the cohort underwent analysis by MasonMD to predict whether NMD would occur.

2.6 Splice site mutation analysis

Several open access splice site mutation analysis tools were reviewed. Some tools such as NetGene2 (Brunak et al., 1991), NNSplice (Reese et al., 1997), MaxENTscan (Yeo and Burge, 2004) and ASSP (Wang and Marin, 2006) required manual insertion of FASTA sequences. This is in contrast to other tools (SpliceAI, MMsplice and SPANR), which can utilise genomic coordinates of variants in a variant call format (.vcf) file as input, which allows analysis of multiple annotated variants in parallel. Unfortunately, SPANR (Xiong et al., 2015 http://tools.genes.toronto.edu) was unavailable for use during the analysis period (April 2021–April 2022).

On direct comparison performed by the developers of each tool, MMSplice outperformed SPANR and MaxENTScan (Cheng et al., 2019), whilst SpliceAI outperformed both NNSplice and MaxENTScan (Jaganathan et al., 2019). Therefore, MMSplice and SpliceAI were chosen to investigate the splice site mutations in our cohort due to superior performance, and also due to their potential for integration into next generation sequencing bioinformatics pipelines, where multiple variants may need to be investigated in parallel in an automated fashion, made possible through appropriate formatting of mutations in a .vcf file.

2.6.1 MMsplice

MMsplice (modular modeling of splicing) is a splice variant analysis tool which predicts pathogenicity of splice site variants and their effects on exon skipping (Cheng et al., 2019). It consists of six individual neural network modules which score six different sequence areas surrounding the exon-intron junction (splice donor site, splice acceptor site, exon 5' and 3' ends, intron 5' and 3' ends). Modules were trained on the GENCODE reference annotation dataset (Frankish et al., 2019) and a massively parallel reporter assay (MPRA) dataset (Rosenberg et al., 2015). The neural network architectures for each module are described in detail by the authors (Cheng et al., 2019). Individual modular scores are then combined in a linear model which underwent further training with the Vex-seq dataset (Adamson et al., 2018) to produce a splice analysis tool which predicts the effects of variants on exon skipping. The output is given as a delta logit (log odds ratio) percentage spliced in (PSI) score, which represents the proportion of transcripts containing a specific exon, compared to transcripts where the exon has been skipped. The authors state that scores of less than -2 or more than +2 predict that the variant has a strong effect. A pathogenicity score from 0 to 1 is also given which is based on the individual modular scores and a logistic regression model from further training data from the ClinVar database.

2.6.2 SpliceAI

SpliceAI is a 32-layer convolutional neural network (CNN) which examines 10,000nt of flanking sequence around a variant of interest to predict its effects on RNA splicing. This model was trained using splice junction annotations from GENCODE (Frankish et al., 2019) and Genome Tissue Expression (GTEx) datasets (Consortium, 2013). When

used by the developers to predict splice junctions in reference pre-mRNA transcripts, it achieved 95% accuracy (Jaganathan et al., 2019).

From variant data formatted as a .vcf file, four delta scores are calculated, which represent the change in probability of a specific position being a splice donor or acceptor. These are:

1. Donor loss (DL): The decrease in probability of a position being used as a splice donor
2. Acceptor loss (AL): The decrease in probability of a position being used as a splice acceptor
3. Donor gain (DG): The increase in probability of a position being used as a splice donor
4. Acceptor gain (AG): The increase in probability of a position being used as a splice acceptor

Distance from the variant for each position with a change in splicing probability is also given.

When used to predict the consequences of splice altering variants, 75% of splice junction predictions with delta scores >0.5 validated against RNA-seq data from GTEx (Jaganathan et al., 2019). Higher delta scores were associated with higher validation rates, and also larger effect sizes (the proportion of novel transcripts produced). Furthermore, accuracy was much greater than three other splice analysis tools (MaxEntScan, NNSplice and GeneSplicer), (Pertea et al., 2001, Yeo and Burge, 2004, Reese et al., 1997).

SpliceAI allows the user to set the maximum distance (-D) between the variant and a potential gain or loss site. The default distance is 50 nucleotides, and the maximum distance is 4999 nucleotides upstream and downstream of the variant. To investigate whether prediction scores for gain or loss sites changed with -D, recurrent analyses were performed with different parameters of -D at 50, 100, 150, 200, 250, 350 and 4999 nucleotides.

2.6.3 Analysis of splice site prediction outcomes

Nucleotide FASTA files of mutant sequences were created by manual annotation according to the predicted splice site mutation consequences. Nucleotide sequences were translated with ExPASy (Gasteiger et al., 2003) and aligned with the reference amino acid sequence using NCBI protein blast (Gish and States, 1993).

2.6.4 Laboratory analysis of RNA transcripts

One case underwent further RNA analysis through RT-PCR which was completed by the NDEB laboratory.

2.7 Immunofluorescence mapping

Patients who had skin biopsies underwent immunofluorescence mapping (IFM) by the NDEB lab to confirm the exact level of blistering, and also to detect the relative amounts of target proteins present. These results were also reviewed if routinely available. Blistered skin was biopsied and sent to the NDEB lab for immunolabelling with GB3 and P3H9-2 antibodies for laminin-332, and COL94 antibody for collagen IV. The full IMF protocol is outlined in the Appendix (p18).

2.8 Phenotyping of Junctional EB cohort

2.8.1 Deep phenotyping of living Junctional EB patients

The phenotype of JEB patients was prospectively characterised in detail through systematically recording affected areas, tissue types and severity. A number of objective scoring systems have been developed and validated for use with EB patients. The Birmingham Epidermolysis Bullosa Severity (BEBS) is a validated tool for assessing disease involvement and severity in specific body sites, including the nails, hair, eyes, mouth, larynx, oesophagus and hands (Moss et al., 2009). Total area of affected skin, chronic wounds and nutritional compromise are also reviewed. As this score was designed to assess all four types of EB, rather than specifically JEB, a pilot phenotyping tool specifically for JEB was developed from the existing BEBS.

The three previous EB expert consensus documents on EB classification and characteristics, along with relevant textbook chapters on JEB were reviewed for characteristic clinical features of intermediate and severe JEB, as well as LOC syndrome (Fine et al., 2008, Fine et al., 2014, Barker et al., 2016, Has et al., 2020a). Discussions were held between expert clinicians in EB (Professor Adrian Heagerty, Professor Iain Chapple, Professor Jo-David Fine and Dr Ajoy Bardhan) regarding pertinent phenotyping criteria. Additional items were added to the BEBS to allow better characterisation of JEB patients and their features. Categories irrelevant to JEB were removed from the score. The phenotyping tool was piloted on three patients for ease of use, and the final version used in the study is shown in the Appendix (p19).

2.8.2. Data collection of deceased JEB patients

Medical records of deceased patients were reviewed and retrospective clinical data was collected which included genetic mutation, gender, age of death, cutaneous and extra-cutaneous clinical features (if documented) and cause of death (if known).

RESULTS

3.1 Genotypes of JEB cohort

3.1.1 Cohort characteristics

17 individuals with JEB met the inclusion criteria and consented to be included in the study. The study cohort consisted of 13 homozygotes and 4 compound heterozygotes with 21 mutations in total (Table 2). Of the 17 individuals, 7 had severe JEB (all deceased), 1 had LOC syndrome (deceased) and 9 had intermediate JEB (all living).

The majority of identified mutations were in the *LAMB3* gene (12 of 21, 57%), and were PTC-causing nonsense or out-of-frame indel mutations (14 of 21, 67%). 5 of 21 mutations (24%) were splice site mutations (Figure 4). Genes sequenced for each individual, along with parental genetic information, if available, are displayed in the Appendix (p20).
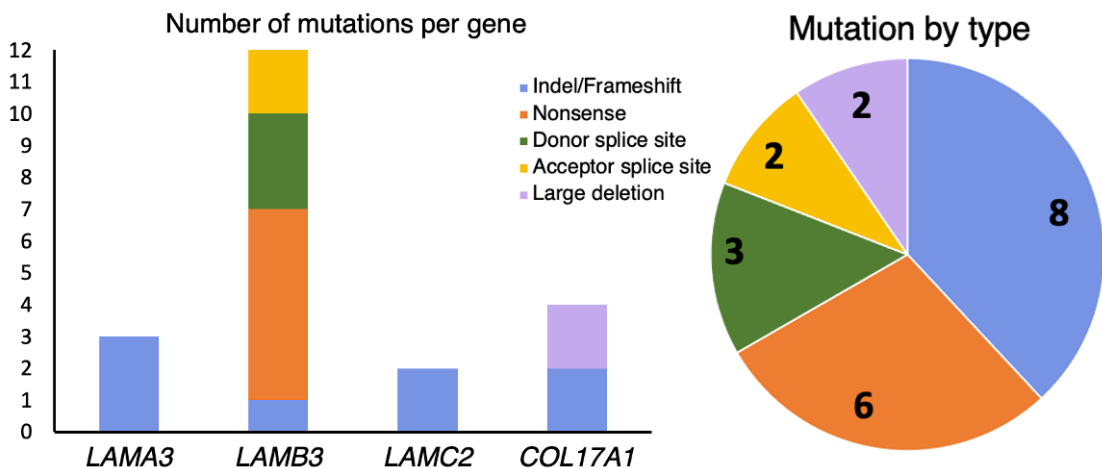
Figure 4: Summary of mutations for all JEB cases. Mutations present in two individuals (*LAMC2* c.132_135del, *COL17A1* c.2910insT and *COL17A1* del3010) are counted twice.

A few individuals had the same genotypes. Cases 12 and 13 had the same four nucleotide deletion in *LAMC2*, Cases 14 and 15 had the same single nucleotide insertion in *COL17A1*, and Cases 16 and 17 had the same large deletion in *COL17A1*. Therefore, of the 21 mutations identified in the cohort, 18 were unique in this cohort.

Case	Gender	Zygosity	Gene	Mutation	Amino acid change	Mutation consequence	JEB subtype
1	F	Homozygous	*LAMA3*	c.2038_2039dupAA	p.K680KfsX45	fs PTC	Severe
2	F	Homozygous	*LAMA3*	c.4338delG	p.L1446LfsX32	fs PTC	Severe
3	M	Homozygous	*LAMA3*	c.151insG	p.V51GfsX4	fs PTC	LOC
4	F	Heterozygous	*LAMB3*	c.727C>T, c.1903C>T	p.Q243X, p.R635X	Nonsense	Severe
5	F	Homozygous	*LAMB3*	c.1702C>T	p.Q568X	Nonsense	Intermediate
6	M	Homozygous	*LAMB3*	c.1186_1196del	p.T396CfsX12	fs PTC	Intermediate
7	F	Homozygous	*LAMB3*	c.2701+1G>A	N/A	Donor splice site mutation	Severe
8	F	Heterozygous	*LAMB3*	c.1705C>T, c.943+2T>C	p.R569X, N/A	Donor splice site and nonsense mutation	Intermediate
9	M	Homozygous	*LAMB3*	c.298+5G>C	N/A	Donor splice site mutation	Intermediate
10	F	Heterozygous	*LAMB3*	c.565-2A>G, c.2914C>T	N/A, p.R972X	Acceptor splice site and nonsense mutation	Severe
11	M	Heterozygous	*LAMB3*	c.3119G>A, c.629-12T>A	p.W1040X, N/A	Acceptor splice site and nonsense mutation	Intermediate
12	F	Homozygous	*LAMC2*	c.132_135delCAGA	p.H44HfsX63	fs PTC	Severe
13	F	Homozygous	*LAMC2*	c.132_135delCAGA	p.H44HfsX63	fs PTC	Severe
14	M	Homozygous	*COL17A1*	c.2910insT	p.P970SfsX8	fs PTC	Intermediate
15	M	Homozygous	*COL17A1*	c.2910insT	p.P970SfsX8	fs PTC	Intermediate
16	M	Homozygous	*COL17A1*	Large deletion of exons 16 and 17	Deletion of 81 amino acids	In frame large deletion	Intermediate
17	F	Homozygous	*COL17A1*	Large deletion of exons 16 and 17	Deletion of 81 amino acids	In frame large deletion	Intermediate

Table 2: Genotypes of JEB cohort. Two individuals (12 and 13) had the same four nucleotide deletion in *LAMC2*, two individuals (14 and 15) had the same single nucleotide insertion in *COL17A1*, and two individuals (16 and 17) had the same large deletion in *COL17A1*. All individuals with severe JEB died before 2 years of age. All individuals with intermediate JEB are alive at time of thesis submission (September 2022). Case 3 died at 6 years of age.

Mutation	Gene	Strand	cDNA mutation	Exon or intron	GRCh37			GRCh38			Variant ID
					Location start	Location end	Accession number	Location start	Location end	Accession number	
1	*LAMA3*	+	c.2038_2039dupAA	Exon 17	Chr18:21487749	Chr18:21487750	NM_000227	Chr18:23907785	Chr18:23907786	NM_000227.6	New
2	*LAMA3*	+	c.4338delG	Exon 32	Chr18:21523890	Chr18:21523890	NM_000227	Chr18:23943926	Chr18:23943926	NM_000227.6	New
3	*LAMA3*	+	c.151insG	Exon 1	Chr18:21453159	Chr18:21453159	NM_000227	Chr18:23873195	Chr18:23873195	NM_000227.6	rs80356678
4a	*LAMB3*	-	c.727C>T	Exon 8	Chr1:209806023	Chr1:209806023	NM_000228	Chr1:209632678	Chr1:209632678	NM_000228.3	rs80356681
4b	*LAMB3*	-	c.1903C>T	Exon 14	Chr1:209799066	Chr1:209799066	NM_000228	Chr1:209625721	Chr1:209625721	NM_000228.3	rs80356682
5	*LAMB3*	-	c.1702C>T	Exon 14	Chr1:209799267	Chr1:209799267	NM_000228	Chr1:209625922	Chr1:209625922	NM_000228.3	New
6	*LAMB3*	-	c.1186_1196delACC GGGCAGTG	Exon 11	Chr1:209801482	Chr1:209801472	NM_000228	Chr1:209628137	Chr1:209628127	NM_000228.3	New
7	*LAMB3*	-	c.2701+1G>A	Intron 18	Chr1:209795880	Chr1:209795880	NM_000228	Chr1:209622535	Chr1:209622535	NM_000228.3	rs1553276110
8a	*LAMB3*	-	c.1705C>T	Exon 14	Chr1:209799264	Chr1:209799264	NM_000228	Chr1:209625919	Chr1:209625919	NM_000228.3	rs201551805
8b	*LAMB3*	-	c.943+2T>C	Intron 9	Chr1:209803958	Chr1:209803958	NM_000228	Chr1:209630613	Chr1:209630613	NM_000228.3	New
9	*LAMB3*	-	c.298+5G>C	Intron 4	Chr1:209811874	Chr1:209811874	NM_000228	Chr1:209638529	Chr1:209638529	NM_000228.3	rs754529975
10a	*LAMB3*	-	c.2914C>T	Exon 20	Chr1:209791389	Chr1:209791389	NM_000228	Chr1:209618044	Chr1:209618044	NM_000228.3	rs747916314
10b	*LAMB3*	-	c.565-2A>G	Intron 6	Chr1:209806480	Chr1:209806480	NM_000228	Chr1:209633135	Chr1:209633135	NM_000228.3	rs370148688
11a	*LAMB3*	-	c.3119G>A	Exon 21	Chr1:209790864	Chr1:209790864	NM_000228	Chr1:209617519	Chr1:209617519	NM_000228.3	rs1057516759
11b	*LAMB3*	-	c.629-12T>A	Intron 7	Chr1:209806133	Chr1:209806133	NM_000228	Chr1:209632788	Chr1:209632788	NM_000228.3	rs754222671
12	*LAMC2*	+	c.132_135delCAGA	Exon 2	Chr1:183177068	Chr1:183177071	NM_005562	Chr1:183207930	Chr1:183207933	NM_005562.3	rs1057516806
13	*LAMC2*	+	c.132_135delCAGA	Exon 2	Chr1:183177068	Chr1:183177071	NM_005562	Chr1:183207930	Chr1:183207933	NM_005562.3	rs1057516806
14	*COL17A1*	-	c.2910insT	Exon 44	Chr10:105798866	Chr10:105798866	NM_000494	Chr10:104039108	Chr10:104039108	NM_000494.4	New
15	*COL17A1*	-	c.2910insT	Exon 44	Chr10:105798866	Chr10:105798866	NM_000494	Chr10:104039108	Chr10:104039108	NM_000494.4	New
16	*COL17A1*	-	3010nt deletion from intron 15 to intron 17	Exons 16, 17	Chr10:105819014	Chr10:105816005	NM_000494	Chr10:104059256	Chr10:104056247	NM_000494.4	New
17	*COL17A1*	-	3010nt deletion from intron 15 to intron 17	Exons 16, 17	Chr10:105819014	Chr10:105816005	NM_000494	Chr10:104059256	Chr10:104056247	NM_000494.4	New

Table 3: Annotation of mutations. Mutation number corresponds to case number. Heterozygous mutations are labelled as a and b.

3.1.2 Annotation and data visualisation of identified mutations

Identified mutations were mapped onto the reference genome (GRCh38) and transcripts (Table 3). As some of the in-silico tools used GRCh37 as a reference, genomic coordinates and accession numbers were also acquired for this version of the genome. All mutations were searched for in ClinVar and HGMD Professional 2021.1 variant databases (April 2022); seven novel, previously unreported mutations were identified (two in *LAMA3*, three in *LAMB3* and two in *COL17A1*). Schematics of mutations mapped to genes, transcripts and proteins are shown in Figures 5 to 8. Predicted InterPro protein domains of protein schematics and their significance scores are shown in the Appendix (p20).

3.1.3 *LAMB3* mutations

Twelve unique mutations in 8 individuals were present in *LAMB3* (Figure 5). These included 7 PTC-introducing mutations (6 nonsense and one frameshift mutation) and five splice site mutations, all of which were located within introns.

Three of the seven nonsense or frameshift mutations resulted in severe JEB. Two of these (pQ243X and p.R635X) occurred in the same compound heterozygote (Case 4), whilst the third (p.R972X) was compound heterozygous with a splice site mutation (c.565-2A>G, Case 10). One mutation (p.Q243X) mapped to the N-terminal domain of the laminin β3 chain (Figure 5), whilst the other two (p.R635X and p.R972X) did not map to any InterPro predicted domains (the region covering residues 575–1172).

Four of seven nonsense or frameshift mutations were associated with intermediate JEB; two were homozygous (p.T396CfsX12 and p.Q568X) and two were heterozygous (p.R569X and W1040X) with splice site mutations. Three of four mutations mapped to the EGF domain (p.T396CfsX12, p.Q568X and p.R569X), whilst the final mutation (W1040X) mapped towards the C-terminus of the β3 chain but did not map to a predicted domain. Splice site mutations are investigated further in Section 3.3.

3.1.4 *LAMA3* mutations

Three mutations in three individuals were identified in *LAMA3*, all of which were homozygous (Figure 6). Two mutations (c.2038_2039dupAA and c.4338delG) were present in both α3a and α3b isoforms and resulted in severe JEB. These mutations mapped to Domain II (p.K680KfsX45) and the G domain (p.L1446LfsX32) in both isoforms. The final mutation (c.151insG) was only present in the α3a isoform, and resulted in LOC syndrome.

3.1.5 *LAMC2* mutations

One unique mutation, present in two individuals, was found in *LAMC2* (c.132_135delCAGA). This was a four-nucleotide deletion which resulted in a frameshift and PTC in the N-terminal EGF domain (p.H44HfsX63). Both individuals were homozygous for this mutation and had severe JEB (Figure 7).
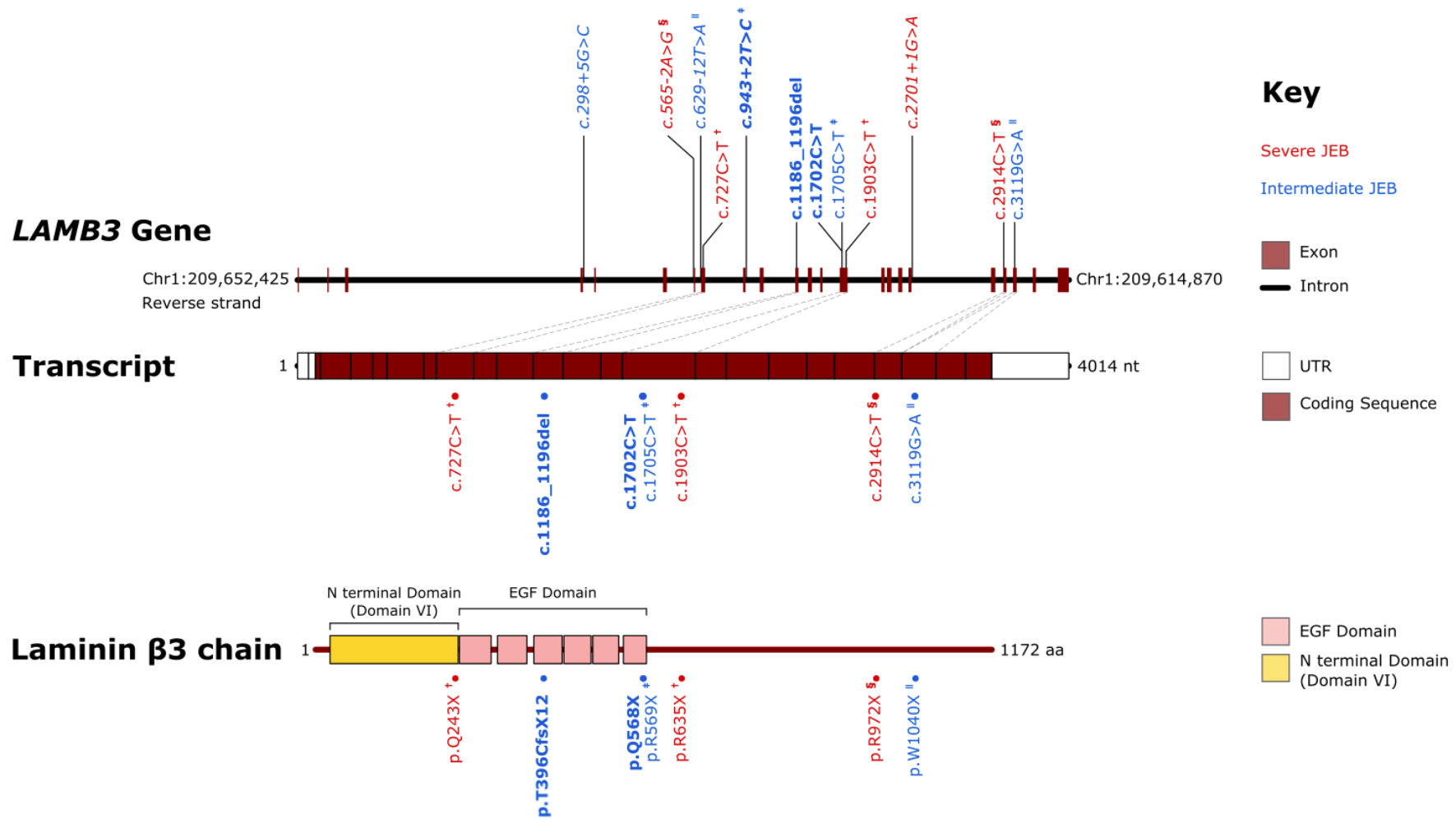
Figure 5: Mutations identified in *LAMB3* and corresponding phenotypes. Novel mutations are shown in bold. Splice site mutations are annotated on genes in italics. Mutations of heterozygotes are marked with the following markers: † Case 4, ‡ Case 8, § Case 10, ‖ Case 11.
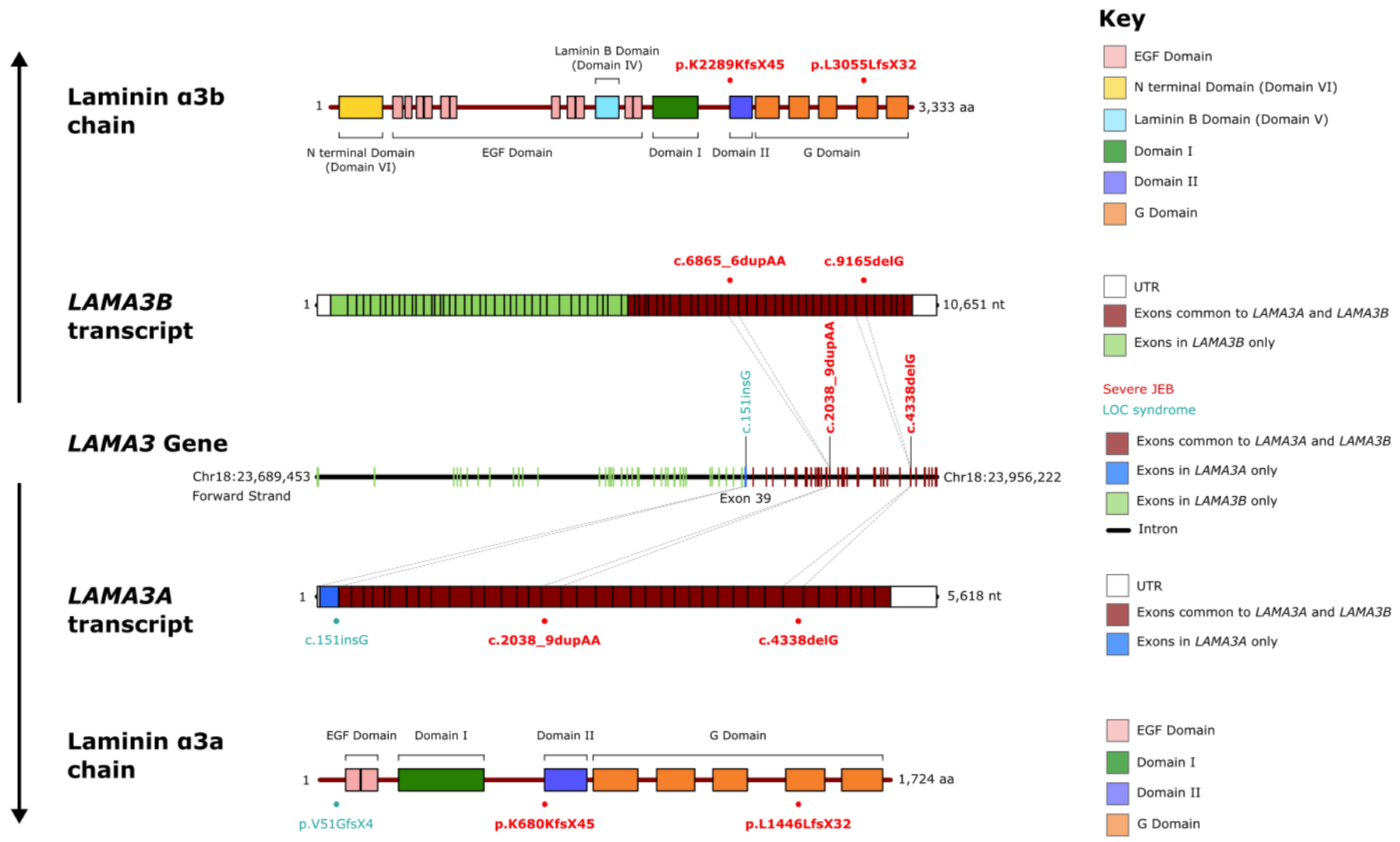
Figure 6: Mutations identified in *LAMA3* and corresponding phenotypes. All individuals were homozygotes. Novel mutations are shown in bold.
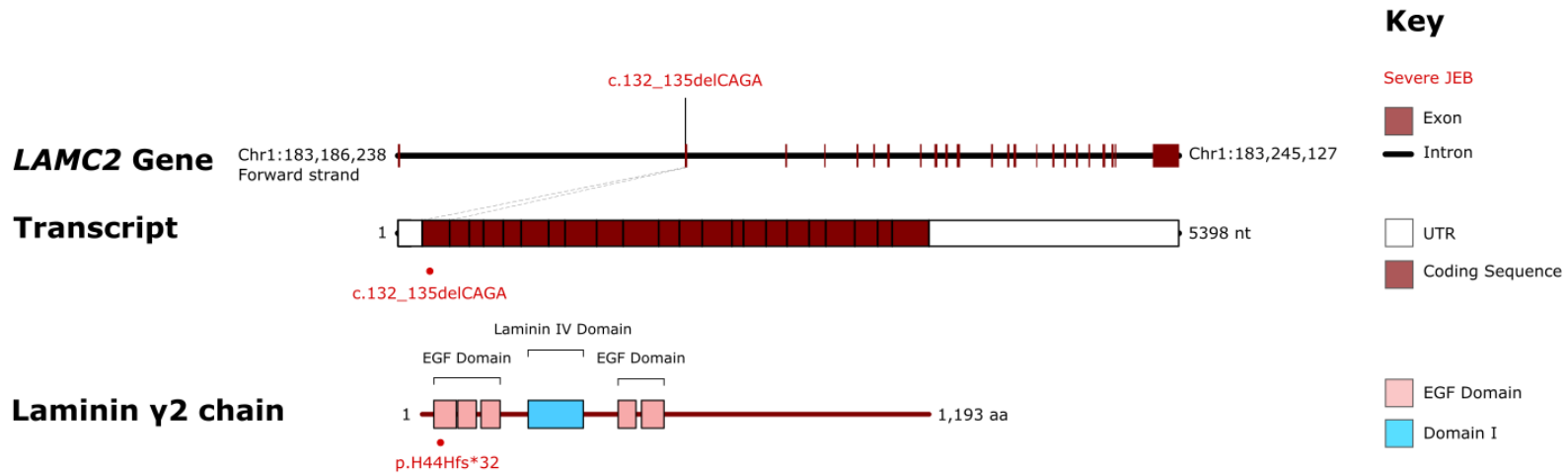
Figure 7: Mutations identified in *LAMC2* and corresponding phenotypes. All individuals were homozygotes.
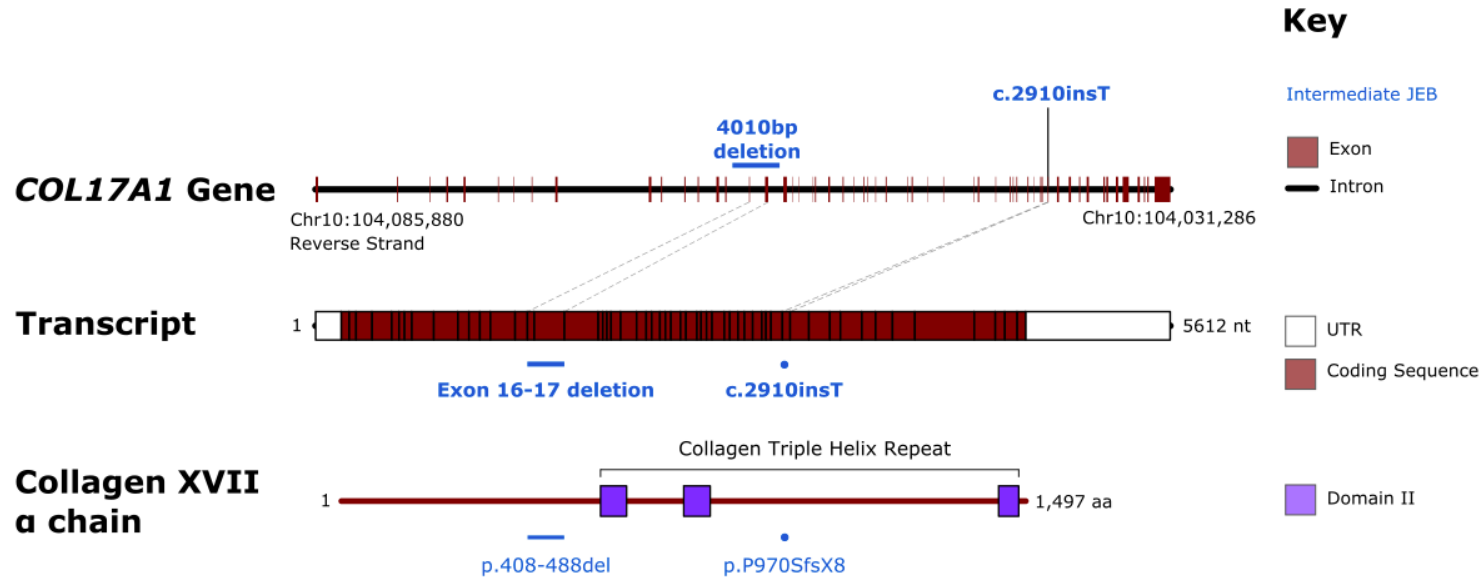
Figure 8: Mutations identified in *COL17A1* and corresponding phenotypes. All individuals are homozygotes. Novel mutations are shown in bold.

3.1.6 *COL17A1* mutations

Two unique mutations in *COL17A1* were identified in four individuals who all had intermediate JEB (Figure 8). Two individuals were homozygotes for the same large deletion of 3,010nt from intron 15 to intron 17. This resulted in deletion of exons 16 and 17 (243nt of exonic sequence), which corresponded to an in-frame deletion of 81 amino acids; these did not map to a predicted domain. The remaining two individuals had a one nucleotide insertion (c.2910insT) which resulted in a frameshift and PTC in the predicted collagen triple helix repeat domain (p.P970SfsX8).

3.2 Immunofluorescence mapping

IFM reports were available for 10 of 17 participants. Reports were reviewed retrospectively and available data are summarised in Table 4. Where present, blistering occurred at the dermo-epidermal junction in all cases, with collagen IV mapping to the base of the blisters, findings which are consistent with JEB. GB3 was the antibody that was most consistently used to detect laminin-332. Staining was completely absent in severe JEB cases, and was reduced or markedly reduced in intermediate JEB cases. Case 9 had two biopsies taken; the first was from an area which was not noticed to blister (non-blistered skin), and the second was from a blistered area. IFM with GB3 of non-blistered skin revealed patchy and substantially reduced staining of laminin-332 in comparison to control. However, the reduction in staining intensity of blistered skin was more marked and revealed almost completely absent staining compared to control skin.

Case	Split location	Laminin-332 (GB3)	Laminin-332 (P3H9-2)	Collagen 4 (COL94)	Collagen 17 (NC16A-3)
1	Antigen mapping is consistent with a split through the lamina lucida.	Complete lack of laminin-332.	–	–	–
2	Cryostat sections show extensive blistering at or close to the dermal-epidermal junction (DEJ) Antigen mapping is consistent with a split through the lamina lucida.	Complete lack of laminin-332 immunoreactivity.	–	–	–
4	Antigen mapping indicates a split within the lamina lucida.	Complete absence of immunoreactivity to laminin-332 using the GB3 antibody.	–	–	–
5	Staining for collagen 4 with COL94 maps to the base of splits.	Very marked reduction in laminin-332 staining which is faint, interrupted and uneven compared to controls.	–	Staining is of similar pattern and intensity to the control.	–
7	No clear split or separation seen in sections viewed	Complete absence of labelling in contrast to bright linear staining in controls	–	Staining is of similar pattern and intensity to control skin	Staining is slightly reduced in comparison to controls with some minor discontinuities
10	No clear split or separation seen in sections viewed	Complete absence of immunostaining at the DEJ compared to bright linear staining in controls	–	Staining is of similar pattern and intensity to control skin	Staining is slightly reduced and patchy in comparison to controls
8	Occasional splits or detachment at or close to the dermal-epidermal junction (DEJ) on cryostat sections. Antigen mapping suggests a split close to the DEJ perhaps within the lamina lucida. Staining for laminin-332 with GB3 and P3H9-2 maps to the base of split areas. Staining for collagen 4 with COL94 maps to the base of splits.	Reduced and intermittent staining at the DEJ compared to controls.	–	Staining is of similar pattern and intensity to the control.	Reduced and intermittent labelling at the DEJ, similar to controls.
9 (non-blistered skin)	Most of the DEJ is intact with just a central microsplit evident. Staining for laminin-332 with GB3 and P3H9-2 maps to the base of split areas. Staining for collagen 4 with COL94 maps to the base of splits.	Patchy and substantially reduced staining in comparison to control.	Staining is reduced and of a similar pattern in comparison to control.	Staining is of similar pattern and intensity to the control.	–
9 (blistered skin)	Cryostat sections show complete detachment of the epidermis from the underlying dermis. Staining for laminin-332 with GB3 and P3H9-2 maps to the base of split areas. Staining for collagen 4 with COL94 maps to the base of splits.	Almost completely absent staining in comparison to control. The reduction in staining intensity is more marked than the non-blistered skin.	Staining is greatly reduced in intensity and patchy in comparison to the control. The reduction in staining intensity is more marked than the non-blistered skin.	Staining is of similar pattern and intensity to the control although the DEJ appears thicker.	–
13	–	Complete absence of laminin-332 immunoreactivity at the DEJ.	–	–	–
15	Antigen mapping shows a plane of cleavage in the lamina lucida of the basement membrane zone.	–	–	–	Absence of staining for type XVII collagen.

Table 4: Immunofluorescence mapping. DEJ = dermo-epidermal junction. Antibodies used are shown in brackets. Severe JEB cases are underlined.

3.3 Prediction of nonsense-mediated decay

MasonMD was used to analyse PTC-introducing mutations and to predict whether NMD of transcripts would occur. Predictions are shown in the Appendix (p22). In summary, MasonMD predicted NMD of all nonsense mutations and frameshift indels (Mutations 1, 2, 4a, 4b, 5, 6, 8a, 10a, 11a, 12–15) except for mutation 3 (*LAMA3* c.151insG, where the mutation was mapped by the tool to an intron. This is because MasonMD was only able to perform analysis for the longer *LAMA3B* transcript (NM_198129.4), where the mutation was mapped to an intron. On manual assessment, this mutation is located within 200bp downstream of the transcription start sit, and therefore satisfies one of the rules for escaping NMD.

3.4 Splice site mutation consequence prediction

Five unique splice site mutations were identified in the cohort. These were all found in *LAMB3* and were intronic (3 donor and 2 acceptor splice site variants). These mutations were analysed using MMSplice (Cheng et al., 2019) and SpliceAI (Jaganathan et al., 2019).

3.4.1 MMSplice analysis

Analyses using MMSplice were performed to investigate the likelihood of the splice site mutations altering normal splicing. Genomic data formatted as .vcf files are displayed in the Appendix (p23). As *LAMB3* is translated from the anti-sense strand of chromosome 1, the bases in the .vcf files are complementary to the reported cDNA mutations, and correspond to the bases on the sense strand of *LAMB3*.

The unformatted output data is shown in p24 of the Appendix. For each variant, output values are automatically given for three transcripts (ENST00000356082.9_2, ENST00000391911.5_1 and ENST00000367030.7_1). For the splice site mutations from our cohort, all MMSplice output values are identical between the three transcripts analysed for each mutation. Key output scores for the principal *LAMB3* transcript (*LAMB3*: ENST00000356082.9_2, NM_000228.3) are displayed in Table 5.

The exons predicted to be affected by the splice site mutations were all single exons either upstream of donor splice site mutations, or downstream of acceptor splice site mutations. All delta_logit_psi values were negative, predicting higher rates of exon exclusion in mutated transcripts.

All variants, apart from c.943+2T>C (Case 8), had delta_logit_psi scores of less than -2, strongly predicting exon skipping to occur as a result of the splice site mutation. Exon skipping in all five cases would result in production of out-of-frame transcripts, as the number of nucleotides skipped in all cases was not a multiple of three (Table 5). Splicing efficiency was also predicted to be reduced in all cases. Furthermore, all mutations were predicted to be pathogenic, with pathogenicity scores of 0.99-1.00.

The delta_logit_psi score of Case 8 was -1.07, which is not predicted to have a strong effect on exon skipping. As this variant had a high pathogenicity score, it was hypothesised whether mechanisms other than exon skipping, such as cryptic splice site activation, could have resulted in this. This was investigated further with SpliceAI.

Variant and annotations					Output scores			Predicted consequence			Phenotype
Case number	Gene name	Mutation	Mutation location	Mutation type	Delta logit psi	Pathogenicity	Efficiency	Predicted skipped exon	Bases	Frame	
7	LAMB3	c.2701+1G>A	Intron 18	Donor splice site mutation	-4.16	1.00	-5.91	Exon 18	145	Out	Severe JEB
8	LAMB3	c.943+2T>C	Intron 9	Donor splice site mutation	-1.07	0.99	-1.62	Exon 9	121	Out	Intermediate JEB
9	LAMB3	c.298+5G>C	Intron 4	Donor splice site mutation	-4.05	1.00	-5.76	Exon 4	115	Out	Intermediate JEB
10	LAMB3	c.565-2A>G	Intron 6	Acceptor splice site mutation	-3.11	1.00	-1.98	Exon 7	64	Out	Severe JEB
11	LAMB3	c.629-12T>A	Intron 7	Acceptor splice site mutation	-3.82	1.00	-2.26	Exon 8	194	Out	Intermediate JEB

Table 5: MMSplice predictions. A delta logit percentage spliced in (PSI) score of less than -2 predicts that the variant has a strong effect.

3.4.2 Analysis of splice site variants using SpliceAI

The effect of splice site variants on mRNA splicing was further investigated with SpliceAI. This tool can analyse splice site variants and predict whether the original splice sites flanking an exon are likely to be disrupted (resulting in exon skipping), or whether a cryptic splice site is more likely to be used as a result of the splice site mutation. The input .vcf file was the same as that used for MMSplice analysis (Appendix p23).

SpliceAI allows the user to set the maximum distance (-D) between the mutation and a potential gain or loss site. This can be any value up to 4999 nucleotides upstream and downstream of the variant. To investigate whether prediction scores for gain or loss sites changed with -D (i.e., the maximum distance between the variant and a gain or loss site), recurrent analyses were performed with different parameters of -D, outlined in Table 6.

-D = 50nt		Delta scores				Positions			
Case	Mutation	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
7	c.2701+1G>A	0	0	0.37	0.99	20	22	20	1
8	c.943+2T>C	0	0	0.32	0.99	-30	20	-25	2
9	c.298+5G>C	0	0	0.01	0.93	-34	-1	-27	5
10	c.565-2A>G	0.1	0.99	0	0	24	-2	-27	3
11	c.629-12T>A	0.95	0.92	0	0	-2	-12	-13	-14
-D = 100nt		Delta scores				Positions			
Case	Mutation	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
7	c.2701+1G>A	0	0	0.37	0.99	20	22	20	1
8	c.943+2T>C	0	0.05	0.32	0.99	-30	98	-25	2
9	c.298+5G>C	0	0	0.75	0.93	-34	98	-59	5
10	c.565-2A>G	0.1	0.99	0	0.37	24	-2	67	-65
11	c.629-12T>A	0.95	0.92	0	0	-2	-12	-13	83

-D = 150nt		Delta scores				Positions			
Case	Mutation	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
7	c.2701+1G>A	0	0.26	0.37	0.99	20	145	20	1
8	c.943+2T>C	0	0.41	0.32	0.99	-30	122	-25	2
9	c.298+5G>C	0	0.17	0.75	0.93	-34	119	-59	5
10	c.565-2A>G	0.3	0.99	0	0.37	115	-2	67	-65
11	c.629-12T>A	0.95	0.92	0	0	-2	-12	-13	83

-D = 200nt		Delta scores				Positions			
Case	Mutation	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
7	c.2701+1G>A	0	0.26	0.37	0.99	20	145	20	1
8	c.943+2T>C	0	0.41	0.32	0.99	-30	122	-25	2
9	c.298+5G>C	0	0.17	0.75	0.93	-34	119	-59	5
10	c.565-2A>G	0.3	0.99	0	0.37	115	-2	67	-65
11	c.629-12T>A	0.95	0.92	0	0	-2	-12	-13	83

-D = 250nt		Delta scores				Positions			
Case	Mutation	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
7	c.2701+1G>A	0	0.26	0.37	0.99	20	145	20	1
8	c.943+2T>C	0	0.41	0.32	0.99	-30	122	-25	2
9	c.298+5G>C	0	0.17	0.75	0.93	-34	119	-59	5
10	c.565-2A>G	0.3	0.99	0	0.37	115	-2	67	-65
11	c.629-12T>A	0.95	0.92	0	0	-2	-12	-205	83

-D = 350nt		Delta scores				Positions			
Case	Mutation	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
7	c.2701+1G>A	0	0.26	0.37	0.99	20	145	20	1
8	c.943+2T>C	0	0.41	0.32	0.99	-30	122	-25	2
9	c.298+5G>C	0	0.17	0.75	0.93	-34	119	-59	5
10	c.565-2A>G	0.3	0.99	0	0.37	115	-2	67	-65
11	c.629-12T>A	0.95	0.92	0	0	-2	-12	-205	83

-D = 4999nt (max)		Delta scores				Positions			
Case	Mutation	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
7	c.2701+1G>A	0	0.26	0.37	0.99	627	145	20	1
8	c.943+2T>C	0	0.41	0.32	0.99	2050	122	-25	2
9	c.298+5G>C	0	0.17	0.75	0.93	-3891	119	-59	5
10	c.565-2A>G	0.3	0.99	0	0.37	115	-2	67	-65
11	c.629-12T>A	0.95	0.92	0	0	-2	-12	-205	1805

Table 6: SpliceAI predictions with varying settings for -D at 50, 100, 150, 200, 250, 350 and 4999 nucleotides. Changes in delta scores from one -D setting to the next are highlighted in red. Corresponding changes in positions of gain or loss sites are also highlighted.

No further delta score changes occurred beyond -D of 150nt. This is likely because the size of -D (150nt) corresponded with the length of exon 18 (145nt), which was the longest exon to be affected by a splice site mutation (Case 7, c.2701+1G>A). When -D is set below 146nt (e.g., at 100nt), the usual acceptor splice site of exon 18 (146 nucleotides away from the c.2701+1G>A donor splice site mutation) is not included in the area analysed. Of the area analysed (100 nucleotides upstream and downstream of the mutation), the acceptor loss score is 0, i.e. the mutation does not decrease the strength of a potential acceptor splice site within 100 nucleotides of the mutation. When -D increases from 100nt to 150nt, the usual acceptor splice site of exon 18 is included in the analysis, and the acceptor loss score increases from 0 to 0.26, suggesting that the c.2701+1G>A mutation disrupts both exon 18 donor and acceptor splice sites.

The scores did not change with -D from 150nt to 4999nt. Using c.2701+1G>A as an example, this suggests that the exon 18 splice sites were affected the most, rather than the splice sites of adjacent exons. This is in agreement with the MMSplice predictions, where the exons predicted to be affected by the splice site mutations were all single exons, rather than multiple exons together. If other splice sites were affected to a greater extent, then the delta scores and positions would continue to change with an increase in -D. However, SpliceAI only gives a single score representing the largest effect for acceptor gain, acceptor loss, donor gain and donor loss, and so other transcripts which are less likely to be produced may not be displayed as a prediction. Regarding Cases 8 and 9, the scores did not change beyond -D of 150nt either, as the affected exons were 121nt (exon 9) and 115nt (exon 4) respectively.

3.4.2.1 Donor splice site mutations

Case 7

This individual was homozygous for a *LAMB3* c.2701+1G>A donor splice site mutation within intron 18. Each prediction from SpliceAI is outlined below:

1. Donor Loss

This was predicted to severely disrupt a donor splice site (donor loss 0.99) one nucleotide downstream of the mutation on genomic DNA (gDNA). On *LAMB3* cDNA, this corresponds to c.2701 (as the *LAMB3* transcript is anti-sense), which is the original junction between exon 18 and intron 18 (Figure 9).
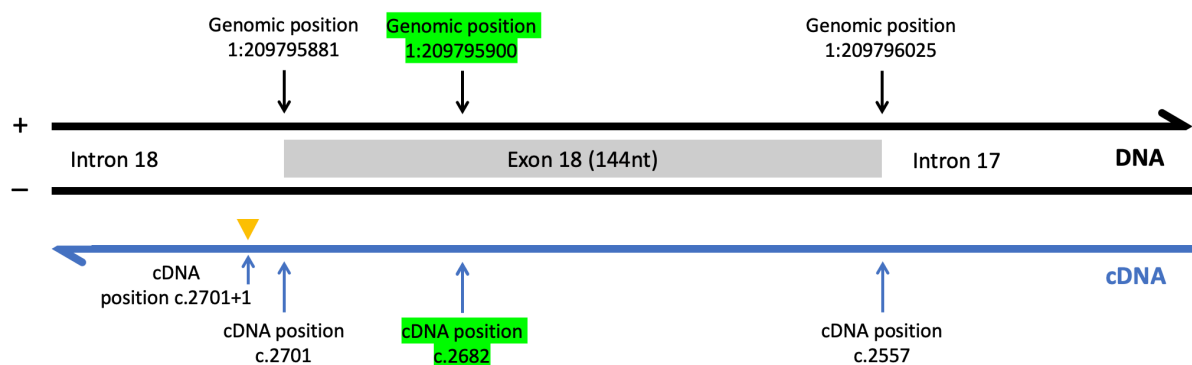
Figure 9: Relationship between genomic coordinates of *LAMB3* (gDNA) and cDNA positions. Genomic positions increase to 5' to 3' as cDNA positions decrease. The c.2701+1G>A mutation is marked by a yellow triangle at 1:209795880. The exon borders are outlined with arrows. The predicted novel cryptic splice site position is highlighted in green. Nucleotide positions are not to scale.

2. Acceptor Loss

Some disruption to an acceptor splice site 145 nucleotides downstream of the mutation was also predicted (acceptor loss of 0.26). This corresponded to the original junction between intron 17 and exon 18 (c.2557).

3. Donor Gain

There was a prediction for donor gain of 0.37 20nt upstream of the mutation (c.2682), and 19nt upstream original donor splice site (Figures 9 and 10, highlighted in green).

4. Acceptor Gain

There was no positive score predicted for a new acceptor splice site (acceptor gain).

```
TGTATGTCAACAGATTAGGGCAGCCGAGGAATCTGCCTCACAGATTCAATCCAGTGCCCAGCGCTTGGAGACCCA
GGTGAGCGCCAGCCGCTCCCAGATGGAGGAAGATGTCAGACGCACACGGCTCCTAATCCAGCAGTCCGGGACTT
CCTAACAGGTGAGCCCA
```

Figure 10: DNA sequence and location of predicted splicing consequences (c.2701+1G>A)
Key: Exon 18 is shown in blue text and highlighted in grey. It is flanked by introns 17 and 18 in black text. The c.2701+1G>A mutation is shown in yellow text and underlined.
The disrupted original splice sites (acceptor loss site 0.26, *LAMB3* c.2557, position 1:209796025; donor loss site 0.99, *LAMB3* c.2701, position 1:209795881) are shown in red text and underlined. The predicted donor gain site (+ 0.37, *LAMB3* c.2682 , position 1:209795900) is highlighted in green and underlined.

These predictions are summarised in a schematic in Figure 11. Disruption of the original exon 18 donor and acceptor splice sites would result in skipping of exon 18, whilst disruption of the exon 18 donor splice site in conjunction with activation of a donor cryptic splice site would result in the use of the novel cryptic splice site to generate a shortened exon 18.
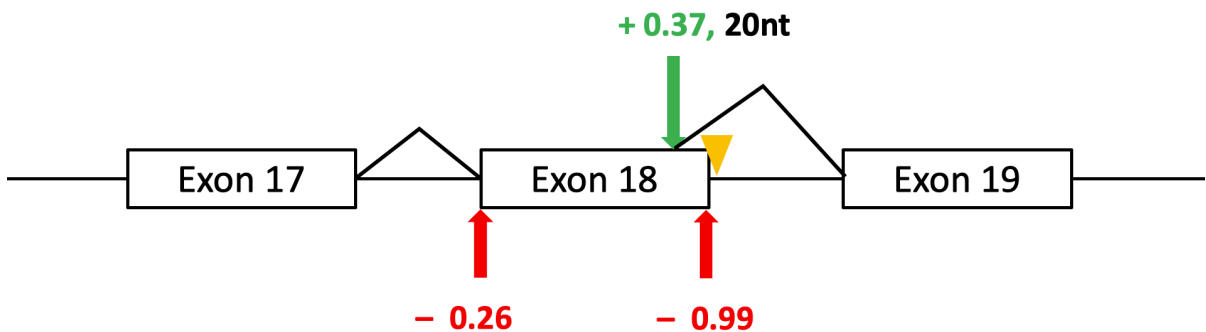
Figure 11: Schematic diagram of predicted splicing outcomes from c.2701+1G>A. The mutation is marked by a yellow triangle. Donor and acceptor loss scores are shown in red with (−). The donor gain score is shown in green with (+). The predicted novel donor site is 20 nucleotides from the original donor splice site.

Overall, the predicted consequences include exon 18 skipping (145nt) or exclusion of 20nt from the 3' end of exon 18 from cryptic splice site activation within this exon. Both predicted outcomes result in production of out-of-frame transcripts.

Case 8

This individual was a compound heterozygote for a nonsense mutation in exon 14 (c.1705C>T), and a donor splice site mutation in intron 9 (c.943+2T>C). The splice site

mutation was predicted to disrupt the original donor splice site (donor loss 0.99) and upstream acceptor splice site (acceptor loss 0.41), along with activation of a cryptic splice site (donor gain 0.32), 27 nucleotides downstream of the original donor splice site, within intron 9 (Figure 12).
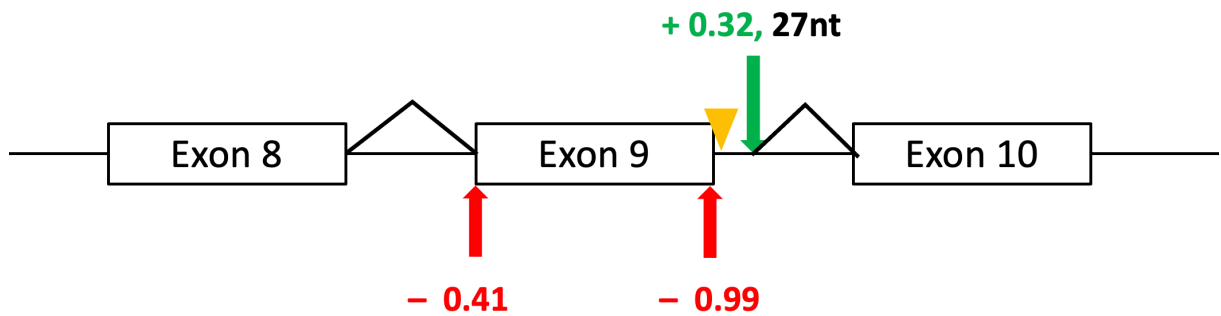
Figure 12: Schematic diagram of predicted splicing outcomes from c.943+2T>C. The novel cryptic donor splice site is 27 nucleotides away from the original donor splice site.

Overall predicted consequences included out-of-frame exon 9 skipping (121nt) or inclusion of 27 additional nucleotides from intron 9, which would result in an in-frame transcript.

Case 9:

This homozygous donor splice site mutation in intron 4 (c.298+5G>C) was predicted to disrupt the original donor splice site (0.93), with slight acceptor splice site disruption (0.17). Cryptic splice site activation within intron 4 was predicted with a donor gain score of 0.75 (Figure 13).
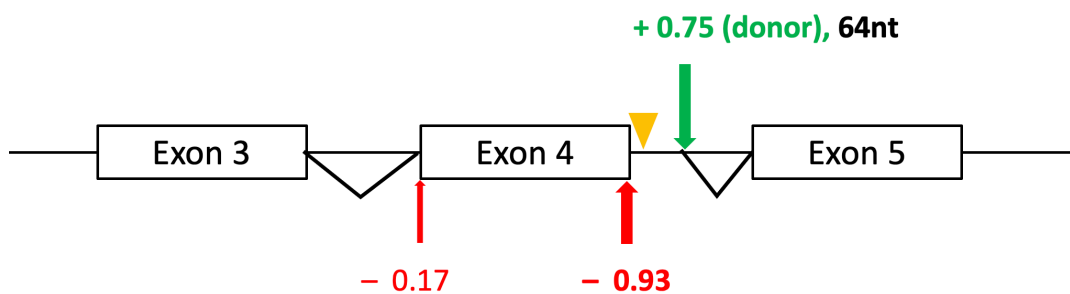
Figure 13: Schematic diagram of predicted splicing outcomes from c.298+5G>C

This was predicted to be 64nt downstream of the original donor splice site, which would lead to inclusion of 64nt from intron 4 and would generate an out-of-frame transcript. Likewise, skipping of exon 4 would also lead to production of an out-of-frame transcript. Both resulting transcripts would be expected to correspond with a severe JEB phenotype. Nevertheless, this individual had a phenotype consistent with intermediate JEB, and so further investigation was warranted (Section 3.4).

3.4.2.2 Acceptor splice site mutations

Case 10

This individual was a compound heterozygote for a nonsense mutation in exon 20 (c.2914C>T) and an acceptor splice site mutation within intron 6 (c.565-2A>G). The splice site mutation was predicted to severely disrupt the original acceptor splice site (0.99 acceptor loss), and to also disrupt the corresponding donor splice site (0.37 donor loss). There was an acceptor gain score of 0.3, 117nt upstream of the original acceptor splice site within intron 6 (Figure 14).
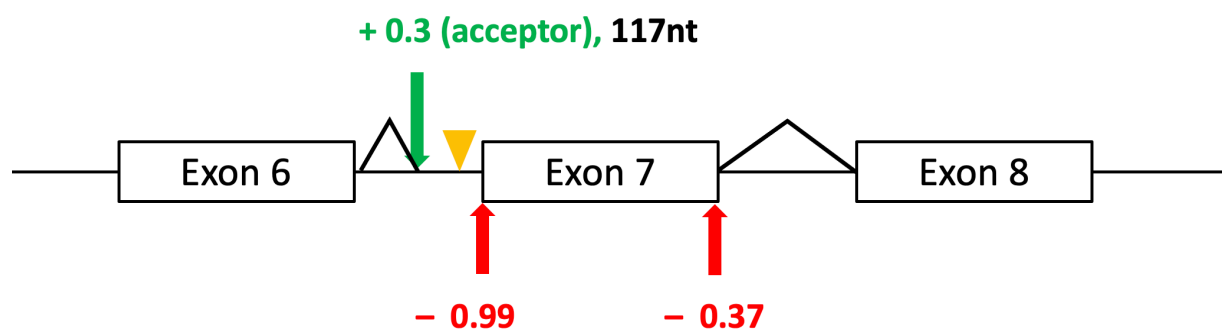
Figure 14: Schematic diagram of predicted splicing outcomes from c.565-2A>G

These two predicted outcomes would result in either out-of-frame exon 7 skipping (64nt) or inclusion of 117nt from intron 6 secondary to cryptic splice site activation. The

latter would generate an in-frame transcript with an additional 117nt (translating to 39 amino acids) inserted between exons 6 and 7.

Case 11

This individual was a compound heterozygote for a nonsense mutation in exon 21 (c.3119G>A), and an acceptor splice site mutation in intron 7 (c.629-12T>A). The splice site mutation was predicted to disrupt the original acceptor splice site (0.92 acceptor loss), with no effect on the corresponding donor splice site (donor loss score of 0). There was, nevertheless, a strong prediction for cryptic splice site activation within intron 7 (0.95 acceptor gain), 10 nucleotides upstream of the original acceptor splice site (Figure 15).
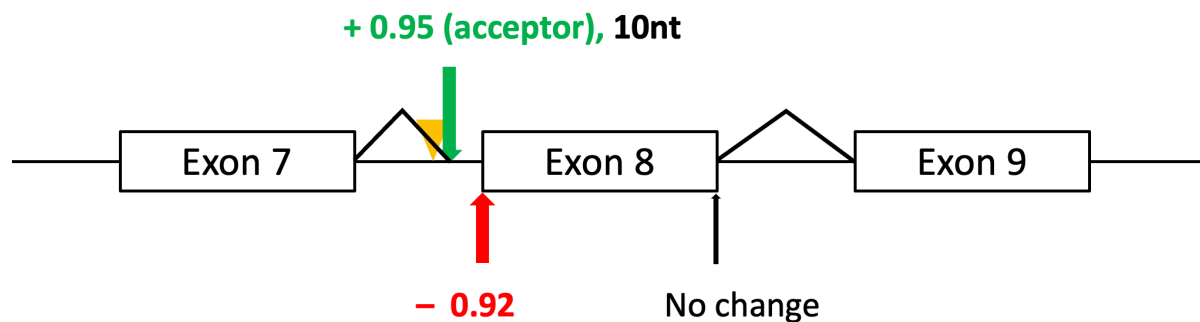
Figure 15: Schematic diagram of predicted splicing outcomes from c.629-12T>A

This would lead to inclusion of 10nt from intron 7 in between exons 7 and 8, which would result in a frameshift.

3.5.3 Summary of predictions from both tools

Reviewing results from MMsplice and SpliceAI together, both exon skipping and cryptic splice site activation are possible outcomes for all cases (Table 7). MMSplice had strong predictions for exon skipping for all cases except for Case 8. SpliceAI predicted exon skipping (to varying degrees) for all cases except for Case 11. SpliceAI also had a low score for exon skipping in Case 9 (0.17). This contrasted with MMSplice, where there were strong predictions for exon skipping in these two cases. SpliceAI predicted cryptic splice site activation (with varying scores) for all cases; the strongest predictions were for Case 9 (0.75) and Case 11 (0.95).

Case	Mutation	Delta logit psi	MMsplice prediction	SpliceAI – exon skipping	SpliceAI – cryptic splice site	Outcome comments
7	c.2701+1G>A	-4.1587118	Out-of-frame exon 18 skipping	Out-of-frame exon 18 skipping (AL=0.26)	Cryptic ss activation resulting in exclusion of 20nt from exon 18 (**DG=0.37**)	Both outcomes out-of-frame
8	c.943+2T>C	-1.0718014	Out-of-frame exon 9 skipping, however not considered to be a strong effect	Out-of-frame exon 9 skipping (**AL= 0.41**)	Cryptic ss activation resulting in inclusion of 27nt from <u>intron</u> 9 (DG=0.32)	1. Cryptic ss activation in-frame 2. Exon skipping out-of-frame
9	c.298+5G>C	-4.0515497	Out-of-frame exon 4 skipping	Out-of-frame exon 4 skipping (AL=0.17)	Cryptic ss activation resulting in inclusion of 64nt from intron 4 (**DG=0.75**)	Both outcomes out-of-frame
10	c.565-2A>G	-3.1108091	Out-of-frame exon 7 skipping	Out-of-frame exon 7 skipping (**DL=0.37**)	Cryptic ss activation resulting in inclusion of 117nt from intron 6 (in-frame, AG=0.3)	1. Cryptic ss activation in-frame 2. Exon skipping out-of-frame
11	c.629-12T>A	-3.8153212	Out-of-frame exon 8 skipping	Predicted to be extremely unlikely (DL=0)	Cryptic ss activation resulting in inclusion of 10 nucleotides from intron 7 (**AG=0.95**)	Both outcomes out-of-frame

Table 7: Summary of predicted splice site outcomes. Abbreviations: ss=splice site, AL=acceptor loss, AG=acceptor gain, DL=donor loss, DG=donor gain. The SpliceAI outcome with the higher score is highlighted in bold.

3.5.4 Predicted splice site mutation outcomes

To evaluate whether the predicted splicing outcomes would result in PTCs, reference *LAMB3* transcript sequences were manually altered to the transcripts predicted by the tools by inserting or deleting DNA sequences that were predicted to be included or excluded by altered splicing. Sequences were translated into a primary amino acid sequence with ExPASY and aligned with the reference amino acid sequence using NCBI protein blast. All out-of-frame transcripts contained a PTC (Table 8), as did the in-frame insertion in Case 10. Here, cryptic splice site activation was predicted to result in the insertion of 117 additional nucleotides between exons 6 and 7, corresponding to 39 additional amino acids from position p.190 to p.228. However, a TGA stop codon is present at c.565-60, and results in a PTC at position p.208 (Figure 16). In contrast, the predicted in-frame transcript generated by cryptic splice site activation for Case 8 contained 9 amino acid insertion (RSFGSPLPW) without a PTC.

Transcript	Mutation	Consequence
7 exon skipping	c.2701+1G>A	p.I853TfsX32
7 cryptic splice site	c.2701+1G>A	p.V895PfsX4
8 exon skipping	c.943+2T>C	p.V275GfsX81
8 cryptic splice site	c.943+2T>C	p.R315Sins9
9 exon skipping	c.298+5G>C	p.W62MfsX2
9 cryptic splice site	c.298+5G>C	p.D100GfsX56
10 exon skipping	c.565-2A>G	p.V189GfsX47
10 cryptic splice site	c.565-2A>G	p.V189ins19X
11 exon skipping	c.629-12T>A	p.E210GfsX17
11 cryptic splice site	c.629-12T>A	p.E210GfsX51

Table 8: Translated splice site mutation predicted outcomes

MRPFFLLCFALPGLLHAQQACSRGACYPPVGDLLVGRTRFLRASSTCGLTKPETYCTQYGEWQMKCCKCDSRQPHNYYSHRVE
NVASSSGPMRWWQSQNDVNPVSLQLDLDRRFQLQEVMMEFQGPMPAGMLIERSSDFGKTWRVYQYLAADCTSTFPRVRQ
GRPQSWQDVRCQSLPQRPNARLNGGKVPGNKDFKCEAYNGKKGFF-

Figure 16: Translated sequence of c.565-2A>G cryptic splice site transcript. Inserted amino acids secondary to cryptic splice site activation are shown in blue.

3.6 RNA analysis

Out-of-frame cryptic splice site activation and out-of-frame exon skipping were predicted for Case 9 which would be expected to result in a severe JEB phenotype. However, he had an intermediate JEB phenotype and therefore cDNA sequences that were available for this case from RT-PCR were reviewed. For reference, the wild type *LAMB3* exon 4, intron 4 and exon 5 sequences are shown below (Figure 17).

Figure 17: *LAMB3* wild type sequence. Exons are outlined in blue letters and highlighted in grey. Intron 4 is outlined in red; this region would usually be spliced out. c.298+5 is underlined; the wild type base (cytosine) is shown.

This individual has a homozygous c.298+5G>C donor splice site mutation and a novel cryptic splice site was predicted within intron 4 at position c.298+64 (donor gain 0.75), which would lead to inclusion of 64nt between exons 4 and 5 after splicing (Transcript A, Figure 18).

Figure 18: Transcript A is predicted by SpliceAI following cryptic splice site activation. Original exons are outlined in blue letters and highlighted in grey. The included 64nt of intron 4 are outlined in blue letters and unhighlighted. The remainder of intron 4 which is spliced out is outlined in red letters. The c.298+5G>C mutation is highlighted in yellow. The novel predicted cryptic splice site (c.295+64) is highlighted in green which leads to the inclusion of 64 additional nucleotides in Transcript A.

RNA studies of white blood cells from blood confirmed the presence of a transcript with an additional 64nt of intron 4 included between exons 4 and 5 (Transcript A, Figure 19).
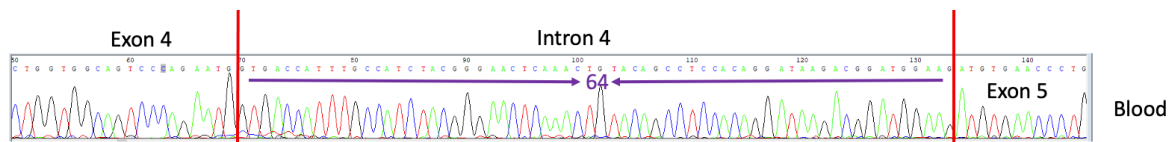
Figure 19: Sanger sequencing electrophoretogram of *LAMB3* cDNA from blood

Interestingly, RT-PCR of biopsied skin detected Transcript A, and also an additional 'minor' transcript was detected (Transcript B, Figure 20). In Transcript B, 60nt were included between exons 4 and 5.
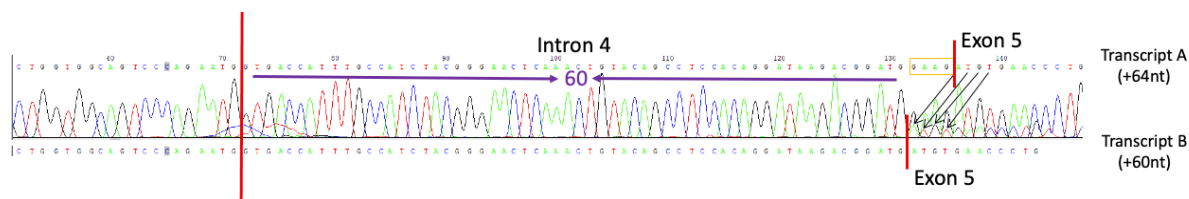
Figure 20: Sanger sequencing electrophoretogram of *LAMB3* cDNA from skin. The GAAG nucleotides from position c.298+61 to c.298+64 are outlined in the orange box

Transcript B would be expected to result in generation of an in-frame transcript, with an additional 20 amino acids (GEHLPSTGTQTVQPPQDKTD) inserted at position p.100, where exons 4 ends and exon 5 begins. The difference between the two transcripts was four nucleotides from position c.298+61 to c.298+64 (GAAG, Figures 20 and 21).

Transcript B

Exon 4 / Intron 4 / Exon 5

Figure 21: Transcript B detected via RT-PCR. Original exons are outlined in blue letters and highlighted in grey. The included 60nt of intron 4 are outlined in blue letters and unhighlighted. The remainder of intron 4 which is spliced out is outlined in red letters. The c.298+5G>C mutation is highlighted in yellow. The novel predicted cryptic splice site (c.295+60) is highlighted in green which produces Transcript B. The four nucleotides (GAAG) which differ between Transcripts A and B are emphasized in bold and underlined.

The SpliceAI predictions for Case 9 are shown again below:

Delta scores				Positions			
Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
0	0.17	**0.75**	0.93	-34	119	**-59**	5

Table 9: SpliceAI predictions for c.298+5G>C. The predicted donor gain score and its corresponding position are underlined

The tool predicted a novel donor splice site 59 nucleotides upstream (as *LAMB3* transcripts are anti-sense) of the mutation at c.298+5G>C, i.e., at cDNA position 298+64. This prediction was completed with -D set at ≥150 nucleotides, and the -59 position was predicted to have the greatest donor gain (0.75). Unfortunately, SpliceAI only displays the position with the greatest donor gain (in this case position -59 with a donor gain of 0.75), and other donor gain sites at different positions (which may have noteworthy donor gain probabilities, although less than 0.75) are not given by the tool.

Production of Transcript B would require a novel donor spice site at position c.298+60. To investigate whether the c.298+5G>C mutation would increase the probability of the

c.298+60 position being used as a splice site (by any degree), targeted SpliceAI analysis was performed with targeted settings of -D for 58nt, which would evaluate the sequence 58 nucleotides either side of the c.298+5G>C mutation. Within intron 4, this would include all nucleotides from 298+1–298+63 in the area analysed, but not position 298+64 or any nucleotides downstream (Figure 22).
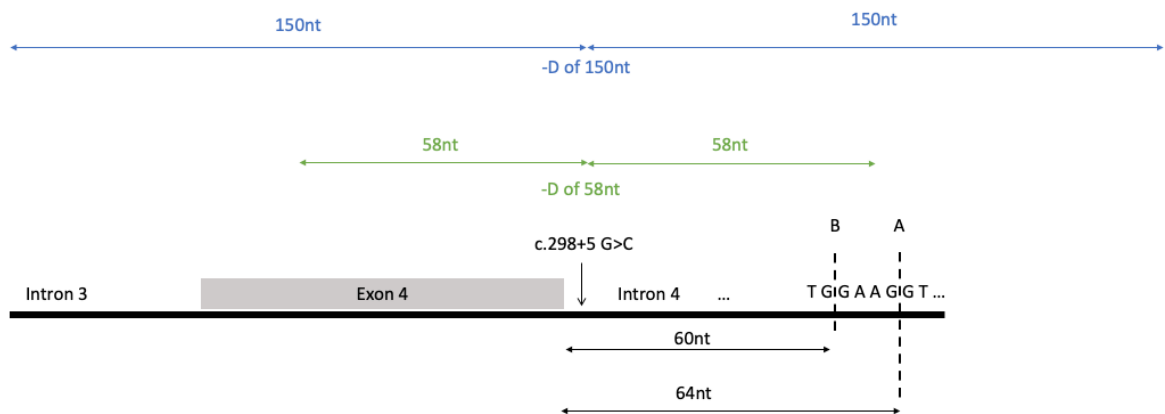
Figure 22: Genomic regions analysed by -D of 58nt (green) and -D of 150nt (blue). Splice site A at c.295+64 corresponds with Transcript A. Splice site B at c.295+60 corresponds with Transcript B. Distances are not to scale.

The results for analysis with -D set at 58nt are displayed below in Table 10, with the donor gain score and corresponding position underlined. A minimal donor gain of 0.01 is predicted at position -27 (27 nucleotides upstream of the mutation) i.e., at c.295+32 (using anti-sense *LAMB3* cDNA notation). There is no donor gain presented by the tool at position -55 i.e., at c.295+60, and even if there were, the gain score would be <0.01 (given that SpliceAI displays the donor gain position that corresponds to the highest donor gain delta score).

-D = 58	Delta scores				Positions			
Position	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
209811874	0	0	<u>0.01</u>	0.93	-34	-1	<u>-27</u>	5

Table 10: Targeted analysis of surrounding region with -D of 58nt

Reviewing the primary DNA sequence, the splice site required to produce Transcript B would be located at TG|GA (splice site B, Figure 22). The GA dinucleotide appears to be poorly suited to being a donor splice site, as it does not follow Chambon's rule which states that the vast majority of introns begin with a GT dinucleotide (Breathnach et al., 1978).

Therefore, it was hypothesised that a second mutation (A>T) could be present in a small subpopulation of cells at position c.295+62 resulting in TG|GT at the splice site junction (Figure 23). This would generate a consensus GT at the splice site and could have increased the likelihood of a novel splice site being generated at this position to produce Transcript B.

Transcript B with second mutation

Figure 23: Transcript B with a second intronic mutation (c.295+62A>T). Original exons are outlined in blue letters and highlighted in grey. The included 60nt of intron 4 are outlined in blue letters and unhighlighted. The remainder of intron 4 which is spliced out is outlined in red letters. The c.298+5G>C mutation is highlighted in yellow. The novel predicted cryptic splice site (c.295+60) is highlighted in green which produces Transcript B. The hypothesised second A>T mutation at position c.295+62 creating a GT dinucleotide is highlighted in cyan.

SpliceAI analysis was re-run with the c.298+62A>T mutation. Unfortunately, SpliceAI can only analyse one mutation at a time, and so this mutation could not be analysed together with the identified c.298+5G>C mutation. Nevertheless, analysis of the c.298+62A>T mutation in isolation predicted activation of a cryptic splice site two

nucleotides downstream of c.298+62 i.e., at c.298+60 (Table 11). The donor gain score of 0.94 suggested that this would be a strong splice site. A summary of the SpliceAI predictions is shown in Figure 24.

Mutation	Delta scores				Positions			
	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
c.298+62A>T	0	0	0.94	0.15	71	-31	2	-2

Table 11: SpliceAI predictions for c.298+62A>T

To confirm whether the hypothesised second mutation was present (c.298+62A>T), intron 4 sequence was reviewed to investigate whether an A>T substitution at this position had occurred. No second mutation was found at position c.298+62 on Sanger sequencing, and the wild type adenine base was present, suggesting that alternative mechanisms may have produced Transcript B, (or that only a very small cell population had the second mutation, which was too small to be detected on Sanger sequencing).
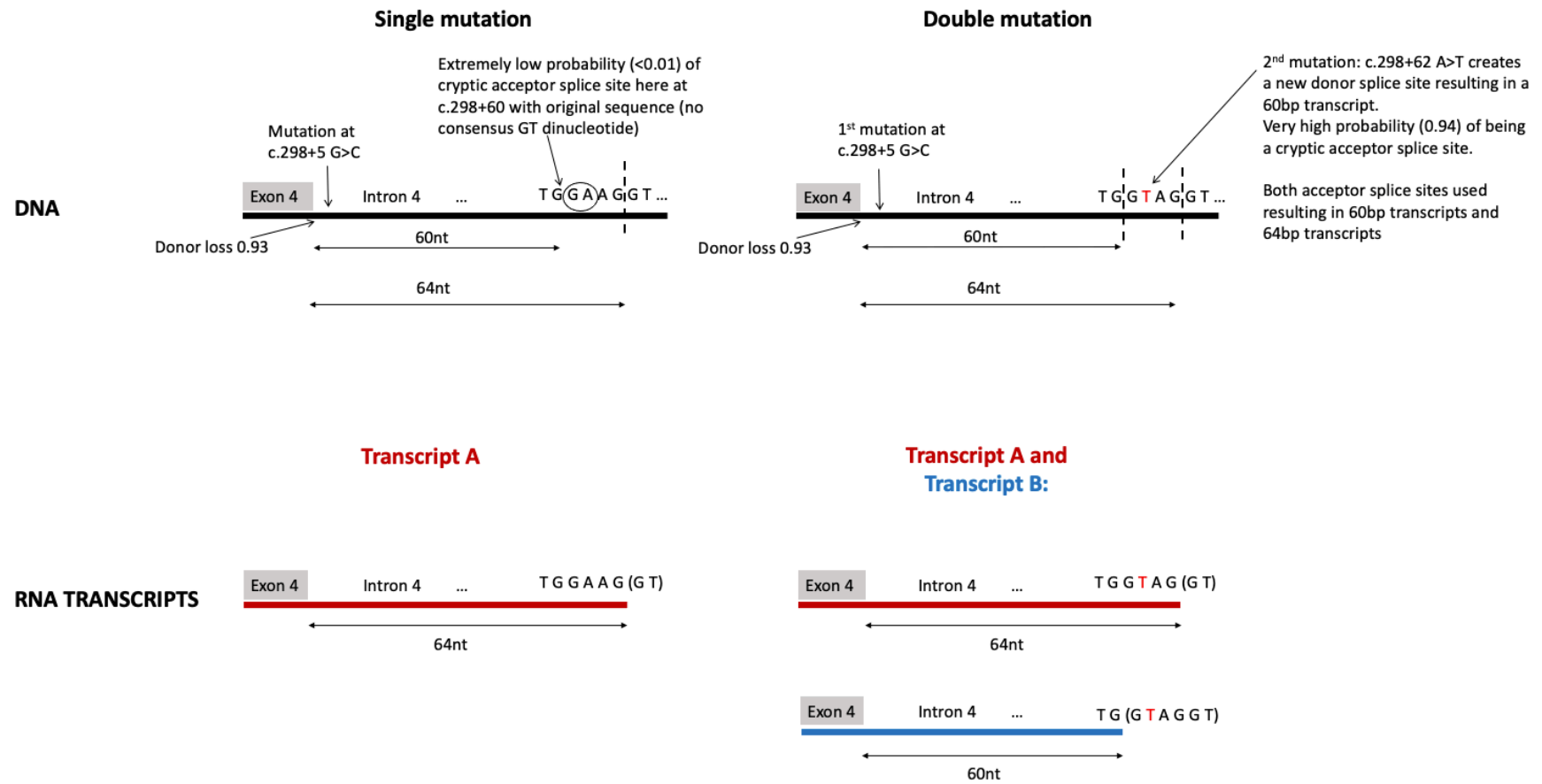
Figure 24: SpliceAI predictions for a novel cryptic splice site at c.298+60. 'Single mutation' refers to c.298+5G>C only, 'double mutation' refers to c.298+5G>C and c.298+62 A>T.

3.7 Effects of exonic mutations on splicing

In addition to altering the primary sequence of a polypeptide, exonic mutations can also affect splicing. Natural in-frame skipping of the mutation-bearing exon has been shown to be a mechanism that can eliminate PTCs, ameliorating JEB severity (McGrath et al., 1999). In order to investigate whether in-frame skipping of the mutation through splicing had occurred, analysis with SpliceAI (-D = 4999nt) was performed on Cases 5 and 6 (who had milder than expected phenotypes) to evaluate whether these mutations had any effect on nearby splicing (Table 12).

Case	Mutation	Delta scores				Positions			
		Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss	Acceptor Gain	Acceptor Loss	Donor Gain	Donor Loss
5	c.1702C>T	0.1	0.04	0.01	0.01	-57	104	-1181	-183
6	c.1186_1196del	0.01	0.01	0	0.01	89	227	-320	493

Table 12: SpliceAI predictions for Cases 5 and 6

There was negligible alteration of splicing for Case 6, with gain and loss scores of 0.01 only. Regarding Case 5, there was modest prediction (0.1) for a novel acceptor cryptic splice site at position c.1759 (161nt from the original intron 13/exon 14 junction), with minimal disruption to the original acceptor splice site (acceptor loss 0.04, Figure 25). Even if a transcript was produced from this, it would be out-of-frame, and translation with Expasy confirmed a PTC two codons downstream from the new acceptor site (p.A533DfsX3).
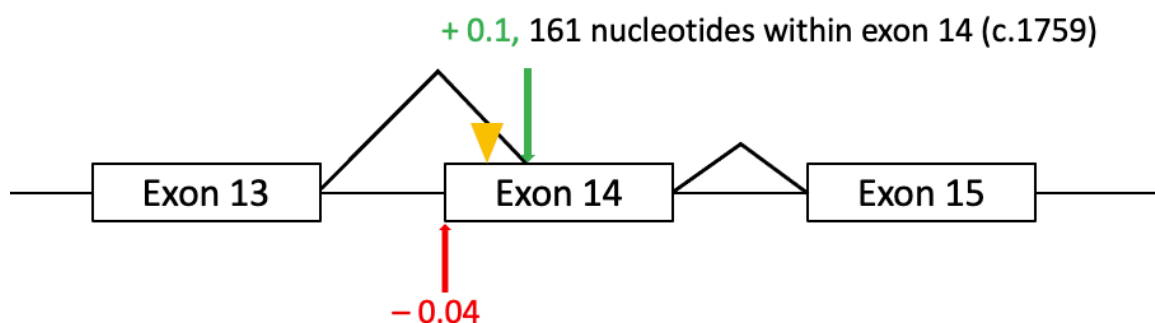
Figure 25: Schematic diagram of predicted splicing outcomes from c.1702C>T

3.8 Deep phenotyping of cohort:

Participants were systematically examined for cutaneous and extra-cutaneous features associated with JEB using a dedicated deep phenotyping tool for JEB (Appendix p19). This was only possible for intermediate JEB participants, as all individuals with severe JEB in the study had passed away. The deep phenotyping findings are shown in Table 13 and 14.

Case	BSA	Chronic wounds	Eyes	Larynx	Nails	Scarring alopecia	Teeth	Granulation tissue	Life threatening illness
	Max score: 10	Max score: 5							Max score: 10
5	4	2	2	0	5	4	4	4	0
6	4	2	3	1	5	2	4	4	0
8	0	0	1	0	2.5	0	4	0	0
9	6	2	0	0	2.875	1	4	0	0
11	3	2	2	0	5	3	3	0	0
14	2	0	1	0	3.5	0	0	0	0
15	4	1	1	0	3.5	0	3	0	0
16	5	2	1	2	4.375	2	0	0	1
17	2	1	1	1	5	5	4	0	0

Key:

% of max score within category
100
80
60
40
20
0
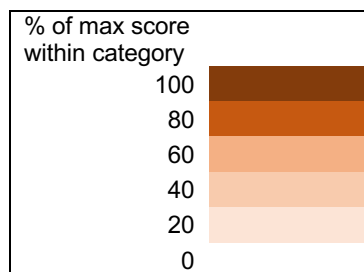
Table 13: Scores of intermediate JEB cases from deep phenotyping. BSA = body surface area affected, this includes blisters, erosions, scabs, healing skin, erythema and atrophic scarring. This is given as a score out of 10 (see Appendix p19). Life threatening illness was scored out of 10. All other categories are scored out of a maximum of 5.

Case	Underweight (BMI <18.5)	Hypoalbuminaemia (35-50 g/L ref)	Anaemia (Hb <130 (males) or <120 (females) g/dL)	Mouth ulcers	Other comments
5	Yes	Yes	Yes	Yes	Atrophic scarring
6	Yes	No	Yes	No	Atrophic scarring
8	No	NR	NR	Yes	10% non-scarring alopecia
9	Yes	Yes	Yes	No	Atrophic scarring
11	No	Yes	No	Yes	Atrophic scarring
14	Yes	No	Yes	Yes	
15	No	No	Yes	Yes	
16	No	No	Yes	Yes	Atrophic scarring and keratoderma
17	Yes	Yes	No	Yes	Atrophic scarring and hypopigmentation

Table 14: Further characteristics of intermediate JEB cases from deep phenotyping. NR = not reported.

Deep phenotyping of nine intermediate JEB individuals revealed substantial variation. Skin area affected varied from 0% (Case 8) to 51–60% (Case 9). Case 8 did not have any skin involvement at time of examination, whereas all other cases did. All individuals' nails were affected with scores ranging from 2.5 to 5 (loss of all nails). Ocular involvement, chronic wounds, tooth damage or loss, oral ulcers (secondary to oral blistering with loss of blister roof), scarring alopecia, atrophic scarring, anaemia and being underweight were also common (present in >50% of cases). Hypoalbuminaemia, laryngeal involvement and presence of granulation tissue were less common (present in <50% of cases). Case 16 was the only individual who experienced a life-threatening illness (sepsis).

JEB MUTATION LANDSCAPE AND GENOTYPE-PHENOTYPE CORRELATION

4.1 Reported JEB mutations

A large proportion of genotypes from our cohort were either nonsense or frameshift mutations. In an attempt to compare these with mutations reported in the literature, nonsense and frameshift indel variants from the HGMD mutation database associated with JEB were retrieved (HGMD Professional 2021.1). These were subsequently mapped onto the reference proteins according to their phenotype category as reported in the database (Figure 26). The number and type of mutations for each gene that were associated with each JEB phenotype is summarised in the Appendix (p2–4).

From Figure 26 it can be observed that mutations within the N-terminal domain of the laminin β chain are associated with both 'EB, Herlitz' and 'EB, junctional'. This is also seen with EGF domain mutations, and the C-terminal region of the protein without predicted domains (residues 575–1172). Nevertheless, there appear to be regions where mutations associated with 'EB, Herlitz' and 'EB, junctional' do not overlap, particularly within the C-terminal region (residues 575–1172).

Mutations reported in *LAMA3, LAMC2* and *COL17A1* were also mapped. Fewer reported mutations and cases were available for *LAMA3* and *LAMC2*, which was also reflected in our cohort. The predicted EGF domain and Laminin IV domain of the laminin γ2 chain contained mutations associated with 'EB, Herlitz' and also 'EB, junctional', as did the predicted EGF and G domains of the laminin α3a chain. Mutations in Domain I and Domain II of the α3a chain were associated with 'EB,

junctional' only. Mutations in *COL17A1* associated with JEB were found both within and outside the Collagen Triple Helix Repeat Domain. Three mutations were notably associated with Herlitz EB.
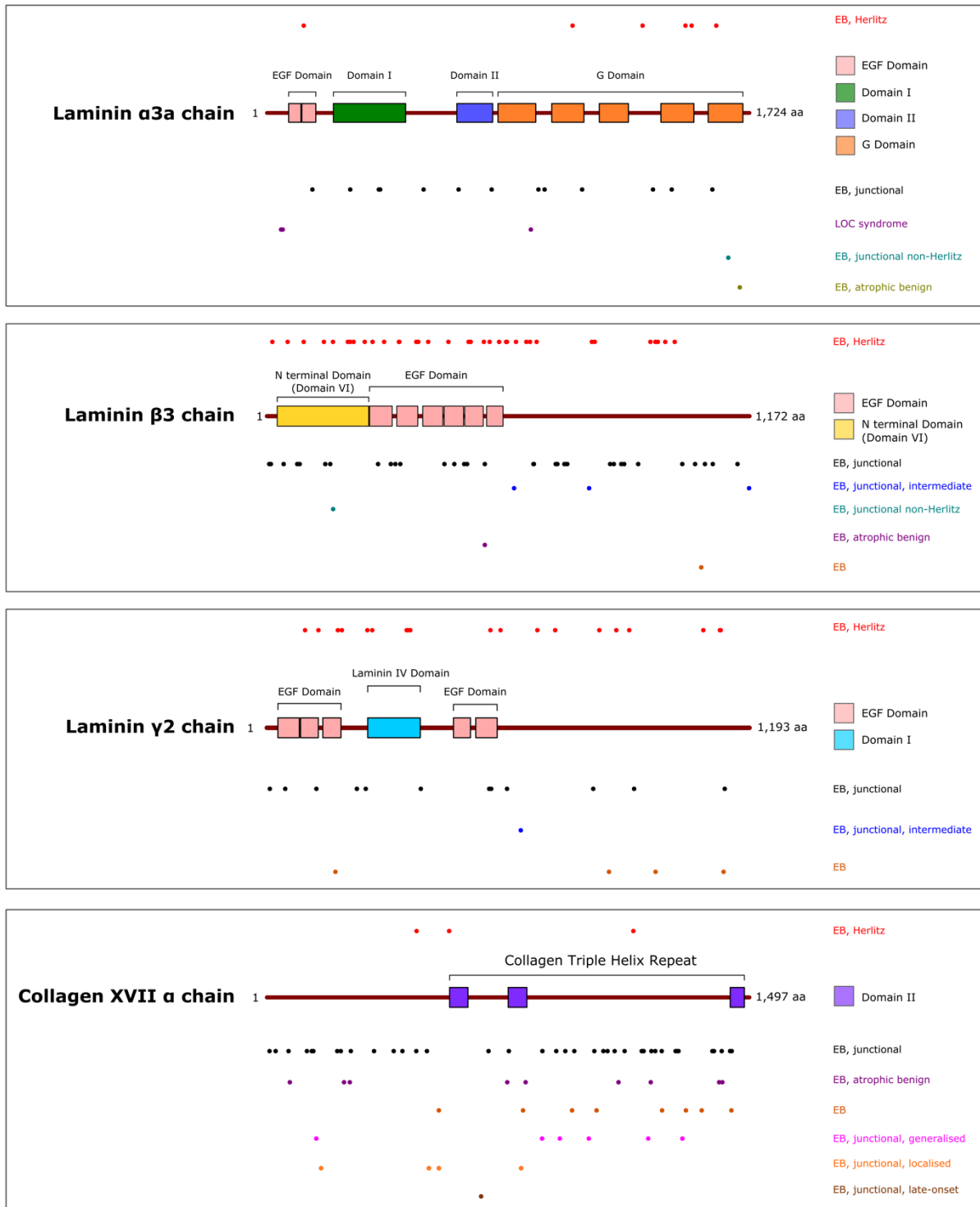
Figure 26: Nonsense and frameshift indel variants reported in HGMD. JEB categories are labelled as they are shown in HGMD.

4.1.1 Limitations of HGMD

However, these figures have limited practical utility and should be interpreted with caution for the following reasons. Unfortunately, a large proportion of mutations reported in the HGMD database were associated with an 'EB, junctional' phenotype, with no further granularity regarding which JEB subtype this was. Some phenotype nomenclature, such as 'EB, Herlitz' and 'EB, junctional non-Herlitz' are outdated and have since been reclassified.

Furthermore, the zygosity of individuals with each mutation was not readily available from the HGMD interface. JEB is a recessive monogenic disorder, where production of only a small amount of partially functional protein from one allele can dramatically ameliorate the severity of an individual's phenotype (Has et al., 2020). Heterozygous individuals with a null mutation may be compensated through a less severe mutation on another allele. For example, p.R569X homozygotes have a severe JEB phenotype (Khan et al., 2021). However, Case 8 has an intermediate JEB phenotype with genotype p.R569X/c.943+2T>C. The c.943+2T>C splice site mutation is predicted by SpliceAI to produce some in-frame transcripts which may ameliorate disease severity for this individual.

The true effect and severity of a mutation can be appreciated either in homozygotes, or in heterozygotes where the exact consequence of the other allele is known. Therefore, knowledge of zygosity and of corresponding mutations (if heterozygous) is essential in the evaluation of the true severity associated with an individual mutation.

The associated mutation on the corresponding allele must also be provided in mutation databases, which was not readily available from the HGMD interface.

Therefore, due to these reasons, mutations from our cohort were not compared with the mutations reported in HGMD using this method. Instead, publications associated with reported mutations in HGMD and ClinVar were reviewed for phenotype information, and also mutations from both alleles.

4.2 Genotype-phenotype correlations

4.2.1 Nonsense and indel mutations

4.2.1.1 Severe JEB

LAMB3

The genotype of Case 4 (p.Q243X/p.R635X) has been reported previously and is also associated with a severe JEB phenotype (Pulkkinen et al., 1997). The p.R635X mutation is the most common mutation associated with JEB in the laminin-332 genes, accounting for 45% of all *LAMB3* severe JEB cases (Nakano et al., 2000).

p.Q243X is also a recurrent mutation, accounting for 5.3% of all *LAMB3* severe JEB cases (Nakano et al., 2000). Homozygous genotypes of p.Q243X and p.R635X result in severe JEB (Nakano et al., 2000). Therefore, it is unsurprising this combination of mutations results in severe JEB. This is reflected in the IFM findings of Case 4, where there was almost absent staining for laminin-332 in comparison to control.

LAMA3

The two cases with *LAMA3* mutations were homozygotes for frameshift indels (p.K680KfsX45 and p.L1446LfsX32), and demonstrated a complete lack of laminin-332 immunoreactivity on IFM, consistent with an expected severe JEB phenotype. Possible explanations for the lack of staining include either NMD of transcripts, or truncated α3a chains that were unable to form laminin-332 heterotrimers.

p.K680KfsX45 has not been reported previously. A homozygous nonsense mutation (p.R782X) affecting the same domain (Domain II, residues 679–807) resulted in severe JEB too (Castori et al., 2008). p.L1446LfsX32 is also a novel mutation; another homozygous exon 17 frameshift mutation (c.4335dupA, p.L1446IfsX11) similarly resulted in a severe JEB phenotype (Ayoub et al., 2005).

LAMC2

The one *LAMC2* mutation identified in our cohort (H44HfsX63) resulted in severe JEB in two cases. This mutation mapped to exon 1, resulting in complete absence of laminin-332 immunoreactivity at the DEJ on IFM, consistent with severe JEB. This mutation is reported in ClinVar but is not linked to any publications in the literature. A similar case with a homozygous p.R95X mutation also resulted in complete absence of staining for laminin-332 with GB3 and severe JEB (Aberdam et al., 1994).

4.2.1.2 Intermediate JEB

LAMB3

The novel nonsense c.1702C>T (p.R568X) mutation in *LAMB3* (Case 5) is notable as it did not result in the expected severe JEB phenotype. Remarkably, individuals homozygous for the nearby mutation *LAMB3* c.1705C>T in the adjacent codon (p.R569X) have been reported to have a severe JEB phenotype with early lethality (Khan et al., 2021). Other cases where the R569X mutation has been reported as a compound heterozygous mutation include Q243X/R569X, which also results in severe JEB (Kivirikko et al., 1996).

Natural in-frame skipping of the mutation bearing exon has been shown to be a mechanism that can eliminate PTCs, ameliorating JEB severity (McGrath et al., 1999). Reviewing the SpliceAI predictions, there was only a very low chance of this occurring for c.1702C>T, and even if it did, would result in an out-of-frame transcript and PTC. Therefore, this mechanism was unlikely to account for the relatively mild phenotype.

c.1186_1196del (p.T396CfsX12) is also a novel mutation (Case 6). A nearby and similar deletion of 11nt (c.1188_1198del, p.G397VfsX11) resulted in severe JEB when present on both alleles (Varki et al., 2006). Examining the cDNA and amino acid sequences together, it can be seen that a similar region is deleted in both cases with the PTCs being in the same location for both cases (Figure 27). The net effect is that only one codon is altered, and this results in two different JEB subtypes. However, whilst there is a GTG codon rather than a GAC codon at c.1185, non-coding, regulatory functions of DNA should also be considered for further investigation.

c.1186_1196del:
```
1168 GCTCCCTGTGACCCAGT GACCGGGCAGTGTG TGTGCAAGGAGCATGTGCAGGGAGAGCG CTGTGA
 390 -A--P--C--D--P--V -             -C --V--Q--G--A--C--A--G--R--A- -L--X-
```

c.1188_1198del:
```
1168 GCTCCCTGTGACCCAGT GACCGGGCAGTGTG TGTGCAAGGAGCATGTGCAGGGAGAGCG CTGTGA
 390 -A--P--C--D--P--V --T            --V--Q--G--A--C--A--G--R--A- -L--X-
```

Figure 27: Aligned nucleotides and amino acids of Case 6 and c.1188_1198del. Homologous nucleotides and amino acids are in black text. Nucleotides and amino acids differing between the two cases are shown in orange. Deleted nucleotides are struck through. Stop codons are shown in red.

c.1186_1196del was not predicted by SpliceAI to alter splicing and prompt skipping of the PTC. Readthrough of *TGA* PTCs has also been reported as a mechanism where translation is preserved. Pacho et al report a *LAMA3* compound heterozygote (R943X/R1159X) where a c.2827C>T mutation introduced a TGA premature stop codon at R943X. Readthrough of the R943X allele (c.2827C>T) occurred, and through site directed mutagenesis, the authors demonstrated that PTC readthrough depended on the genetic sequence flanking the TGA stop codon, and determined the consensus sequence (A/T)(A/G)(T/C) **TGA** CTA for PTC readthrough (Pacho et al., 2011).

The PTC and flanking sequence of Case 6 is shown below, which does not match the consensus sequence suggested by Pacho et al. In fact, a cytosine nucleotide at position -3, guanine at position -1, cytosine at position +2 and thymine at +3 were all demonstrated to markedly reduce readthrough from occurring.

```
Position      -3 -2 -1  0  +1 +2 +3
Nucleotide     C  T  G  TGA  C  C  T
```

Figure 28: PTC and surrounding sequence of Case 6. The TGA stop codon is shown in red.

Therefore, according to the consensus sequence suggested by Pacho et al, PTC readthrough was unlikely to have occurred in Case 6, and also for Case 5, where a

TAG PTC was present. Whilst other consensus sequences may exist which could facilitate PTC readthrough, given the widespread skin involvement and severity in these two individuals, this is unlikely; the case reported by Pacho et al produced full length *LAMA3* transcript and laminin-332 on IFM, and was phenotypically very mild by 5 years of age. Therefore, it is likely that these two variants result in intermediate JEB through other mechanisms. Analysis of the RNA transcripts of these individuals will undoubtedly shed light on this.

COL17A1

In line with classical genotype-phenotype correlation paradigms, Cases 14, 15, 16 and 17 had intermediate JEB phenotypes. All *COL17A1* mutations in our cohort were novel. Although the intron 15-17 deletion (exons 16 and 17, p.408-488) was in-frame, there was widespread skin and extracutaneous involvement. The COL17 transmembrane domain (p.467-489) is almost completely deleted, which may explain the relatively severe intermediate JEB phenotype. Herisse et al report a severe intermediate JEB phenotype with generalised blistering, nail and dental dystrophy and severe mucosal involvement secondary to a deletion from intron 16 to intron 17 (834 nucleotides, c.1268-267 to c.1465+369) which resulted in negative staining for COL17 on IFM (Herisse et al., 2021).

Regarding the novel p.P970SfsX8 mutation, a similar p.G954AfsX112 (c.2860delG) mutation resulted in negative COL17 IFM, alopecia, dystrophic nails, oral blistering and generalised skin blistering (Kiritsi et al., 2011).

LOC syndrome

As expected, the homozygous c.151insG genotype in Case 3 resulted in LOC syndrome, and this individual passed away at six years of age following an out of hospital cardiac arrest. Although a PTC is introduced 7 nucleotides downstream of the mutation in exon 1, the transcript has been demonstrated to escape NMD as an alternative ATG start codon is utilised six exons downstream (McLean et al., 2003). The net overall is an N-terminal truncation of the laminin α3a polypeptide which lacks 226 residues, and this gives rise to the characteristic LOC phenotype.

4.2.2 *LAMB3* Splice site mutations

4.2.2.1 Severe JEB

Case 7 (c.2701+1G>A homozygote)

Both tools predicted out-of-frame skipping of exon 18 (I853TfsX32), and SpliceAI also predicted cryptic splice site activation which would also generate an out-of-frame transcript (V895PfsX4). Both predicted out-of-frame transcripts correspond to lack of staining for laminin-332 with GB3 on IFM and the severe JEB phenotype of the individual. This mutation was also found in a newborn with skin fragility and nail dystrophy (Laroussi et al., 2017), although there was no comment on which type of JEB this individual developed.

Case 10 (p.R972X, c.565-2A>G heterozygote)

Homozygotes for p.R972X, and also c.565-2A>G have been reported to have a severe JEB phenotype (Nakano et al., 2000, Pulkkinen et al., 1997). Therefore, in combination, these two mutations would be expected to result in severe JEB, which

correlates with the phenotype of Case 10. Nonetheless, additional insight is provided by the splice site analysis, which would be especially valuable if the c.565-2A>G mutation had not been reported previously. Out-of-frame exon 7 skipping was predicted by SpliceAI and MMsplice resulting in V189GfsX47. SpliceAI also predicted in-frame cryptic splice site activation 117nt within intron 6 which resulted in the inclusion of 19 amino acids and a stop codon, and thus both predicted outcomes of c.565-2A>G resulted in PTCs, explaining the lack of staining for laminin-332 with GB3 on IFM and the severe JEB phenotype.

4.2.2.2 Intermediate JEB phenotypes

Case 11 (p.W1040X, c.629-12T>A heterozygote)

This genotype has been reported by Chen et al in a 1-day-old male patient (Patient 1) who was alive at the time of study publication. The described phenotype for this individual was severe JEB, although no IFM data was available (Chen et al., 2020).

The c.629-12T>A mutation was also found in Patient 12 of Chen's cohort in combination with p.R635X in a 12-year-old intermediate JEB individual. As described previously, transcripts produced from p.R635X undergo NMD, and homozygotes with this mutation have a severe JEB phenotype (Nakano et al., 2000, Pulkkinen et al., 1994). Therefore, it is likely that the c.629-12T>A mutation produces some functional polypeptide, and that W1040X is a loss of function mutation. Patients 2 (p.Q730X/p.W1040X) and 4 (p.L7X/p.W1040X) from Chen's cohort provide further evidence for this, as the W1040X mutation in combination with another nonsense mutation has resulted in death and severe JEB in both cases.

The c.629-12T>A mutation was also reported by Hou et al. Using RT-PCR, it was demonstrated that this mutation did indeed activate a cryptic splice site within intron 7, which led to the inclusion of 10 additional nucleotides (Hou et al., 2021). This matched the SpliceAI prediction of cryptic splice site activation within intron 7 (and disproved the MMSplice prediction of exon skipping). Additionally, using microfluidic electrophoresis, it was also demonstrated that this was a leaky splice site that is not utilised in all splicing events, resulting in alternate splicing occurring where wild type transcript was also produced in a 1:3–1:7 ratio to mutant transcript. Therefore, this could provide a potential explanation for Case 8's intermediate JEB phenotype (and Patient 11 in Chen et al's cohort with the genotype p.R635X/c.629-12T>A with a JEB-intermediate phenotype and alive at 12 years of age). Regarding the SpliceAI predictions, the acceptor loss delta score was -0.92, which is slightly less than the acceptor loss prediction of -0.99 in Case 10, and the donor loss prediction of -0.99 in Cases 7 and 8. This lower value may reflect some leakiness of the splice site.

Regarding Patient 1 from Chen et al's cohort, as severe and intermediate JEB can be difficult to distinguish in early life based on clinical presentation alone, and as this individual is still young with no IFM data available, it is possible that the authors misclassified this case and an intermediate JEB phenotype may become apparent in the future. The authors were contacted regarding this case; they confirmed that no IFM was available and that he passed away at 19 months. It may be possible that this patient died from causes not completely dependent on genotype.

Case 8 (p.R569X, c.943+2T>C heterozygote)

Individuals homozygous for p.R569X have been reported to have a severe JEB phenotype with early lethality (Khan et al., 2021). Other cases where the R569X mutation has been reported as a compound heterozygous mutation include Q243X/R569X, which also results in severe JEB (Kivirikko et al., 1996).

Thus the c.943+2T>C splice site mutation is likely to produce some partially functional laminin β3 chain. This supports the SpliceAI prediction of an in-frame transcript where 27 additional nucleotides are included from intron 9, producing a slightly lengthened 1181 amino acid laminin β3 chain, that could demonstrate reduced and intermittent (but not absent) staining at the DEJ compared to controls. Further support for this hypothesis includes the relatively low MMSplice score for exon skipping, which would result in an out-of-frame transcript. The resultant mutant protein from cryptic splice site activation, with 9 additional amino acids inserted at position 135, maps to the EGF domain and gives rise to an interesting phenotype where her skin was mostly spared The functional effects of this mutation do not appear to affect skin greatly but may be important in extra-cutaneous tissues such as teeth, eyes and nails, which were affected in her case.

Case 9 (c.298+5G>C homozygote)

The c.298+5G>C mutation has been reported in NCBI, but is not associated with any papers which may contain details regarding phenotype. Although this individual had a relatively large skin area affected, there were areas which did not blister, perhaps

suggesting revertant mosaicism. This is where the genetic defect is corrected naturally, and second site mutations have been demonstrated to be a mechanism for this (Pasmooij et al., 2007). Case 9 produced an additional 62bp transcript in cells from non-blistered skin areas, and reviewing the sequence, the donor splice site would have to be at Splice site B, 60nt within the intron (Figures 22-24). The wild type sequence at this position was not compatible with Chambon's rule (Breathnach et al., 1978), and therefore it was hypothesised whether a somatic mutation created a new splice site in a population of cells, allowing Transcript B to be produced. However, no second mutation could be detected on Sanger Sequencing. Further investigation, perhaps with more sensitive methods such as single cell sequencing, may provide further clues and help elucidate mechanisms for this interesting case. Alternatively, GT–TG splicing has been reported in the literature (Szafranski et al., 2007), and this is another possibility for consideration, although there was a very low SpliceAI prediction score (<0.01) for splicing to occur with this sequence.

DISCUSSION

5.1 PTC-introducing mutations and nonsense-mediated decay

A large number of mutations in our JEB cohort (12 mutations from 13 individuals) were nonsense mutations or out-of-frame insertions or deletions which introduced PTCs into RNA transcripts. These would be expected to either translate into truncated proteins, or be degraded by nonsense-mediated decay (NMD). The three rules used by MasonMD to predict NMD occurrence are predicted to account for almost three quarters of NMD variance, however the remainder of transcripts which escape NMD are not accounted for (Lindeboom et al., 2016). Therefore, more accurate in-silico approaches to predict whether NMD will occur are required, which is supported by findings from our JEB cohort.

Considering the cases in our cohort, only the c.151insG mutation would be predicted to escape NMD according to the three rules used by MasonMD, with all other nonsense and frameshift indels predicted to produce transcripts undergoing NMD. Indeed, some mutations identified in our cohort, such as p.R635X, have been demonstrated to undergo NMD completely, with undetectable levels of *LAMB3* RNA on Northern blotting (Pulkkinen et al., 1994). Nevertheless, as some amount of laminin-332 (albeit reduced) is detected on immunostaining in Case 5 (c.1702C>T; p.Q568X), it is expected that the PTC containing transcript has escaped NMD to some degree. Whilst Case 6 (c.1186_1196del; p.Thr396CysfsX12) did not have IFM results available, due to his intermediate phenotype, it can be speculated that the transcript escapes NMD to some degree to produce some partially functional laminin β3 chain polypeptide.

These two cases from our cohort demonstrate that decision tree prediction tools such as MasonMD are not sufficient to accurately predict NMD. This is also supported by cases reported in the literature (Buchroithner et al., 2004). Buchroithner et al describe a PTC in exon 21 of *LAMB3* caused by a c.3009C>T mutation. This PTC was 269 bp upstream of the last exon/exon border, and according to the rules above, would be expected to undergo NMD. However, it was demonstrated that the transcript escaped NMD and produced a slightly truncated laminin β3 polypeptide. Therefore, more sophisticated NMD prediction tools are required, which may also be able to provide quantitative predictions of the amount of transcripts which undergo NMD, or alternatively laboratory methods such as RT-PCR or RNA-seq can be used to establish whether NMD occurs, although these are more labour intensive and expensive.

Regarding the severe JEB cases in our cohort with PTC containing mutations, as it is currently not possible to accurately predict in which cases NMD occurs, it is not possible to speculate whether the lethal phenotype is secondary to NMD of mRNA transcripts or due to loss of key residues and domains due to truncation. For example, whilst p.R972X results in severe JEB, further investigation will be required to determine whether this is due to loss of residues p.973–p.1172, or if NMD occurs, preventing translation of the entire *LAMB3* transcript. Therefore, it is not possible to comment on which domains or residues are essential for protein function at this stage. Relevant areas for future investigation include collating cases where NMD escape occurs, and to evaluate whether there are any NMD resistant areas in the transcripts.

5.2 Homozygous mutations

13 of 17 individuals in our cohort were homozygotes, many of whom possessed previously unreported mutations. Potential explanations for homozygosity of rare or previously unreported mutations include uniparental isodisomy and consanguineous partnerships between parents of probands. Uniparental isodisomy is a potential disease-causing process where an individual inherits two copies of a chromosome from a single homolog of one parent, as a result of nondisjunction in meiosis II (Benn, 2021). This process results in the individual having two identical alleles for the loci of the chromosome where this has occurred, and can cause recessive Mendelian genetic disease if disease-causing recessive alleles are inherited in this fashion (Liehr, 2022). Although we did not specifically ask the families of probands whether consanguinity was present in this predominately retrospective study, this could be further explored in future prospective studies.

5.3 Missense mutations

No missense mutations were found in our cohort. Reviewing data from HGMD Professional 2021.1, missense mutations in *LAMA3*, *LAMB3*, *LAMC2* and *COL17A1* are associated with JEB, although less frequently than nonsense or frameshift mutations.

In the future, as more cases of JEB secondary to missense mutations are reported, analysis of these mutations may give unique insight into protein structure and function, as unlike PTC-introducing mutations where it is unclear if NMD occurs or whether the resulting protein is truncated, missense mutations directly alter one residue of the

protein only in a specific location. The mutation mapping pipeline developed as part of this project will be useful for analysis of these mutations. Additionally, tools such as AlphaFold2 can be used to analyse the effects of missense mutations on a protein's overall 3D structure (Jumper et al., 2021).

5.4 Splice site mutation analysis

5.4.1 Limitations of in-silico splice site prediction tools

Comparing the tools, MMSplice was limited in that it was unable to predict whether and where cryptic splicing would occur. Furthermore, for Cases 9 and 11, MMsplice strongly predicted exon skipping whilst SpliceAI favoured cryptic splice site activation. Considering the RNA data for these two cases, SpliceAI was correct (to some extent) with its predictions for cryptic splice site activation for both cases, although in reality exon skipping and cryptic splice site activation are unlikely to be mutually exclusive, with multiple transcripts often being generated from a splice site mutation; this is discussed further below.

It is important to note that these tools were used to analyse five splice site mutations only. Confirmatory RNA data was available in two cases, and in the future, it is hoped that RNA analysis for Case 8 (c.1705C>T/c.943+2T>C) will be performed. RNA results from the deceased severe JEB cases are not available, which is a limitation of retrospectively reviewing available data. Investigation and validation with larger numbers of cases will be required to explore the accuracy of these tools in predicting transcript outcomes from splice site mutations. With a larger number of cases, algorithm outputs could be further refined. For example, it is pertinent to investigate

whether SpliceAI delta scores correlate with how much of each transcript is being produced, which will affect the amount of functional or partially functional protein produced, influencing phenotype.

In a recent study where in-silico tools were used to classify 249 variants of unknown significance into whether they impacted splicing or not, SpliceAI outperformed seven other in-silico splice mutation prediction tools, including MMSplice (Rowlands et al., 2021). For this task, the sensitivity (0.91) and specificity (0.90) of SpliceAI was greatest with a threshold of any delta score being greater than 0.145. Similar accuracy was found when SpliceAI was used to classify variants in the *NF1* gene; it achieved a sensitivity of 94.5 and specificity of 94.3 with a threshold of 0.22 (Ha et al., 2021).

However, in both studies the authors did not detail whether the correct transcripts were predicted; accuracy was only measured through the number of true positives (splicing impacted) and true negatives (splicing not impacted).

A further limitation of SpliceAI is that only the single greatest delta score for each category (donor gain or loss, acceptor gain or loss) is shown. This conceals other predictions which may not be as great, but may still have a significant biological effect by producing different transcripts. In reality, multiple transcripts may be generated following a splice site mutation in various different quantities (Nakano et al., 2002, Kiritsi et al., 2011). Certain transcripts may not be the most abundant, but may produce a functionally significant amount of protein to alleviate JEB severity (Hou et al., 2021). Therefore, techniques such as RNA-seq, which are able to not only detect multiple

transcripts for each gene but also quantify the amount of each one, would be well suited to further investigation at the RNA level. The quantification of transcripts is vital as the amount of partially functional peptide produced influences the resulting phenotype. An additional benefit of RNA-seq is that up or down regulation of other genes encoding proteins which interact with laminin-332 can also be evaluated (Wang et al., 2009).

Finally, isoform expression in different tissues varies, and therefore splice site mutations may have potentially different effects in different tissues. More sophisticated models in the future may address tissue specific splice site mutation prediction (Jaganathan et al., 2019), and data from the human cell atlas could potentially be used for algorithm development (Regev et al., 2017).

5.4.2 RNA analysis provides valuable data at the molecular level

Current EB diagnostic services in the UK focus on DNA sequencing to confirm the genetic defect, which allows classification into EB type depending on which gene the mutation is present in. IFM can also confirm the level of split, and presence of target protein. This study highlights that evaluating gene transcripts is an important further step in genotype-phenotype correlation. Potential techniques which can be used to investigate resultant transcripts include:

1. In-silico prediction e.g. with SpliceAI

The accuracy of transcript prediction in JEB is yet to be explored and formally validated; however, in the two cases from our cohort where RT-PCR data was

available (Cases 9 and 11), SpliceAI did correctly predict the main transcript in both cases. Once established, in-silico tools have the advantage of being quick and cheap, but require validation. Once validated, they may be suitable for routine clinical use. They may also facilitate selection of cases for further laboratory investigation and guide which regions of the gene or transcript to sequence for confirmatory analyses.

2. RT-PCR

This can identify the transcripts that are produced as the result of a splice site mutation, as have been done for Case 9. The length of the transcript, and its reading frame can help guide prediction of disease severity and phenotype. However, it is low throughput, and is unable to quantify the amount of each transcript produced (Wang et al., 2009).

3. RNA-seq

Importantly, this technique not only allows identification of transcripts, but will also allow quantification of each transcript identified (Wang et al., 2009). This is relevant in JEB as the amount of partially-functional protein produced may correlate with disease severity and phenotype (Has et al., 2020a). Furthermore, RNA-seq gives a global overview of gene expression within a tissue or cell (if single cell RNA-seq), and will therefore give important information about other genes. For example, genes encoding other hemidesmosomal or basement membrane proteins may be up or downregulated to compensate for non-functional or partially functional laminin-332. Additionally, RNA-seq data could be used as a dataset to train, validate or fine-tune deep learning AI algorithms for splice site mutation outcome prediction. However, RNA-seq is the most expensive technique, and whilst it generates the most data, requires a significant

amount of bioinformatic data processing. Currently, it is perhaps most suited for research purposes but not routine clinical use.

Other cases in our cohort where RT-PCR or RNA-seq would be especially useful include Cases 5 and 6, where biallelic null mutations have resulted in intermediate JEB. Again, investigations at the RNA level are likely to shed light on the molecular mechanisms underpinning amelioration of expected severity, and in the future, these mechanisms could be investigated further with the aim of developing novel therapeutic agents that can exploit these mechanisms.

5.4.3 Utilisation of in-silico splice site prediction tools for personalised gene therapy

Accurate prediction of how altered sequence affects modulation of exon skipping is important in the development of personalised gene therapies for EB. This is especially relevant for antisense oligonucleotides (ASOs), a number of which have recently been licenced for clinical use in rare genetic diseases, such as spinal muscular atrophy and Duchenne muscular dystrophy (Dhuri et al., 2020). ASOs are short strands of synthesised DNA or RNA which are complementary to a specific mRNA sequence. Through binding to and silencing key mRNA motifs required for splicing (such as splice sites, splicing enhancers and silencers), ASOs can modulate exon skipping to include or exclude targeted exons, and any mutations that they may contain (Rinaldi and Wood, 2018).

5.4.3.1 Modulation of splice site location

Ablinger et al describe a JEB genotype consisting of a homozygous acceptor splice site mutation (c.380-1G>A) at the intron 6/exon 7 junction of *COL17A1* which disrupts the consensus splice acceptor AG dinucleotide and results in cryptic splice site activation 16 nucleotides downstream (c.395) within exon 7, where another AG dinucleotide was present (Ablinger et al., 2021). This resulted in the production of an out-of-frame transcript and an intermediate JEB phenotype. ASOs were designed to bind to the cryptic splice site, preventing its use and ultimately resulting in splicing of intron 6, exon 7 and intron 7. The overall result was in-frame skipping of exon 7, and restoration of the open reading frame. This restored COL17 expression in vitro, highlighting this as a potential therapeutic strategy for JEB patients.

This approach can also be adapted for use in participants from our cohort, for example in Case 11. In the mutated sequence, a novel cryptic splice site is predicted by SpliceAI to be formed at position c.629-10, which is stronger than the original splice site. Similar to the case described by Ablinger et al, the novel cryptic splice site at c.629-10 could be targeted and inactivated by an ASO, which may result in increased use of the original leaky splice site (where the native AG acceptor dinucleotide remains), producing greater amounts or normal mRNA transcript. As illustrated by these examples, in-silico approaches can be used to identify cases which may be suitable for particular gene therapy strategies.
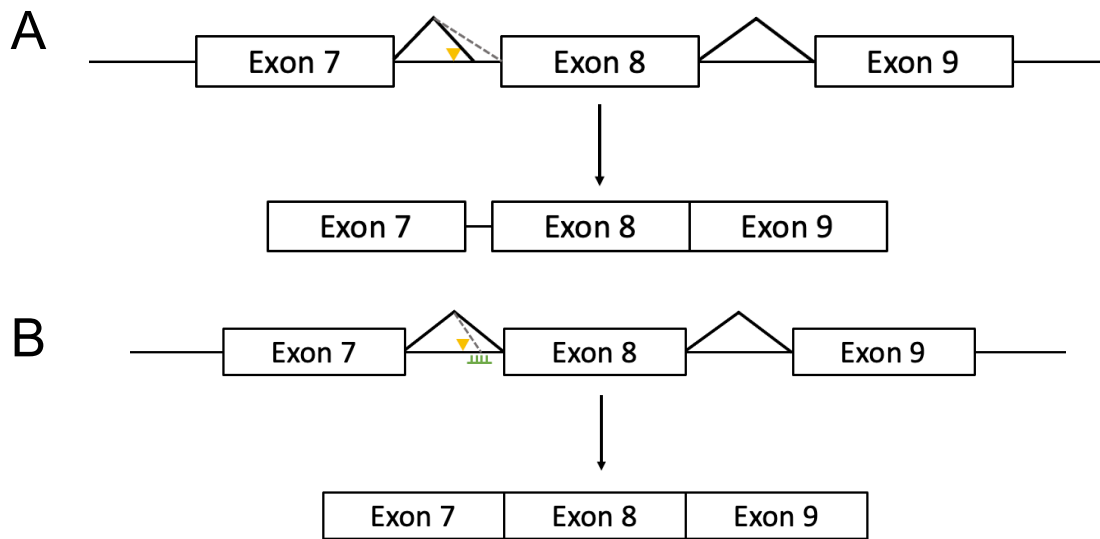
Figure 29: Targeted ASO therapy to direct splicing. A) The c.629-12T>A mutation (yellow triangle) results in activation of a cryptic splice site at c.629-10 which leads to inclusion of 10 additional nucleotides from intron 7 which results in an out-of-frame transcript predominantly. B) Targeting of the cryptic splice site by an ASO (green) may reduce the likelihood of it being used, resulting in greater use of the original splice site, producing greater amounts of in-frame wild type transcript.

5.4.3.2 Targeted skipping of null mutation containing exons

ASOs can also be used to modulate splicing so that in-frame exons that contain disease-causing mutations (such as missense, nonsense or small frameshift indels) are skipped. Using keratinocytes and fibroblasts from an RDEB patient with a c.7828C>T (p.R2610X) mutation in exon 105 of *COL7A1*, Bremer et al designed ASOs to target exonic splicing enhancers (ESEs) in exon 105 of *COL7A1* (Bremer et al., 2016). Without treatment, the mutation resulted in NMD of transcripts and the absence of *COL7A1* expression on immunostaining. With ASOs, splicing was modulated so that exon 105 skipping and exclusion of the c.7828C>T mutation occurred, which resulted in production of an exon 105 deleted transcript that was in-frame and translated into type VII collagen protein visible on IFM. This was also replicated in vivo using a mouse model.

In this case, in-silico prediction allowed the identification and targeting of ESEs necessary for exon 105 splicing. The authors comment that exon 105 skipping is likely to be tolerated, as it is only one of 84 in-frame exons that encode nine of 454 Gly-X-Y repeats which make up the Collagen 7 triple helix domain, where length is not essential for function. This was further supported by real world examples of relatively mild RDEB phenotypes, where natural in-frame *COL7A1* exon skipping occurred (Toyonaga et al., 2015).

The majority of ASO applications in EB so far involve skipping of entire in-frame exons. This approach is only feasible for targeting exons that are in-frame, which are relatively abundant in *COL17A1* and *COL7A1*, but less so in *LAMA3*, *LAMB3* or *LAMC2,* although not completely non-existent. Nevertheless, with accurate prediction of splicing modulation, mutation containing regions within out-of-frame exons could also be targeted by activating in-frame exonic cryptic splice sites, which would splice out nonsense or indel mutations (Figure 30).

Alternatively, multiple exons could be skipped to produce a resultant in-frame transcript. This has been reported to occur in natural in-frame exon skipping of *LAMC2*, where the p.R245X mutation generated alternately spliced products where either exon 6 alone was skipped, or exons 4-7 were skipped together (Nakano et al., 2002). Both transcripts preserved the reading frame, although when considering therapeutic strategies, skipping of multiple exons is less desirable as larger regions of the polypeptide would be deleted.
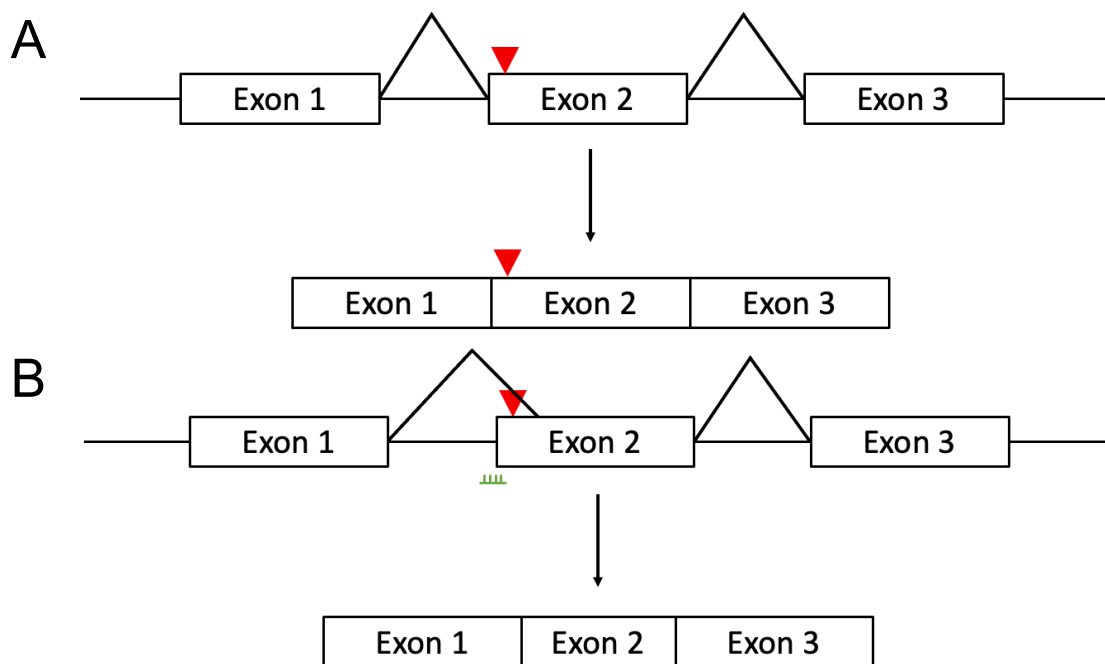
Figure 30: ASO modulation may allow targeted in-frame skipping of mutations. A nonsense or indel mutation is represented by the red triangle. A) The mutation is included with normal splicing which leads to a PTC being present in the transcript. B) ASO modulation of splicing results in activation of an in-frame exonic cryptic splice site. The 5' end of exon 2 is spliced out, which contains the mutation. The remainder of the transcript is in-frame.

5.5 JEB Phenotyping tool

Cutaneous, nail, mucosal and systemic extra-cutaneous manifestations of JEB were evident in individuals with intermediate JEB who underwent deep phenotyping. Variation in severity of characteristics within and between categories highlighted different phenotypes within the intermediate JEB subtype. Six individuals were found to have atrophic scarring (a feature not included in the BEBS), and this characteristic may be considered for inclusion into future JEB phenotyping systems.

When designing the phenotyping tool, subjective features such as pain and itch were excluded and characteristics which could be scored objectively were selected. However, examination at one timepoint gave only a snapshot of disease activity and

severity, and consideration must be given to the possibility of characteristics changing or progressing over time. For example, whilst Case 8 did not have any skin involvement at time of examination, she did have blistering previously. Therefore, longitudinal phenotyping at standardised time points may allow more accurate deep phenotype assessment and monitoring of disease course. Individuals with the same genotype (14 and 15, 16 and 17) had similar, but slightly different phenotypes. This could have been due to changes with age, or the presence of disease modifying factors (Section 5.6), or both. Another limitation of the tool was that whilst scores allowed comparison between individuals within a category e.g., severity of nail involvement, scores could not be compared between categories to reflect overall disease severity, for example loss of all nails (nails score=5) does not equate to laryngeal obstruction (larynx score=5).

The phenotyping tool could be refined by using a structured methodology such as the Delphi method (Engelman et al., 2018) to achieve consensus opinion on which parameters to include. Furthermore, formal testing and validation would be required before widespread use.

5.6 Rationale for a JEB genotype-phenotype database

With advances in sequencing technologies, novel mutations and further exceptions to existing genotype-phenotype correlations will inevitably be detected. Due to the rarity of JEB, collation of cases is necessary for identification of disease patterns. Genotypes and corresponding phenotypes are often mentioned in individual articles, and as the

data is disparate, it is arduous for clinicians and researchers to efficiently search and access data on genotypes and phenotypes in a single place. Therefore, mutation databases are required for comparison and analysis of data (Bardhan et al., 2020). Evaluation of current databases, such as HGMD, have revealed that nomenclature is outdated, limited phenotype information exists, and importantly data regarding both alleles (i.e. the full genotype) is not reported. Whilst a database for RDEB has been established (van den Akker et al., 2011), currently, no publicly available databases or registries exist specifically for JEB.

Creation of a mutation database will allow collation, comparison and analysis of existing genetic variants in a systematic fashion, and will also allow data sharing for the clinical and scientific EB community, rather than silo data in particular institutions. It will allow easy and comprehensive access to reported cases, and this will help prognostication of individuals with mutations that are already reported in the database. Numerous cases in our cohort (particularly Cases 4 and 11) demonstrate that reported mutations can be used as a guide for predicting the phenotypes of individuals who have the same mutation. Additionally, establishing a database of pathogenic mutations in JEB in the appropriate format with uniform annotations allows systematic testing of in-silico tools for accuracy and validation, and also provides an up-to-date dataset for subsequent bioinformatic analyses of how mutations affect corresponding protein structure and function.

Furthermore, when available in the future, targeted precision therapies for patients will be evaluated on a case by case basis and will depend on an individual's specific

molecular defect (Condrat et al., 2018). A mutation database will allow stratification and sub-classification of genetically and molecularly heterogenous patients into further subgroups based on their precise genetic abnormalities and consequently enable matching of patients to appropriate treatments in the future. As mentioned, examples of therapies currently in development include ASO-mediated exon skipping of in-frame exons which contain pathogenic mutations, (Bremer et al., 2016, Bornert et al., 2021), which will only be feasible for non-essential exons that can be spliced out without major deleterious consequences for resulting protein structure and function.

5.7 Further considerations for the future

A limitation of this study was that it was retrospective, and therefore a complete dataset for each participant which included DNA, RNA, IFM and deep phenotype data was unavailable (Table 15). This limited the number of hypothesised relationships that could be confirmed and sadly not all predicted transcripts from the in-silico tools could be validated by laboratory studies. Future prospective studies systematically collecting data in each of these areas, along with greater case numbers, will allow more meaningful comparison and elucidation of genotype-phenotype correlation.

There are many layers and facets to gene expression and protein function. Other processes which have not yet been discussed but may potentially influence phenotype include:

1. Histone modifications and DNA methylation (Cavalli and Heard, 2019)
2. RNA modifications and transcript stability e.g., modifications via M6 methyl-adenosine (Meyer and Jaffrey, 2014)

Case	DNA mutation consequence	RNA prediction (splice site)	RNA	IFM	JEB subtype	Deep phenotype
1	LAMA3 fs PTC	N/A	Unavailable	Complete absence laminin-332	Severe	Unavailable (Deceased)
2	LAMA3 fs PTC	N/A	Unavailable	Complete absence laminin-332	Severe	Unavailable (Deceased)
3	LAMA3 fs PTC	N/A	Unavailable	Unavailable	Severe	Unavailable (Deceased)
4	LAMB3 Nonsense	N/A	Unavailable	Complete absence laminin-332	Severe	Unavailable (Deceased)
5	LAMB3 Nonsense	N/A	Unavailable	Marked reduction in staining	Intermediate	Yes
6	LAMB3 fs PTC	N/A	Unavailable	Unavailable	Intermediate	Yes
7	LAMB3 Splice	Out of frame	Unavailable	Unavailable	Severe	Unavailable (Deceased)
8	LAMB3 Splice / nonsense	Out of frame and in frame both possible	Unavailable	Reduction in staining	Intermediate	Yes
9	LAMB3 Splice	Out of frame	Both in and out of frame transcripts generated	Substantially reduced staining (overall)	Intermediate	Yes
10	LAMB3 Splice / nonsense	Out of frame and in frame with PTC	Unavailable	Unavailable	Severe	Unavailable (Deceased)
11	LAMB3 Splice / nonsense	Out of frame	Unavailable	Unavailable	Intermediate	Yes
12	LAMC2 fs PTC	N/A	Unavailable	Unavailable	Severe	Unavailable (Deceased)
13	LAMC2 fs PTC	N/A	Unavailable	Complete absence laminin-332	Severe	Unavailable (Deceased)
14	COL17A1 fs PTC	N/A	Unavailable	Unavailable	Intermediate	Yes
15	COL17A1 fs PTC	N/A	Unavailable	Absence of COL17	Intermediate	Yes
16	COL17A1 Large deletion	N/A	Unavailable	Unavailable	Intermediate	Yes
17	COL17A1 Large deletion	N/A	Unavailable	Unavailable	Intermediate	Yes

Table 15: Summary of available data for this study. Unavailable data is highlighted in grey.

3. Post-translational protein modifications e.g., COL17 phosphorylation, glycosylation and ectodomain cleavage (Has et al., 2018)

4. Other disease modifying factors such as cytokine expression and signalling. For example, Odorisio et al found that differing modulation of TGFβ signalling resulted in divergent phenotypes in monozygotic RDEB twins (Odorisio et al., 2014)

These are beyond the scope of this thesis but are relevant areas for exploration in the future.

CONCLUSION

Accurate genotype-phenotype correlation has important clinical implications for JEB patients. Accurate predictions are challenging as a number of cases have a phenotype that is less severe than expected from their genetic defect, and this is also demonstrated in our cohort. The findings of this project highlight that review of reported cases in the literature allows phenotype prediction and this can be streamlined with collation of cases in a dedicated JEB database. In-silico approaches are promising for analysis of functional effects of mutations and currently may directly identify candidates for confirmatory laboratory investigation. When further refined and validated, these in-silico tools may be suitable for routine clinical use. Concerning novel mutations, investigation of RNA transcripts will help to further elucidate genotype-phenotype correlations. Finally, the systematic approaches used in this study can be applied to available genetic data in the literature, and also other types of EB and other mendelian genetic diseases.

REFERENCES

ABERDAM, D., GALLIANO, M. F., VAILLY, J., PULKKINEN, L., BONIFAS, J., CHRISTIANO, A. M., TRYGGVASON, K., UITTO, J., EPSTEIN, E. H., JR., ORTONNE, J. P. & ET AL. 1994. Herlitz's junctional epidermolysis bullosa is linked to mutations in the gene (LAMC2) for the gamma 2 subunit of nicein/kalinin (LAMININ-5). *Nat Genet,* 6**,** 299-304.

ABLINGER, M., LETTNER, T., FRIEDL, N., POTOCKI, H., PALMETZHOFER, T., KOLLER, U., ILLMER, J., LIEMBERGER, B., HAINZL, S., KLAUSEGGER, A., REISENBERGER, M., LAMBERT, J., VAN GELE, M., DESMET, E., VAN MAELSAEKE, E., WIMMER, M., ZAUNER, R., BAUER, J. W. & WALLY, V. 2021. Personalized Development of Antisense Oligonucleotides for Exon Skipping Restores Type XVII Collagen Expression in Junctional Epidermolysis Bullosa. *Int J Mol Sci,* 22.

ADAMSON, S. I., ZHAN, L. & GRAVELEY, B. R. 2018. Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. *Genome Biol,* 19**,** 71.

ANGELIS, A., KANAVOS, P., LÓPEZ-BASTIDA, J., LINERTOVÁ, R., OLIVA-MORENO, J., SERRANO-AGUILAR, P., POSADA-DE-LA-PAZ, M., TARUSCIO, D., SCHIEPPATI, A., ISKROV, G., BRODSZKY, V., VON DER SCHULENBURG, J. M., CHEVREUL, K., PERSSON, U., FATTORE, G. & NETWORK, B.-R. R. 2016. Social/economic costs and health-related quality of life in patients with epidermolysis bullosa in Europe. *Eur J Health Econ,* 17 Suppl 1**,** 31-42.

AYOUB, N., TOMB, R., CHARLESWORTH, A. & MENEGUZZI, G. 2005. [Junctional epidermolysis bullosa. Identification of a new mutation in two Lebanese families]. *Ann Dermatol Venereol,* 132**,** 550-3.

BARDHAN, A., BRUCKNER-TUDERMAN, L., CHAPPLE, I. L. C., FINE, J. D., HARPER, N., HAS, C., MAGIN, T. M., MARINKOVICH, M. P., MARSHALL, J. F., MCGRATH, J. A., MELLERIO, J. E., POLSON, R. & HEAGERTY, A. H. 2020. Epidermolysis bullosa. *Nat Rev Dis Primers,* 6**,** 78.

BARKER, J., BLEIKER, T. O., CHALMERS, R., GRIFFITHS, C. E. M. & CREAMER, D. 2016. *Rook's textbook of dermatology*, John Wiley & Sons.

BENN, P. 2021. Uniparental disomy: Origin, frequency, and clinical significance. *Prenat Diagn,* 41**,** 564-572.

BLUM, M., CHANG, H. Y., CHUGURANSKY, S., GREGO, T., KANDASAAMY, S., MITCHELL, A., NUKA, G., PAYSAN-LAFOSSE, T., QURESHI, M., RAJ, S., RICHARDSON, L., SALAZAR, G. A., WILLIAMS, L., BORK, P., BRIDGE, A., GOUGH, J., HAFT, D. H., LETUNIC, I., MARCHLER-BAUER, A., MI, H., NATALE, D. A., NECCI, M., ORENGO, C. A., PANDURANGAN, A. P., RIVOIRE, C., SIGRIST, C. J. A., SILLITOE, I., THANKI, N., THOMAS, P. D., TOSATTO, S. C. E., WU, C. H., BATEMAN, A. & FINN, R. D. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res,* 49**,** D344-D354.

BORNERT, O., HOGERVORST, M., NAUROY, P., BISCHOF, J., SWILDENS, J., ATHANASIOU, I., TUFA, S. F., KEENE, D. R., KIRITSI, D., HAINZL, S., MURAUER, E. M., MARINKOVICH, M. P., PLATENBURG, G., HAUSSER, I., WALLY, V., RITSEMA, T., KOLLER, U., HAISMA, E. M. & NYSTROM, A. 2021. QR-313, an Antisense Oligonucleotide, Shows Therapeutic Efficacy for Treatment of Dominant and Recessive Dystrophic Epidermolysis Bullosa: A Preclinical Study. *J Invest Dermatol,* 141**,** 883-893 e6.

BREATHNACH, R., BENOIST, C., O'HARE, K., GANNON, F. & CHAMBON, P. 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc Natl Acad Sci U S A,* 75**,** 4853-7.

BREMER, J., BORNERT, O., NYSTROM, A., GOSTYNSKI, A., JONKMAN, M. F., AARTSMA-RUS, A., VAN DEN AKKER, P. C. & PASMOOIJ, A. M. 2016. Antisense Oligonucleotide-mediated Exon Skipping as a Systemic Therapeutic Approach for Recessive Dystrophic Epidermolysis Bullosa. *Mol Ther Nucleic Acids,* 5**,** e379.

BRUNAK, S., ENGELBRECHT, J. & KNUDSEN, S. 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol,* 220**,** 49-65.

BUCHROITHNER, B., KLAUSEGGER, A., EBSCHNER, U., ANTON-LAMPRECHT, I., POHLA-GUBO, G., LANSCHUETZER, C. M., LAIMER, M., HINTNER, H. & BAUER, J. W. 2004. Analysis of the LAMB3 gene in a junctional epidermolysis bullosa patient reveals exonic splicing and allele-specific nonsense-mediated mRNA decay. *Lab Invest,* 84**,** 1279-88.

CASTORI, M., FLORIDDIA, G., DE LUCA, N., PASCUCCI, M., GHIRRI, P., BOCCALETTI, V., EL HACHEM, M., ZAMBRUNO, G. & CASTIGLIA, D. 2008. Herlitz junctional epidermolysis bullosa: laminin-5 mutational profile and carrier frequency in the Italian population. *Br J Dermatol,* 158**,** 38-44.

CAVALLI, G. & HEARD, E. 2019. Advances in epigenetics link genetics to the environment and disease. *Nature,* 571**,** 489-499.

CHEN, F., HUANG, L., LI, C., ZHANG, J., YANG, W., ZHANG, B., LI, H., DENG, D., LIANG, J., SHEN, J., YAO, Z. & LI, M. 2020. Next-generation sequencing through multigene panel testing for the diagnosis of hereditary epidermolysis bullosa in Chinese population. *Clin Genet,* 98**,** 179-184.

CHENG, J., NGUYEN, T. Y. D., CYGAN, K. J., CELIK, M. H., FAIRBROTHER, W. G., AVSEC, Z. & GAGNEUR, J. 2019. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol,* 20**,** 48.

CONDRAT, I., HE, Y., COSGAREA, R. & HAS, C. 2018. Junctional Epidermolysis Bullosa: Allelic Heterogeneity and Mutation Stratification for Precision Medicine. *Front Med (Lausanne),* 5**,** 363.

CONSORTIUM, G. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet,* 45**,** 580-5.

DHURI, K., BECHTOLD, C., QUIJANO, E., PHAM, H., GUPTA, A., VIKRAM, A. & BAHAL, R. 2020. Antisense Oligonucleotides: An Emerging Area in Drug Discovery and Development. *J Clin Med,* 9.

ENGELMAN, D., FULLER, L. C., STEER, A. C. & INTERNATIONAL ALLIANCE FOR THE CONTROL OF SCABIES DELPHI, P. 2018. Consensus criteria for the diagnosis of scabies: A Delphi study of international experts. *PLoS Negl Trop Dis,* 12**,** e0006549.

FINE, J. D. 2016. Epidemiology of Inherited Epidermolysis Bullosa Based on Incidence and Prevalence Estimates From the National Epidermolysis Bullosa Registry. *JAMA Dermatol,* 152**,** 1231-1238.

FINE, J. D., BRUCKNER-TUDERMAN, L., EADY, R. A., BAUER, E. A., BAUER, J. W., HAS, C., HEAGERTY, A., HINTNER, H., HOVNANIAN, A., JONKMAN, M. F., LEIGH, I., MARINKOVICH, M. P., MARTINEZ, A. E., MCGRATH, J. A., MELLERIO, J. E., MOSS, C., MURRELL, D. F., SHIMIZU, H., UITTO, J., WOODLEY, D. & ZAMBRUNO, G. 2014. Inherited epidermolysis bullosa: updated recommendations on diagnosis and classification. *J Am Acad Dermatol,* 70**,** 1103-26.

FINE, J. D., EADY, R. A., BAUER, E. A., BAUER, J. W., BRUCKNER-TUDERMAN, L., HEAGERTY, A., HINTNER, H., HOVNANIAN, A., JONKMAN, M. F., LEIGH, I., MCGRATH, J. A., MELLERIO, J. E., MURRELL, D. F., SHIMIZU, H., UITTO, J., VAHLQUIST, A., WOODLEY, D. & ZAMBRUNO, G. 2008. The classification of inherited epidermolysis bullosa (EB): Report of the Third International Consensus Meeting on Diagnosis and Classification of EB. *J Am Acad Dermatol,* 58**,** 931-50.

FINE, J. D. & MELLERIO, J. E. 2009a. Extracutaneous manifestations and complications of inherited epidermolysis bullosa: part I. Epithelial associated tissues. *J Am Acad Dermatol,* 61**,** 367-84; quiz 385-6.

FINE, J. D. & MELLERIO, J. E. 2009b. Extracutaneous manifestations and complications of inherited epidermolysis bullosa: part II. Other organs. *J Am Acad Dermatol,* 61**,** 387-402; quiz 403-4.

FRANKISH, A., DIEKHANS, M., FERREIRA, A. M., JOHNSON, R., JUNGREIS, I., LOVELAND, J., MUDGE, J. M., SISU, C., WRIGHT, J., ARMSTRONG, J., BARNES, I., BERRY, A., BIGNELL, A., CARBONELL SALA, S., CHRAST, J., CUNNINGHAM, F., DI DOMENICO, T., DONALDSON, S., FIDDES, I. T., GARCIA GIRON, C., GONZALEZ, J. M., GREGO, T., HARDY, M., HOURLIER, T., HUNT, T., IZUOGU, O. G., LAGARDE, J., MARTIN, F. J., MARTINEZ, L., MOHANAN, S., MUIR, P., NAVARRO, F. C. P., PARKER, A., PEI, B., POZO, F., RUFFIER, M., SCHMITT, B. M., STAPLETON, E., SUNER, M. M., SYCHEVA, I., USZCZYNSKA-RATAJCZAK, B., XU, J., YATES, A., ZERBINO, D., ZHANG, Y., AKEN, B., CHOUDHARY, J. S., GERSTEIN, M., GUIGO, R., HUBBARD, T. J. P., KELLIS, M., PATEN, B., REYMOND, A., TRESS, M. L. & FLICEK, P. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res,* 47**,** D766-D773.

GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. & BAIROCH, A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res,* 31**,** 3784-8.

GISH, W. & STATES, D. J. 1993. Identification of protein coding regions by database similarity search. *Nat Genet,* 3**,** 266-72.

HA, C., KIM, J. W. & JANG, J. H. 2021. Performance Evaluation of SpliceAI for the Prediction of Splicing of NF1 Variants. *Genes (Basel),* 12.

HAMMERSEN, J., HAS, C., NAUMANN-BARTSCH, N., STACHEL, D., KIRITSI, D., SÖDER, S., TARDIEU, M., METZLER, M., BRUCKNER-TUDERMAN, L. & SCHNEIDER, H. 2016. Genotype, Clinical Course, and Therapeutic Decision Making in 76 Infants with Severe Generalized Junctional Epidermolysis Bullosa. *J Invest Dermatol,* 136**,** 2150-2157.

HAS, C., BAUER, J. W., BODEMER, C., BOLLING, M. C., BRUCKNER-TUDERMAN, L., DIEM, A., FINE, J. D., HEAGERTY, A., HOVNANIAN, A., MARINKOVICH, M. P., MARTINEZ, A. E., MCGRATH, J. A., MOSS, C., MURRELL, D. F., PALISSON, F., SCHWIEGER-BRIEL, A., SPRECHER, E., TAMAI, K., UITTO, J., WOODLEY, D. T., ZAMBRUNO, G. & MELLERIO, J. E. 2020a. Consensus reclassification of inherited epidermolysis bullosa and other disorders with skin fragility. *Br J Dermatol*.

HAS, C. & BRUCKNER-TUDERMAN, L. 2014. The genetics of skin fragility. *Annu Rev Genomics Hum Genet,* 15**,** 245-68.

HAS, C. & HE, Y. 2016. Research Techniques Made Simple: Immunofluorescence Antigen Mapping in Epidermolysis Bullosa. *J Invest Dermatol,* 136**,** e65-e71.

HAS, C., LIU, L., BOLLING, M. C., CHARLESWORTH, A. V., EL HACHEM, M., ESCÁMEZ, M. J., FUENTES, I., BÜCHEL, S., HIREMAGALORE, R., POHLA-GUBO, G., VAN DEN AKKER, P. C., WERTHEIM-TYSAROWSKA, K. & ZAMBRUNO, G. 2020b. Clinical practice guidelines for laboratory diagnosis of epidermolysis bullosa. *Br J Dermatol,* 182**,** 574-592.

HAS, C., NYSTROM, A., SAEIDIAN, A. H., BRUCKNER-TUDERMAN, L. & UITTO, J. 2018. Epidermolysis bullosa: Molecular pathology of connective tissue components in the cutaneous basement membrane zone. *Matrix Biol,* 71-72**,** 313-329.

HEAGERTY, A. H., KENNEDY, A. R., EADY, R. A., HSI, B. L., VERRANDO, P., YEH, C. J. & ORTONNE, J. P. 1986. GB3 monoclonal antibody for diagnosis of junctional epidermolysis bullosa. *Lancet,* 1**,** 860.

HERISSE, A. L., CHARLESWORTH, A., BELLON, N., LECLERC-MERCIER, S., BOURRAT, E., HADJ-RABIA, S., BODEMER, C., LACOUR, J. P. & CHIAVERINI, C. 2021. Genotypic and phenotypic analysis of 34 cases of inherited junctional epidermolysis bullosa caused by COL17A1 mutations. *Br J Dermatol*.

HOU, P. C., NATSUGA, K., TU, W. T., HUANG, H. Y., CHEN, B., CHEN, L. Y., CHEN, W. R., HONG, Y. K., TANG, Y. A., LEE, J. Y., CHEN, P. C., SUN, H. S., MCGRATH, J. A. & HSU, C. K. 2021. Complexity of Transcriptional and Translational Interference of Laminin-332 Subunits in Junctional Epidermolysis Bullosa with LAMB3 Mutations. *Acta Derm Venereol,* 101**,** adv00522.

HOWE, K. L., ACHUTHAN, P., ALLEN, J., ALLEN, J., ALVAREZ-JARRETA, J., AMODE, M. R., ARMEAN, I. M., AZOV, A. G., BENNETT, R., BHAI, J., BILLIS, K., BODDU, S., CHARKHCHI, M., CUMMINS, C., DA RIN FIORETTO, L., DAVIDSON, C., DODIYA, K., EL HOUDAIGUI, B., FATIMA, R., GALL, A., GARCIA GIRON, C., GREGO, T., GUIJARRO-CLARKE, C., HAGGERTY, L., HEMROM, A., HOURLIER, T., IZUOGU, O. G., JUETTEMANN, T., KAIKALA, V., KAY, M., LAVIDAS, I., LE, T., LEMOS, D., GONZALEZ MARTINEZ, J., MARUGAN, J. C., MAUREL, T., MCMAHON, A. C., MOHANAN, S., MOORE, B., MUFFATO, M., OHEH, D. N., PARASCHAS, D., PARKER, A., PARTON, A., PROSOVETSKAIA, I., SAKTHIVEL, M. P., SALAM, A. I. A., SCHMITT, B. M., SCHUILENBURG, H., SHEPPARD, D., STEED, E., SZPAK, M., SZUBA, M., TAYLOR, K., THORMANN, A., THREADGOLD, G., WALTS, B., WINTERBOTTOM, A., CHAKIACHVILI, M., CHAUBAL, A., DE SILVA, N., FLINT, B., FRANKISH, A., HUNT, S. E., GR, I. I., LANGRIDGE, N., LOVELAND, J. E., MARTIN, F. J., MUDGE, J. M., MORALES, J., PERRY, E., RUFFIER, M., TATE, J., THYBERT, D., TREVANION, S. J., CUNNINGHAM, F., YATES, A. D., ZERBINO, D. R. & FLICEK, P. 2021. Ensembl 2021. *Nucleic Acids Res,* 49**,** D884-D891.

HU, Z., YAU, C. & AHMED, A. A. 2017. A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay. *Nat Commun,* 8**,** 15943.

HUG, N., LONGMAN, D. & CACERES, J. F. 2016. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res,* 44**,** 1483-95.

HUNT, S. E., MCLAREN, W., GIL, L., THORMANN, A., SCHUILENBURG, H., SHEPPARD, D., PARTON, A., ARMEAN, I. M., TREVANION, S. J., FLICEK, P. & CUNNINGHAM, F. 2018. Ensembl variation resources. *Database (Oxford),* 2018.

JACKOW, J., LOFFEK, S., NYSTROM, A., BRUCKNER-TUDERMAN, L. & FRANZKE, C. W. 2016. Collagen XVII Shedding Suppresses Re-Epithelialization by Directing Keratinocyte Migration and Dampening mTOR Signaling. *J Invest Dermatol,* 136**,** 1031-1041.

JAGANATHAN, K., KYRIAZOPOULOU PANAGIOTOPOULOU, S., MCRAE, J. F., DARBANDI, S. F., KNOWLES, D., LI, Y. I., KOSMICKI, J. A., ARBELAEZ, J., CUI, W., SCHWARTZ, G. B., CHOW, E. D., KANTERAKIS, E., GAO, H., KIA, A., BATZOGLOU, S., SANDERS, S. J. & FARH, K. K. 2019. Predicting Splicing from Primary Sequence with Deep Learning. *Cell,* 176**,** 535-548.e24.

JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ZIDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T., PETERSEN, S., REIMAN, D., CLANCY, E., ZIELINSKI, M., STEINEGGER, M., PACHOLSKA, M., BERGHAMMER, T., BODENSTEIN, S., SILVER, D., VINYALS, O., SENIOR, A. W., KAVUKCUOGLU, K., KOHLI, P. & HASSABIS, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596**,** 583-589.

KHAN, F. F., KHAN, N., REHMAN, S., EJAZ, A., ALI, U., ERFAN, M., AHMED, Z. M. & NAEEM, M. 2021. Identification and Computational Analysis of Novel Pathogenic Variants in Pakistani Families with Diverse Epidermolysis Bullosa Phenotypes. *Biomolecules,* 11.

KIRITSI, D., HAS, C. & BRUCKNER-TUDERMAN, L. 2013. Laminin 332 in junctional epidermolysis bullosa. *Cell Adh Migr,* 7**,** 135-41.

KIRITSI, D., HUILAJA, L., FRANZKE, C. W., KOKKONEN, N., PAZZAGLI, C., SCHWIEGER-BRIEL, A., LARMAS, M., BRUCKNER-TUDERMAN, L., HAS, C. & TASANEN, K. 2015. Junctional epidermolysis bullosa with LAMB3 splice-site mutations. *Acta Derm Venereol,* 95**,** 849-51.

KIRITSI, D., KERN, J. S., SCHUMANN, H., KOHLHASE, J., HAS, C. & BRUCKNER-TUDERMAN, L. 2011. Molecular mechanisms of phenotypic variability in junctional epidermolysis bullosa. *J Med Genet,* 48**,** 450-7.

KIVIRIKKO, S., MCGRATH, J. A., PULKKINEN, L., UITTO, J. & CHRISTIANO, A. M. 1996. Mutational hotspots in the LAMB3 gene in the lethal (Herlitz) type of junctional epidermolysis bullosa. *Hum Mol Genet,* 5**,** 231-7.

KOSTER, J., GEERTS, D., FAVRE, B., BORRADORI, L. & SONNENBERG, A. 2003. Analysis of the interactions between BP180, BP230, plectin and the integrin alpha6beta4 important for hemidesmosome assembly. *J Cell Sci,* 116**,** 387-99.

LANDRUM, M. J., LEE, J. M., BENSON, M., BROWN, G. R., CHAO, C., CHITIPIRALLA, S., GU, B., HART, J., HOFFMAN, D., JANG, W., KARAPETYAN, K., KATZ, K., LIU, C., MADDIPATLA, Z., MALHEIRO, A., MCDANIEL, K., OVETSKY, M., RILEY, G., ZHOU, G., HOLMES, J. B., KATTMAN, B. L. & MAGLOTT, D. R. 2018. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res,* 46**,** D1062-D1067.

LAROUSSI, N., MESSAOUD, O., CHARGUI, M., BEN FAYALA, C., ELAHLAFI, A., MOKNI, M., BASHAMBOO, A., MCELREAVEY, K., BOUBAKER, M. S., YACOUB YOUSSEF, H. & ABDELHAK, S. 2017. Identification of a Novel Mutation of LAMB3 Gene in a Lybian Patient with Hereditary Epidermolysis Bullosa by Whole Exome Sequencing. *Ann Dermatol,* 29**,** 243-246.

LIEHR, T. 2022. Uniparental disomy is a chromosomic disorder in the first place. *Mol Cytogenet,* 15**,** 5.

LINDEBOOM, R. G., SUPEK, F. & LEHNER, B. 2016. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet,* 48**,** 1112-8.

MATSUMURA, H., MOHRI, Y., BINH, N. T., MORINAGA, H., FUKUDA, M., ITO, M., KURATA, S., HOEIJMAKERS, J. & NISHIMURA, E. K. 2016. Hair follicle aging is driven by transepidermal elimination of stem cells via COL17A1 proteolysis. *Science,* 351**,** aad4395.

MCGRATH, J. A., ASHTON, G. H., MELLERIO, J. E., SALAS-ALANIS, J. C., SWENSSON, O., MCMILLAN, J. R. & EADY, R. A. 1999. Moderation of phenotypic severity in dystrophic and junctional forms of epidermolysis bullosa through in-frame skipping of exons containing non-sense or frameshift mutations. *J Invest Dermatol,* 113**,** 314-21.

MCLEAN, W. H., IRVINE, A. D., HAMILL, K. J., WHITTOCK, N. V., COLEMAN-CAMPBELL, C. M., MELLERIO, J. E., ASHTON, G. S., DOPPING-HEPENSTAL, P. J., EADY, R. A., JAMIL, T., PHILLIPS, R., SHABBIR, S. G., HAROON, T. S., KHURSHID, K., MOORE, J. E., PAGE, B., DARLING, J., ATHERTON, D. J., VAN STEENSEL, M. A., MUNRO, C. S., SMITH, F. J. & MCGRATH, J. A. 2003. An unusual N-terminal deletion of the laminin alpha3a isoform leads to the chronic granulation tissue disorder laryngo-onycho-cutaneous syndrome. *Hum Mol Genet,* 12**,** 2395-409.

MEYER, K. D. & JAFFREY, S. R. 2014. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat Rev Mol Cell Biol,* 15**,** 313-26.

MONTES, M., SANFORD, B. L., COMISKEY, D. F. & CHANDLER, D. S. 2019. RNA Splicing and Disease: Animal Models to Therapies. *Trends Genet,* 35**,** 68-87.

MOSS, C., WONG, A. & DAVIES, P. 2009. The Birmingham Epidermolysis Bullosa Severity score: development and validation. *Br J Dermatol,* 160**,** 1057-65.

NAKANO, A., CHAO, S. C., PULKKINEN, L., MURRELL, D., BRUCKNER-TUDERMAN, L., PFENDNER, E. & UITTO, J. 2002. Laminin 5 mutations in junctional epidermolysis bullosa: molecular basis of Herlitz vs. non-Herlitz phenotypes. *Hum Genet,* 110**,** 41-51.

NAKANO, A., PFENDNER, E., HASHIMOTO, I. & UITTO, J. 2000. Herlitz junctional epidermolysis bullosa: novel and recurrent mutations in the LAMB3 gene and the population carrier frequency. *J Invest Dermatol,* 115**,** 493-8.

NYSTROM, A., VELATI, D., MITTAPALLI, V. R., FRITSCH, A., KERN, J. S. & BRUCKNER-TUDERMAN, L. 2013. Collagen VII plays a dual role in wound healing. *J Clin Invest,* 123**,** 3498-509.

O'TOOLE, E. A., MARINKOVICH, M. P., HOEFFLER, W. K., FURTHMAYR, H. & WOODLEY, D. T. 1997. Laminin-5 inhibits human keratinocyte migration. *Exp Cell Res,* 233**,** 330-9.

ODORISIO, T., DI SALVIO, M., ORECCHIA, A., DI ZENZO, G., PICCINNI, E., CIANFARANI, F., TRAVAGLIONE, A., UVA, P., BELLEI, B., CONTI, A., ZAMBRUNO, G. & CASTIGLIA, D. 2014. Monozygotic twins discordant for recessive dystrophic epidermolysis bullosa phenotype highlight the role of TGF-β signalling in modifying disease severity. *Hum Mol Genet,* 23**,** 3907-22.

PACHO, F., ZAMBRUNO, G., CALABRESI, V., KIRITSI, D. & SCHNEIDER, H. 2011. Efficiency of translation termination in humans is highly dependent upon nucleotides in the neighbourhood of a (premature) termination codon. *J Med Genet,* 48**,** 640-4.

PASMOOIJ, A. M., PAS, H. H., BOLLING, M. C. & JONKMAN, M. F. 2007. Revertant mosaicism in junctional epidermolysis bullosa due to multiple correcting second-site mutations in LAMB3. *J Clin Invest,* 117**,** 1240-8.

PERTEA, M., LIN, X. & SALZBERG, S. L. 2001. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res,* 29**,** 1185-90.

PULKKINEN, L., CHRISTIANO, A. M., GERECKE, D., WAGMAN, D. W., BURGESON, R. E., PITTELKOW, M. R. & UITTO, J. 1994. A homozygous nonsense mutation in the beta 3 chain gene of laminin 5 (LAMB3) in Herlitz junctional epidermolysis bullosa. *Genomics,* 24**,** 357-60.

PULKKINEN, L., MENEGUZZI, G., MCGRATH, J. A., XU, Y., BLANCHET-BARDON, C., ORTONNE, J. P., CHRISTIANO, A. M. & UITTO, J. 1997. Predominance of the recurrent mutation R635X in the LAMB3 gene in European patients with Herlitz junctional epidermolysis bullosa has implications for mutation detection strategy. *J Invest Dermatol,* 109**,** 232-7.

REESE, M. G., EECKMAN, F. H., KULP, D. & HAUSSLER, D. 1997. Improved splice site detection in Genie. *J Comput Biol,* 4**,** 311-23.

REGEV, A., TEICHMANN, S. A., LANDER, E. S., AMIT, I., BENOIST, C., BIRNEY, E., BODENMILLER, B., CAMPBELL, P., CARNINCI, P., CLATWORTHY, M., CLEVERS, H., DEPLANCKE, B., DUNHAM, I., EBERWINE, J., EILS, R., ENARD, W., FARMER, A., FUGGER, L., GOTTGENS, B., HACOHEN, N., HANIFFA, M., HEMBERG, M., KIM, S., KLENERMAN, P., KRIEGSTEIN, A., LEIN, E., LINNARSSON, S., LUNDBERG, E., LUNDEBERG, J., MAJUMDER, P., MARIONI, J. C., MERAD, M., MHLANGA, M., NAWIJN, M., NETEA, M., NOLAN, G., PE'ER, D., PHILLIPAKIS, A., PONTING, C. P., QUAKE, S., REIK, W., ROZENBLATT-ROSEN, O., SANES, J., SATIJA, R., SCHUMACHER, T. N., SHALEK, A., SHAPIRO, E., SHARMA, P., SHIN, J. W., STEGLE, O., STRATTON, M., STUBBINGTON, M. J. T., THEIS, F. J., UHLEN, M., VAN OUDENAARDEN, A., WAGNER, A., WATT, F., WEISSMAN, J., WOLD, B., XAVIER, R., YOSEF, N. & HUMAN CELL ATLAS MEETING, P. 2017. The Human Cell Atlas. *Elife,* 6.

RINALDI, C. & WOOD, M. J. A. 2018. Antisense oligonucleotides: the next frontier for treatment of neurological disorders. *Nat Rev Neurol,* 14**,** 9-21.

ROSENBERG, A. B., PATWARDHAN, R. P., SHENDURE, J. & SEELIG, G. 2015. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell,* 163**,** 698-711.

ROWLANDS, C., THOMAS, H. B., LORD, J., WAI, H. A., ARNO, G., BEAMAN, G., SERGOUNIOTIS, P., GOMES-SILVA, B., CAMPBELL, C., GOSSAN, N., HARDCASTLE, C., WEBB, K., O'CALLAGHAN, C., HIRST, R. A., RAMSDEN, S., JONES, E., CLAYTON-SMITH, J., WEBSTER, A. R., GENOMICS ENGLAND RESEARCH, C., DOUGLAS, A. G. L., O'KEEFE, R. T., NEWMAN, W. G., BARALLE, D., BLACK, G. C. M. & ELLINGFORD, J. M. 2021. Comparison of in silico strategies to prioritize rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. *Sci Rep,* 11**,** 20607.

RUZZI, L., PAS, H., POSTERARO, P., MAZZANTI, C., DIDONA, B., OWARIBE, K., MENEGUZZI, G., ZAMBRUNO, G., CASTIGLIA, D. & D'ALESSIO, M. 2001. A homozygous nonsense mutation in type XVII collagen gene (COL17A1) uncovers an alternatively spliced mRNA accounting for an unusually mild form of non-Herlitz junctional epidermolysis bullosa. *J Invest Dermatol,* 116, 182-7.

SCHOFIELD, O. M., FINE, J. D., VERRANDO, P., HEAGERTY, A. H., ORTONNE, J. P. & EADY, R. A. 1990. GB3 monoclonal antibody for the diagnosis of junctional epidermolysis bullosa: results of a multicenter study. *J Am Acad Dermatol,* 23**,** 1078-83.

SHABBIR, G., HASSAN, M. & KAZMI, A. 1986. Laryngo-onycho-cutaneous syndrome: a study of 22 cases. *Biomedica,* 2**,** 15-25.

STENSON, P. D., MORT, M., BALL, E. V., CHAPMAN, M., EVANS, K., AZEVEDO, L., HAYDEN, M., HEYWOOD, S., MILLAR, D. S., PHILLIPS, A. D. & COOPER, D. N. 2020. The Human Gene Mutation Database (HGMD((R))): optimizing its use in a clinical diagnostic or research setting. *Hum Genet,* 139**,** 1197-1207.

SZAFRANSKI, K., SCHINDLER, S., TAUDIEN, S., HILLER, M., HUSE, K., JAHN, N., SCHREIBER, S., BACKOFEN, R. & PLATZER, M. 2007. Violating the splicing rules: TG dinucleotides function as alternative 3' splice sites in U2-dependent introns. *Genome Biol,* 8**,** R154.

TOYONAGA, E., NISHIE, W., KOMINE, M., MURATA, S., SHINKUMA, S., NATSUGA, K., NAKAMURA, H., OHTSUKI, M. & SHIMIZU, H. 2015. Skipped exon in COL7A1 determines the clinical phenotypes of dystrophic epidermolysis bullosa. *Br J Dermatol,* 172**,** 1141-4.

TURNER, T. N. 2015. Plot Protein: Visualization of Mutations. 3.0.0 ed. https://github.com/tycheleturner/plot-protein.

UITTO, J., BRUCKNER-TUDERMAN, L., MCGRATH, J. A., RIEDL, R. & ROBINSON, C. 2018. EB2017-Progress in Epidermolysis Bullosa Research toward Treatment and Cure. *J Invest Dermatol,* 138**,** 1010-1016.

VAILLY, J., PULKKINEN, L., CHRISTIANO, A. M., TRYGGVASON, K., UITTO, J., ORTONNE, J. P. & MENEGUZZI, G. 1995. Identification of a homozygous exon-skipping mutation in the LAMC2 gene in a patient with Herlitz's junctional epidermolysis bullosa. *J Invest Dermatol,* 104**,** 434-7.

VAN DEN AKKER, P. C., JONKMAN, M. F., RENGAW, T., BRUCKNER-TUDERMAN, L., HAS, C., BAUER, J. W., KLAUSEGGER, A., ZAMBRUNO, G., CASTIGLIA, D., MELLERIO, J. E., MCGRATH, J. A., VAN ESSEN, A. J., HOFSTRA, R. M. & SWERTZ, M. A. 2011. The international dystrophic epidermolysis bullosa patient registry: an online database of dystrophic epidermolysis bullosa patients and their COL7A1 mutations. *Hum Mutat,* 32**,** 1100-7.

VARKI, R., SADOWSKI, S., PFENDNER, E. & UITTO, J. 2006. Epidermolysis bullosa. I. Molecular genetics of the junctional and hemidesmosomal variants. *J Med Genet,* 43**,** 641-52.

WANG, M. & MARIN, A. 2006. Characterization and prediction of alternative splice sites. *Gene,* 366**,** 219-27.

WANG, Z. & BURGE, C. B. 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA,* 14**,** 802-13.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet,* 10**,** 57-63.

WATANABE, M., NATSUGA, K., NISHIE, W., KOBAYASHI, Y., DONATI, G., SUZUKI, S., FUJIMURA, Y., TSUKIYAMA, T., UJIIE, H., SHINKUMA, S., NAKAMURA, H., MURAKAMI, M., OZAKI, M., NAGAYAMA, M., WATT, F. M. & SHIMIZU, H. 2017. Type XVII collagen coordinates proliferation in the interfollicular epidermis. *Elife,* 6.

WEN, D., BALACCO, D. L., BARDHAN, A., HARPER, N., WALSH, D., RYAN, G., LIU, L., GUY, A., MCGRATH, J. A., OGBOLI, M. & HEAGERTY, A. H. M. 2022. Localized autosomal recessive epidermolysis bullosa simplex arising from a novel homozygous frameshift mutation in DST (BPAG1). *Clin Exp Dermatol,* 47**,** 497-502.

XIONG, H. Y., ALIPANAHI, B., LEE, L. J., BRETSCHNEIDER, H., MERICO, D., YUEN, R. K., HUA, Y., GUEROUSSOV, S., NAJAFABADI, H. S., HUGHES, T. R., MORRIS, Q., BARASH, Y., KRAINER, A. R., JOJIC, N., SCHERER, S. W., BLENCOWE, B. J. & FREY, B. J. 2015. RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science,* 347**,** 1254806.

YEO, G. & BURGE, C. B. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol,* 11**,** 377-94.

ZHU, J., MAYEDA, A. & KRAINER, A. R. 2001. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell,* 8**,** 1351-61.

APPENDIX

1. Mutations associated with JEB

Gene	Phenotype	Nonsense	Missense	Splicing	Small deletions	Small insertions	Small indels	Gross deletions	Gross insertions	Total
LAMB3 Total mutations: 123	Epidermolysis bullosa, junctional	12	12	18	14	6	1	1	2	66
	Epidermolysis bullosa, Herlitz	19	3	7	15	3	0	0	1	48
	Epidermolysis bullosa, junctional, intermediate	0	1	0	2	0	0	1	0	4
	Epidermolysis bullosa	0	0	2	1	0	0	0	0	3
	Epidermolysis bullosa, atrophic benign	0	0	0	1	0	0	0	0	1
	Epidermolysis bullosa, junctional non-Herlitz	0	0	0	1	0	0	0	0	1
LAMA3 Total mutations: 54	Epidermolysis bullosa, junctional	4	6	6	6	3	0	2	0	27
	Epidermolysis bullosa, Herlitz	9	2	3	8	0	0	0	0	22
	Laryngo-onycho-cutaneous syndrome	2	0	0	0	1	0	0	0	3
	Epidermolysis bullosa, generalised intermediate	1	0	0	0	0	0	0	0	1
	Epidermolysis bullosa, junctional non-Herlitz	0	0	0	1	0	0	0	0	1
LAMC2 Total mutations: 48	Epidermolysis bullosa, Herlitz	9	1	4	10	1	1	0	0	26
	Epidermolysis bullosa, junctional	6	0	2	4	1	2	0	1	16
	Epidermolysis bullosa	4	0	0	0	0	0	0	0	4
	Epidermolysis bullosa, junctional, intermediate	1	1	0	0	0	0	0	0	2
COL17A1 Total mutations: 114	Epidermolysis bullosa, junctional	13	7	9	20	11	2	0	1	63
	Epidermolysis bullosa	3	2	4	3	2	0	1	0	15
	Epidermolysis bullosa, atrophic benign	3	1	5	2	4	0	0	0	15
	Epidermolysis bullosa, junctional, localised	1	2	2	1	2	0	0	0	8
	Epidermolysis bullosa, junctional, generalised	1	0	1	2	3	0	0	0	7
	Epidermolysis bullosa, Herlitz	2	0	1	1	0	0	0	0	4
	Epidermolysis bullosa, junctional with prurigo-like lesions	0	0	1	0	0	0	0	0	1
	Epidermolysis bullosa, junctional, late-onset	0	0	0	1	0	0	0	0	1
Total		90	38	65	93	37	6	5	5	339

Table A1: Reported mutations associated with JEB in LAMB3, LAMA3, LAMC2 and COL17A1 (HGMD 2021.1). (HGMD 2021.1). EB phenotypes and mutation types are as labelled in HGMD. Mutation type statistics HGMD 23.4.22.xlsx
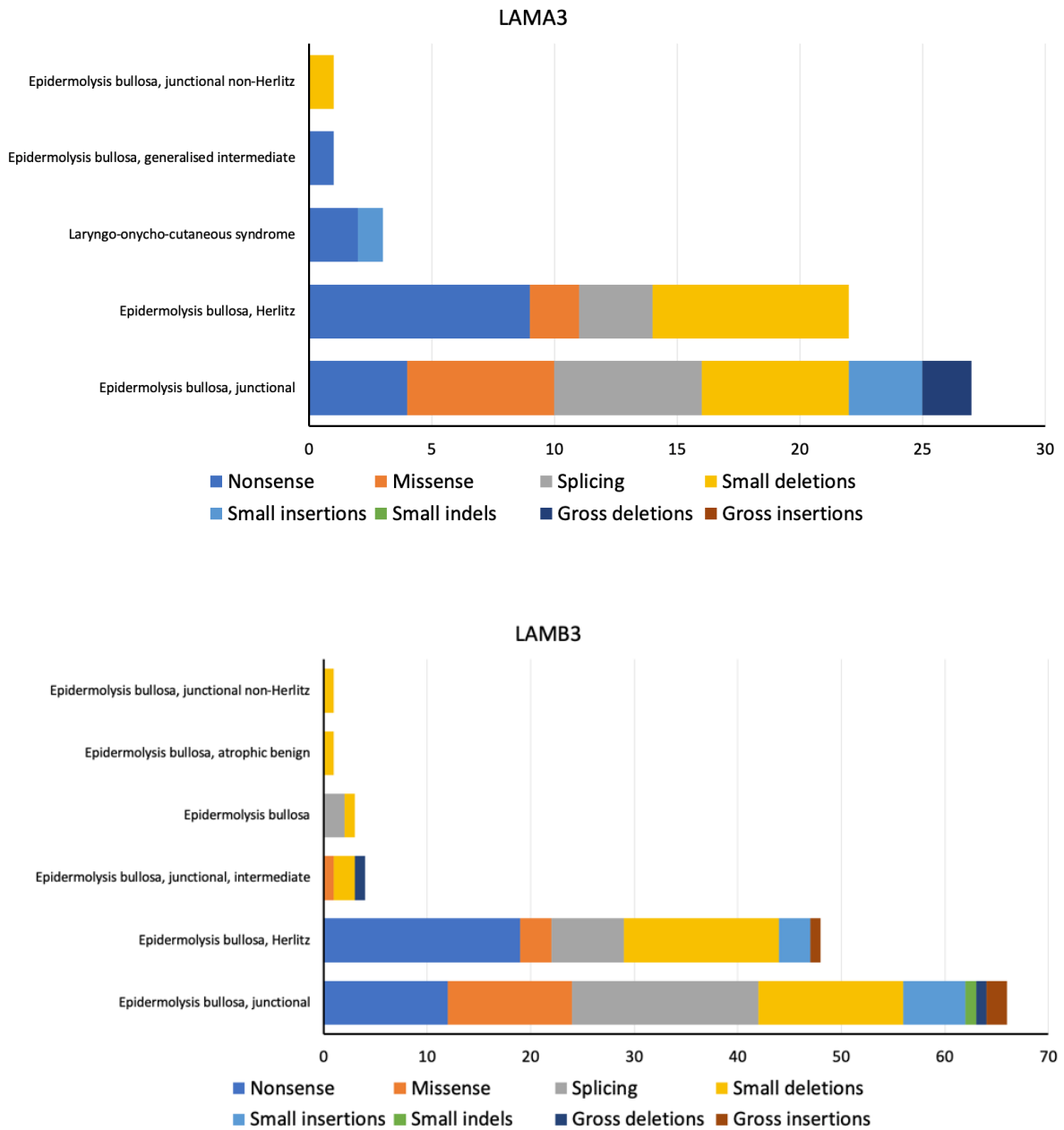
LAMA3

LAMB3

Figure A1: Frequency and type of mutations associated with JEB for LAMA3 and LAMB3. Mutation types and phenotypes are labelled as in HGMD (HGMD Professional 2021.1).
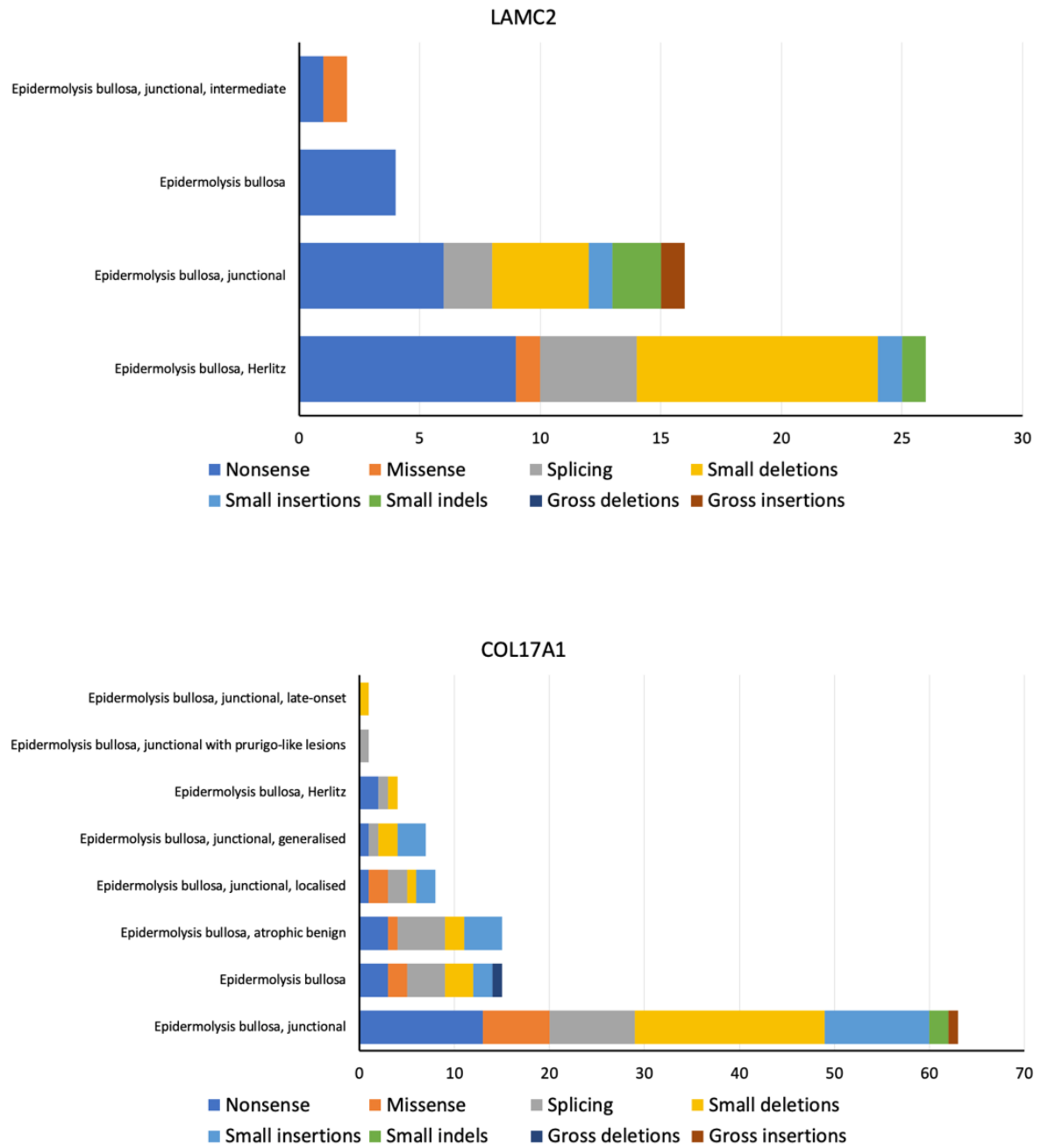
Figure A2: Frequency and type of mutations associated with JEB for LAMC2 and COL17A1. Mutation types and phenotypes are labelled as in HGMD (HGMD Professional 2021.1).

4

2.1 Patient Groups Flow Chart
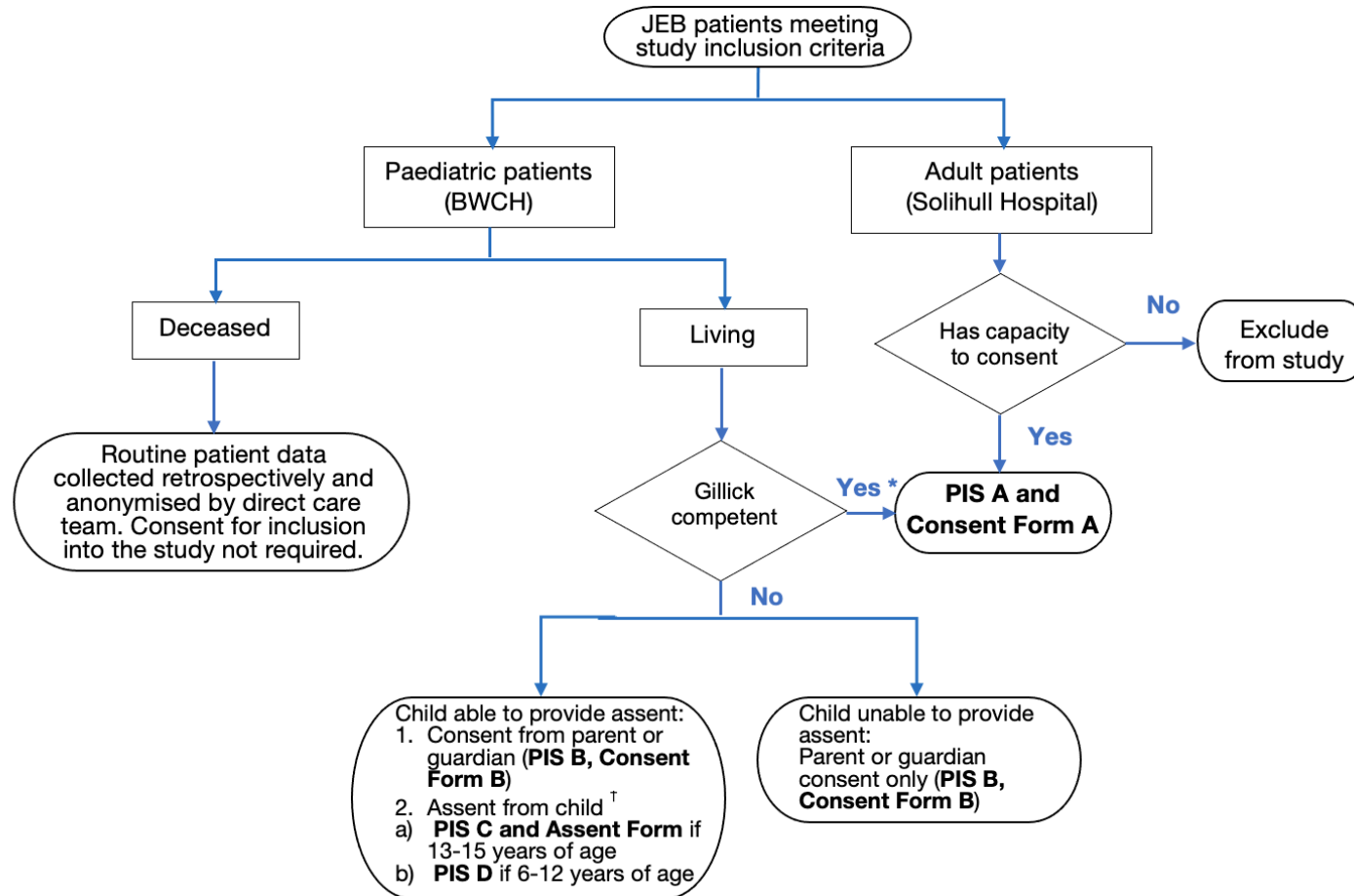
Figure A3: Flowchart to illustrate participant groups and which consent form and patient information sheet to use for each.

Abbreviations: JEB = junctional epidermolysis bullosa, PIS = patient information sheet, BWCH = Birmingham Womens and Children's Hospital, NOK = next of kin. * For Gillick competent children, attempts for parents or guardians to also consent for their child's inclusion into the study will be made. Parents to complete PIS B and Consent Form B. † Suggested ages of children for PIS C and PIS D are shown. The exact PIS to be used is at the clinician's discretion based on the child's understanding.

PIS A and Consent Form A: For adult JEB patients with capacity, or Gillick competent JEB patients. PIS B and Consent form B: For parents or guardians consenting on behalf of living children. PIS C or PIS D: For children lacking capacity, but able to provide assent. Assent forms are included in PIS C and PIS D. In addition, parents to complete PIS B and Consent Form B.

2.2 Patient Information Sheet A
(PIS A)

Version 1.10 – 18th November 2021

Study title: Genotype-phenotype correlation in junctional epidermolysis bullosa

Name of lead researcher: Professor Adrian Heagerty
IRAS ID: 290183
Patient information sheet version: 1.10
Release date: 18/11/2021

Introduction
We would like to invite you to participate in a study that is collecting genetic and clinical data about patients with junctional epidermolysis bullosa (EB) and laryngo-onycho-cutaneous (LOC) syndrome. Before you decide whether you would like to take part, we would like to outline why the research is being done and what it would involve for you. Please read this information carefully, and discuss it with others such as family, friends or your GP if you wish. Please contact the research team if you have any further questions after reading this information sheet.

What is the purpose of the study?
Epidermolysis bullosa is caused by alterations in one's DNA (our genes). This results in proteins found in the skin and other areas of the body not working as they should do, which gives rise to blistering and other conditions. We know that different types of DNA alterations can result in different characteristics and severity, but we still have much to learn about this relationship. The purpose of this project is to examine the relationship between the DNA alterations and severity of disease features.

This research will help in establishing the precise relationship between genetic defects and the clinical characteristics of EB patients. Ultimately, this can help us understand some of the processes that result in disease, and can also help us predict the likely course of the disease from genetic information. Computer modelling and analysis of genetic defects will also be carried out, to help us understand how genetic alterations affect protein structure and function.

What information will be collected?
As part of a study, a member of your direct care team will review your medical records for the information required for the study. Genetic information collected from participants will include which gene is affected, and the specific DNA alteration involved. This information will be obtained from your medical records by a member of the EB clinical team. No direct identifiers, such as name or date of birth will be available to anyone outside of your EB clinical team.

Clinical information will be collected from your medical records and following detailed examination if additional details are required. This will include characteristics such as how much skin is affected by blistering, whether there are any changes in the nails, teeth, eyes, hair or windpipe, and if individuals have developed scarring, skin cancer, or other life-threatening complications. Skin biopsy results, if available from previous reports, will also be collected. If completed previously, results from questionnaires about the impact that EB has had on your life will be collected.

Who can take part?
Patients with junctional epidermolysis bullosa from both genders and of any age can participate in this study. Data from patients with junctional epidermolysis bullosa who have recently passed away can also be used in this study.

Do I have to take part?
It is entirely your decision whether to participate or not. Should you meet all criteria for the study at this stage and are interested in participating, you will be invited to sign a consent form before taking part in the study. If you do not meet the criteria for the study, the assessor will inform you of your ineligibility, and will not perform any further research investigations on you for the purposes of this study. Any decision to decline from participating in the study will not in any way affect your future care within the health service. If you change your mind after agreeing to take part, you are free to withdraw from the study at any time without giving a reason. Withdrawal from the study will not in any way affect your future care within the health service.

What will my participation involve?
By consenting to the study, a member of your healthcare team will review your medical records for the results of any genetic tests, and also for clinical characteristics such as those mentioned above. If further clinical information is required regarding EB features, a doctor will undertake a detailed clinical examination at your next clinic appointment or home visit.

What will happen to the data collected?
All information collected about you will be processed in accordance with the Data Protection Act 2018 and General Data Protection Regulation (GDPR) and will be kept strictly confidential. Your medical records will only be available to your care team at your hospital. However, your medical records and data collected during this study may be looked at by the hospital's Research and Development Department and where necessary by regulatory authorities to check that the study is being performed correctly.

The raw data collected as part of this study will be documented in a paper form which will be kept in the hospital Dermatology department in a research file kept in a locked filing cabinet, within a secure office, along with signed consent forms. This data will also be saved on an electronic patient record accessible only by those directly involved in your care. This data will be kept for 10 years in keeping with the Sponsor's (University Hospitals Birmingham NHS Foundation Trust) policy.

All personal identifiable information will be removed after this point and you will be assigned a unique code number. Relevant data such as genetic information and clinical characteristics including age and gender will be extracted without identifiable information and shared with other researchers at the University of Birmingham, who will analyse the data.

What are the possible benefits of taking part?
It is unlikely that patients will benefit directly from taking part in this study. However, participation will allow the clinical and scientific community to have a better understanding of the disease mechanisms underlying junctional epidermolysis bullosa. In the long term, this could lead to the development of new treatments for EB in the future which could benefit you and other EB patients.

What are the possible disadvantages and risks of taking part?
If required, we expect clinical examination and the quality of life questionnaire to take 20-30 minutes to complete. However, we will try and arrange for this to be completed during a routine clinical appointment or home visit, so you won't have to come to hospital especially for this.

What are my choices about how my information is used?
We need to manage your records in specific ways for the research to be reliable. This means that we won't be able to let you see or change the data we hold about you. You can stop being a part of the study at any time without giving a reason, but we will keep information about you that we already have, unless you specifically request for it not to be used.

Where can I find more about how my information is used?
University Hospitals Birmingham NHS Foundation Trust is the UK Sponsor for the study, and the contact for any queries or concerns regarding the use of your information is: Professor Adrian Heagerty.
Further information on how your data may be used by us can be found on the following website: www.hra.nhs.uk/information-about-patients/   or you can ask a member of the research team.

University Hospitals Birmingham NHS Foundation Trust has produced a privacy notice that explains what we do with your personal information where we are or have provided care to you. It tells you:
- the information we collect about you
- how we store this information
- how long we retain it
- who we may share it with
- for which legal purpose we may share it

This information can be found at: https://www.uhb.nhs.uk/privacy-notice/research

What if I have concerns about the study?
If you have any concerns about the study you can speak to a member of the research team whose contact details are provided at the end of this information sheet. If you are not satisfied

with their answers or have any further concerns you can contact the local Patients Advisory and Liaison Service (PALS).
Solihull Hospital (University Hospitals Birmingham NHS Foundation Trust): 0121 424 0808
Birmingham Children's Hospital (Birmingham Women's and Children's NHS Foundation Trust): 0121 333 8403

If you wish to complain about how our data has been handled, you may do so by contacting the Data Protection Officer.
Solihull Hospital (University Hospitals Birmingham NHS Foundation Trust): informationgovernance@uhb.nhs.uk.
Birmingham Children's Hospital (Birmingham Women's and Children's NHS Foundation Trust): Data Protection Officer, Birmingham Children's Hospital, Steelhouse Lane, Birmingham B4 6NH

If you are not happy with their response or believe that your data is being processed in a way that is not lawful, you can complain to the Information Commissioner's Office (ICO) via their website on www.ico.org.uk or by telephone: 0303 123 1113.

What will happen to the results of the research?
The results of this study may be published in peer review scientific journals, postgraduate degree theses, and may also be presented at scientific national and international conferences; we will publish these results in a way that no one will be able to identify you. Participants wishing to have access to the results prior to publication can contact the researchers.

What if something goes wrong?
We do not expect anything to go wrong during the study. However, if you have any concerns about the way you have been approached or treated during this study, please contact:
Dr David Wen or Professor Adrian Heagerty at: eb.team@nhs.net.

Who is funding and organizing the research?
This study is funded by the Dystrophic Epidermolysis Research Association (DeBRA-UK) and is organised by the Department of Dermatology at Solihull Hospital (a site of University Hospitals Birmingham NHS Foundation Trust).

Who has reviewed the study?
This study has been reviewed by a Research Ethics Committee.

Contact and further information
If you want any further information about this particular study, you can contact the lead researcher:

Prof. Adrian Heagerty, BSc(Hons) MB BS MD FRCP
Consultant Dermatologist and Honorary Chair in Dermatology at University of Birmingham
Department of Dermatology, Solihull Hospital, University Hospitals Birmingham

Thank you for taking the time to read this information.

2.3 Parent/Guardian Information Sheet B
(PIS B)

Version 1.10 – 18th November 2021

Study title: Genotype-phenotype correlation in junctional epidermolysis bullosa

Name of lead researcher: Professor Adrian Heagerty
IRAS ID: 290183
Patient information sheet version: 1.10
Release date: 18/11/2021

Introduction
We would like to ask for your permission to include data about your child in a study that is collecting genetic and clinical information about patients with junctional epidermolysis bullosa (EB) and laryngo-onycho-cutaneous (LOC) syndrome. Before you decide whether you would like to provide consent, we would like to outline why the research is being done and what it would involve for you and your child. Please read this information carefully, and discuss it with others such as family, friends or your GP if you wish. Please contact the research team if you have any further questions after reading this information sheet.

What is the purpose of the study?
Epidermolysis bullosa is caused by alterations in one's DNA (our genes). This results in proteins found in the skin and other areas of the body not working as they should do, which gives rise to blistering and other conditions. We know that different types of DNA alterations can result in different characteristics and severity, but we still have much to learn about this relationship. The purpose of this project is to examine the relationship between the DNA alterations and severity of disease features.

This research will help in establishing the precise relationship between genetic defects and the clinical characteristics of EB patients. Ultimately, this can help us understand some of the processes that result in disease, and can also help us predict the likely course of the disease from genetic information. Computer modelling and analysis of genetic defects will also be carried out, to help us understand how genetic alterations affect protein structure and function.

What information will be collected?
Information will be collected from reviewing participants' medical records and results of any investigations. This will be completed by either a member of the EB clinical team at Birmingham Children's Hospital. Genetic information collected will include which gene is affected, and the specific DNA alteration involved. No direct patient identifiers such as name or date of birth will be available to anyone outside of your EB clinical team.

Clinical information will be collected from medical records and following detailed examination by an EB professional usually involved in your child's care if additional details are required. This will include characteristics such as how much skin is affected by blistering, whether there are any changes in the nails, teeth, eyes, hair or windpipe, and if individuals have developed scarring, skin cancer, or other life-threatening complications. Skin biopsy results, if available from previous reports, will also be collected. If completed previously, results from questionnaires about the impact that EB has had on your child's life will be collected.

Who can take part?
Patients with junctional epidermolysis bullosa from both genders and of any age can participate in this study. Data from patients with junctional epidermolysis bullosa who have recently passed away can also be used in this study.

Does my child have to take part?
It is entirely you and (if they are old enough to understand) your child's decision whether they participate or not. Should your child meet all criteria for the study at this stage, you will be invited to sign a consent form on behalf of your child before being included in the study. If you change your mind after agreeing for your child's details to be used, you are free to withdraw consent from the study at any time without giving a reason. Any decision to withdraw will not in any way affect any future care within the health service.

What will participation involve?
By consenting to the study, your child's medical records will be reviewed for the results of any genetic tests, and also for clinical characteristics such as those mentioned above. This will be completed by the EB team at Birmingham Children's Hospital. If further clinical information is required regarding EB features, a doctor will undertake a detailed clinical examination at your child's next clinic appointment or home visit.

What will happen to the data collected?
All information collected will be processed in accordance with the Data Protection Act 2018 and General Data Protection Regulation (GDPR) and will be kept strictly confidential. Your child's medical records will only be available to their care team. However, medical records and data collected during this study may be looked at by the hospital's Research and Development Department and where necessary by regulatory authorities to check that the study is being performed correctly.

The raw data collected as part of this study will be documented in a paper form which will be kept in the hospital Dermatology department in a research file kept in a locked filing cabinet, within a secure office, along with signed consent forms. This data will also be saved on an electronic patient record accessible only by those directly involved in your child's care. This data will be kept for 10 years in keeping with the Sponsor's (University Hospitals Birmingham NHS Foundation Trust) policy.

All personal identifiable information will be removed after this point and your child will be assigned a unique code number. Relevant data such as genetic information and clinical

characteristics including age and gender will be extracted without identifiable information and shared with other researchers at the University of Birmingham, who will analyse the data.

What are the possible benefits of taking part?
It is unlikely that participants will benefit directly from taking part in this study. However, participation will allow the clinical and scientific community to have a better understanding of the disease mechanisms underlying junctional epidermolysis bullosa. In the long term, this could lead to the development of new treatments for EB in the future which could benefit your child and other EB patients.

What are the possible disadvantages and risks of taking part?
If required, we expect clinical examination and the quality of life questionnaire to take 20-30 minutes to complete. However, we will try and arrange for this to be completed during a routine clinical appointment or home visit, so your child won't have to come to hospital especially for this.

What are my choices about how my child's information is used?
We need to manage the records in specific ways for the research to be reliable. This means that we won't be able to let you see or change the data we hold about your child. You can withdraw consent from the study at any time without giving a reason, but we will keep information about your child that we already have, unless you specifically request for it not to be used.

Where can I find more about how my child's information is used?
University Hospitals Birmingham NHS Foundation Trust is the UK Sponsor for the study, and the contact for any queries or concerns regarding the use of your information is: Professor Adrian Heagerty.
Further information on how your data may be used by us can be found on the following website: www.hra.nhs.uk/information-about-patients/ or you can ask a member of the research team.

University Hospitals Birmingham NHS Foundation Trust has produced a privacy notice that explains what we do with your personal information where we are or have provided care to you. It tells you:
- the information we collect about you
- how we store this information
- how long we retain it
- who we may share it with
- for which legal purpose we may share it

This information can be found at: https://www.uhb.nhs.uk/privacy-notice/research

What if I have concerns about the study?
If you have any concerns about the study you can speak to a member of the research team whose contact details are provided at the end of this information sheet. If you are not satisfied with their answers or have any further concerns you can contact the local Patients Advisory and Liaison Service (PALS).

Solihull Hospital (University Hospitals Birmingham NHS Foundation Trust): 0121 424 0808
Birmingham Children's Hospital (Birmingham Women's and Children's NHS Foundation Trust): 0121 333 8403

If you wish to complain about how our data has been handled, you may do so by contacting the Data Protection Officer.
Solihull Hospital (University Hospitals Birmingham NHS Foundation Trust): informationgovernance@uhb.nhs.uk.
Birmingham Children's Hospital (Birmingham Women's and Children's NHS Foundation Trust): Data Protection Officer, Birmingham Children's Hospital, Steelhouse Lane, Birmingham B4 6NH

If you are not happy with their response or believe that your data is being processed in a way that is not lawful, you can complain to the Information Commissioner's Office (ICO) via their website on www.ico.org.uk or by telephone: 0303 123 1113.

What will happen to the results of the research?
The results of this study may be published in peer review scientific journals, postgraduate degree theses, and may be presented at scientific national and international conferences; we will publish these results in a way that no one will be able to identify participants. Parents of participants wishing to have access to the results prior to publication can contact the researchers.

What if something goes wrong?
We do not expect anything to go wrong during the study. However, if you have any concerns about the way you have been approached or treated during this study, please contact:
Dr David Wen or Professor Adrian Heagerty at: eb.team@nhs.net.

Who is funding and organizing the research?
This study is funded by the Dystrophic Epidermolysis Research Association (DeBRA-UK) and is organised by the Department of Dermatology at Solihull Hospital (a site of University Hospitals Birmingham NHS Foundation Trust).

Who has reviewed the study?
This study has been reviewed by a Research Ethics Committee.

Contact and further information
If you want any further information about this particular study, you can contact the lead researcher:
Prof. Adrian Heagerty, BSc(Hons) MB BS MD FRCP
Consultant Dermatologist and Honorary Chair in Dermatology at University of Birmingham
Department of Dermatology, Solihull Hospital, University Hospitals Birmingham

Thank you for taking the time to read this information.

2.4 Patient Information Sheet C (PIS C)
For children aged 13 – 15

Version 1.6 – 18th November 2021

Study title: Genotype-phenotype correlation in junctional epidermolysis bullosa

Name of lead researcher: Professor Adrian Heagerty
IRAS ID: 290183
Patient information sheet version: 1.6
Release date: 18/11/2021

Introduction

- We are completing a research project to learn more about children and adults with junctional epidermolysis bullosa (EB) and would like to invite you to take part.

- Before you decide if you want to take part, we would like to explain why the research is being done and what it would involve for you.

- Please read this information carefully, and discuss it with others such as family, friends or your GP if you wish.

- Please contact the research team if you have any further questions after reading this information sheet.

What is the study about?

- EB is caused by problems with some of the building blocks (proteins) that make up our skin.

- These building blocks are damaged because of changes (mutations) in the instructions to make them (the DNA).

- This can lead to blisters and problems with other parts of the body such as hair, nails, eyes and teeth.

- Different people can have different changes (mutations) to their DNA, and this can lead to different types of EB.

- We would like to look at some of your blood test results and look at what changes have occurred in your DNA.

- We would also like to examine you and ask you some questions about how EB has affected your life and what you are able to do.

- Your blood test results and DNA changes will also be analysed by computer programs to investigate how the proteins in your skin are affected. No extra blood tests will need to be taken for this.

- This can help us understand how and why people with EB are affected differently by the disease.

Why have I been chosen?

- You have been diagnosed with junctional EB. We would like to collect information about all children from our clinic with this condition.

Do I have to take part?

- It is completely up to you and your parents to decide if you want to take part in this project or not.

- If you would like to take part, you and your parents will have to fill in a form (a consent form) to confirm this. If you change your mind, you can withdraw from the study at any time.

- If you decide not to take part, your usual medical care will carry on as usual. Your medical care in the future won't be affected by your decision.

What will happen if I choose to take part?

- A doctor or nurse from the EB team will look through your medical records to find out what change to your DNA has occurred and record this. They will also look through your medical records to review what features your EB has, such as how much skin is involved, and whether there are any changes in your nails, teeth, eyes, hair or windpipe.

- If further information is required, EB doctor or nurse will examine you at your next appointment (either in clinic or at home).

- No additional blood tests will be needed as part of this research project.

What will happen to information about me?

- Your medical records may be accessed by your EB care team as part of the project.

- Staff from the study organizer (University Hospitals Birmingham) may also access your medical records to check that the study is being done correctly.

- Information collected about you will be saved on a password protected hospital server and can only be accessed by the staff members mentioned above. Paper forms will be kept in a locked cabinet in a locked hospital office which requires swipe card access.

- Computer analysis of your DNA will be performed by researchers at the University of Birmingham. When information is sent to them, details which can identify you such as your name and birthday will be removed.

- The results of this project may be presented at scientific conferences and published in medical journals. Your name, and anything that could identify you, such as your birthday, won't be included.

Will taking part benefit me?

- Taking part in this project probably won't help you directly right now. But the information collected will help EB doctors and scientists understand the disease processes in EB better.

- In the long term, this could lead to the development of new treatments for EB in the future which could benefit you and other EB patients.

2.5 EB Study Project Information Sheet

- PIS D: For children aged 6 - 12
- Version 1.2 - 2/6/2021

- This project is being completed to learn more about a skin condition called EB.

- We would like to invite you to take part in this project because you have EB.

- We would like to put some information about your skin condition on a computer.

- This information will be analysed by scientists.

- This could lead to more treatments for EB in the future.

- You won't have to come to hospital any more than usual and you won't have any extra tests or medicine.

- If you don't want to be in the project then you don't have to.

- We will ask for permission from Mum and Dad to include you in this project.

- The doctors and nurses are here to answer any questions you have.

- Please talk to Mum and Dad about if you want to be in this project and circle your choice below:

✔ I want to be in the project

✖ I don't want to be in the project

- My name is:_____ The date is:_____

3.1 Sanger sequencing protocol

1. Add 20 - 50ng of PCR products or 100ng of cloned PCR products into each tube/well

2. Add 1µl of corresponds primer (**from one direction only**)

3. Add 1µl of BigDye v3.1 mixture

4. Add 1µl of 5X sequencing buffer

5. Add dH2O to bring up the total volume of 10µl

6. Mix well

7. Place the tube or plate into Centrifuge set for a short spin (in order to collect all the solution into the bottom of the tube/well)

8. Place the tube or PCR plate into the ABI GeneAmp 9700 PCR System and start reaction with the following cycle program:

        96°C for 10s
        50°C for 5s
        60°C for 45s
        Repeat this cycle for 25 times.

9. add 2 volumes of solution P (see Appendix 1) into each sequence reaction product, mix well by pipetting it up and down if using a 96 well plate, or vortex if using a single tube;

10. Add 2.2 volumes (of above) of 100% ethanol into each sample and mix well.

11. For sequence reaction products, incubate the mixture at room temperature for 15 minutes.

12. Centrifuge 20 minutes at 3700rpm if using Eppendorf benchtop Centrifuge 5804 or 13000rpm if using Eppendorf microcentrifuge 5424R.

13. Discarded the supernatant.

14. Add 200ul of 70% ethanol.

15. Centrifuge for 12 minutes at using Eppendorf benchtop Centrifuge 5804 or 13000rpm if using Eppendorf microcentrifuge 5424R.

16. Discarded the supernatant

17. Add 20µl of HiDi into each reaction products, mix well and load it onto the Sequencer.

3.2 Immunofluorescence mapping protocol

Sample preparation before immunolabelling.

1) Wash sample with PBS by rotating it in cold PBS for 1 hour;

2) On a small piece of cork tile, carefully orientate the sample with semi-frozen OCT medium to ensure the surface of the sample contains all layers of the skin;

3) Carefully transfer the cork with the skin sample into a pre-colled n-hexane to solidify the medium completely.

4) Transfer the frozen sample into cryotube that has been pre-cooled in LN_2.

5) Cryosection the sample at 4um thickness, leave the section dry at RT for few minutes, continue with the immunolabelling.

Immunolabelling (All antibodies are prepared in PBS/BSA)

1) Wash away OCT medium by soaking the slides in PBS for 5 minutes at RT;

2) Pre-block sample with 33% normal goat serum in 1% BSA made in PBS for 2 minutes at RT;

3) Add primary antibody onto each section;

4) Add PBS/BSA to one section as negative control;

5) Incubate the sample with antibodies at 37C incubator for 1 hour;

6) Wash the slides with PBS at RT for 10 minutes, with 2 changes in between;

7) Add secondary antibody onto each section, including the section for negative control;

8) Incubator at 37C for 1 hour;

9) Wash the slide with PBS at RT for 10 minutes, with 2 changes in between;

10) Drain, tap and blot the slides, and add mounting medium.

4. JEB phenotyping tool

Study ID: RRK7326   Date: Subject ID:

Area of body surface area (BSA) affected (%):

0	0%	6	51-60%
1	1-10%	7	61-70%
2	11-20%	8	71-80%
3	21-30%	9	81-90%
4	31-40%	10	91-100%
5	41-50%		

Chronic wounds (wounds present for >6 mo) (0-5)

0   None
1   1-10%
2   11-20%
3   21-29%
4   31-40%
5   41% or higher

Eyes (0-5)
Presence, extent and severity of:
☐ Conjunctival erythema and suffusion
☐ Corneal scarring (with end result = blindness in one or both eyes)
☐ Trichiasis and mechanical keratitis secondary to entropion formation
☐ Dryness secondary to ectropion formation
☐ Damage to the tear ducts

Larynx (0-5)
0   No problems from EB
1   Occasional hoarseness
2   Frequent hoarseness
3   Persistent hoarseness
4   Stridor
5   Laryngeal obstruction

Nails (0-5)
For score: Total lost nails ÷ 4 + total dystrophic nails ÷ 8

	Left hand	Right hand	Left foot	Right foot
Lost nails				
Dysmorphic				
Normal				

Scarring alopecia due to EB (0-5)
0   No alopecia
1   1-19% scalp involvement
2   20-39% scalp involvement
3   40-59% scalp involvement
4   60-79% scalp involvement
5   80-100% scalp involvement

Nutritional compromise (0-3)
• Height:_____
• Body weight:_____

• BMI:_____
☐ Underweight
☐ Hypoalbuminaemia
☐ Anaemia

Mouth (0-1)
☐ Soft tissue blisters and erosions

Teeth (0-5)
0   No problems from EB (normal teeth)
1   Localised or Generalised pitting
2   Enamel completely absent
3   Severe tooth wear and/or caries damage
4   Some teeth lost
5   All teeth lost

Granulation tissue (0-5)
0   None
1   <1% Body surface area
2   1-2% BSA
3   2-5% BSA
4   5-10% BSA
5   >10% BSA

Life threatening illness (0-10):
☐ Sepsis
☐ Renal failure
☐ Heart failure
☐ SCC
Total number of SCCs (up to 5):
SCC spread:
☐ No extracutaneous spread (0)
☐ Local/regional/LN spread (1)
☐ Distant metastatic spread (2)
• Other (freetext):

Other items for consideration (0-6):
☐ GU tract abnormalities
☐ Milia
☐ Hypertrophic or atrophic scarring
☐ Keratoderma
☐ Hyperhidrosis
☐ Absent dermatoglyphs (LOC)

Supplementary notes:
BSA affected includes:
Blisters, erosions, scabs, healing skin, erythema, atrophic scarring
Does not include:
Post-inflammatory pigmentation

5. Additional genotyping information

Case	Zygosity	Gene	Mutation	Genes sequenced	Parental mutations
1	Homozygous	*LAMA3*	c.2038_2039dupAA	PCR and bidirectional Sanger sequencing of the LAMB3, LAMC2 and LAMA3 genes	Mutation present in mother and father who were heterozygous carriers
2	Homozygous	*LAMA3*	c.4338delG	PCR and bidirectional Sanger sequencing of the LAMB3, LAMC2 and LAMA3 genes	Mutation present in mother and father who were heterozygous carriers
3	Homozygous	*LAMA3*	c.151insG	PCR and bidirectional sequencing of the LAMA3 gene due to clinical suspicion of LOC syndrome	Not available
4	Heterozygous	*LAMB3*	c.727C>T, c.1903C>T	PCR and bidirectional Sanger sequencing of the LAMB3 gene	Mutation present in mother and father who were heterozygous carriers of each mutation
5	Homozygous	*LAMB3*	c.1702C>T	PCR and bidirectional Sanger sequencing of the LAMB3 gene	Not available
6	Homozygous	*LAMB3*	c.1186_1196del	Not available	Not available
7	Homozygous	*LAMB3*	c.2701+1G>A	PCR and bidirectional Sanger sequencing of the LAMB3, LAMC2 and LAMA3 genes	Mutation present in mother and father who were heterozygous carriers
8	Heterozygous	*LAMB3*	c.1705C>T, c.943+2T>C	PCR and direct sequencing of the COL17A1 and LAMB3 genes	Not available
9	Homozygous	*LAMB3*	c.298+5G>C	PCR and bidirectional Sanger sequencing of the LAMB3 gene	Mutation present in mother and father who were heterozygous carriers
10	Heterozygous	*LAMB3*	c.565-2A>G, c.2914C>T	PCR and bidirectional Sanger sequencing of the LAMB3 and LAMC2 genes	Mutation present in mother and father who were heterozygous carriers of each mutation
11	Heterozygous	*LAMB3*	c.3119G>A, c.629-12T>A	PCR and bidirectional Sanger sequencing of the LAMB3 gene	Not available
12	Homozygous	*LAMC2*	c.132_135delCAGA	PCR and bidirectional Sanger sequencing of exon 2 of the LAMC2 gene	Not available
13	Homozygous	*LAMC2*	c.132_135delCAGA	PCR and bidirectional Sanger sequencing of the LAMB3, LAMC2 and LAMA3 genes	Mutation present in mother and father who were heterozygous carriers
14	Homozygous	*COL17A1*	c.2910insT	PCR and bidirectional Sanger sequencing of the COL17A1 gene	Mutation present in mother and father who were heterozygous carriers
15	Homozygous	*COL17A1*	c.2910insT	PCR and bidirectional Sanger sequencing of exon 44 of COL17A1 gene	Sibling
16	Homozygous	*COL17A1*	Large deletion of exons 16 and 17	PCR of exons 14 to 19 of the COL17A1 gene (forward primer exon 14 and reverse primer of exon 19)	Mutation present in mother and father who were heterozygous carriers
17	Homozygous	*COL17A1*	Large deletion of exons 16 and 17	Initial PCR and bidirectional Sanger sequencing did not detect any mutations in COL17A1. WES was performed which showed a lack of reads covering exons 16 and 17 of the COL17A1 gene. Subsequent PCR and Sanger sequencing confirmed a homozygous intragenic deletion.	Not mentioned

Table A2: Additional genotyping information outlining which genes were sequenced and whether paternal mutations were available.
Individuals 12 and 15 had siblings who had a genetic diagnosis of JEB already and so targeted sequencing was performed to confirm the presence of mutations. The NDEB laboratory comment that it is unclear why the original sample did not contain the mutation but nevertheless the repeat sample testing has provided a definitive answer.

6. Predicted domains by InterPro

Architecture name	InterPro reference	Pfam reference	Start site	End site	Significance
LAMB3					
Laminin_N-terminal(Domain VI)	IPR008211	PF00055	26	248	6.30E-52
Laminin_EGF_domain	IPR002049	PF00053	250	305	3.10E-06
Laminin_EGF_domain	IPR002049	PF00053	316	367	6.20E-07
Laminin_EGF_domain	IPR002049	PF00053	379	428	5.40E-12
Laminin_EGF_domain	IPR002049	PF00053	431	478	1.80E-10
Laminin EGF domain	IPR002049	PF00053	481	526	1.20E-05
Laminin_EGF_domain	IPR002049	PF00053	534	574	8.70E-07
LAMA3A					
Laminin EGF domain	IPR002049	PF00053	78	122	1.30E-07
Laminin EGF domain	IPR002049	PF00053	125	175	3.60E-07
Laminin Domain I	IPR009254	PF06008	238	496	3.00E-92
Laminin Domain II	IPR010307	PF06009	679	807	9.50E-39
Laminin G domain	IPR001791	PF00054	826	961	1.70E-05
Laminin G domain	IPR001791	PF02210	1018	1133	3.60E-14
Laminin G domain	IPR001791	PF02210	1188	1292	3.60E-07
Laminin G domain	IPR001791	PF02210	1408	1526	1.20E-20
Laminin G domain	IPR001791	PF02210	1577	1701	9.80E-22
LAMA3B					
Laminin N-terminal (Domain VI)	IPR008211	PF00055	48	297	5.60E-75
Laminin EGF domain	IPR002049	PF00053	356	412	1.20E-05
Laminin EGF domain	IPR002049	PF00053	426	465	1.00E-06
Laminin EGF domain	IPR002049	PF00053	491	531	2.30E-08
Laminin EGF domain	IPR002049	PF00053	536	581	2.00E-09
Laminin EGF domain	IPR002049	PF00053	631	681	3.20E-07
Laminin EGF domain	IPR002049	PF00053	684	722	7.50E-05
Laminin EGF domain	IPR002049	PF00053	1266	1314	4.60E-10
Laminin EGF domain	IPR002049	PF00053	1356	1401	1.30E-09
Laminin EGF domain	IPR002049	PF00053	1405	1453	2.30E-08
Laminin B (Domain IV)	IPR000034	PF00052	1518	1652	1.40E-30
Laminin EGF domain	IPR002049	PF00053	1654	1677	0.014
Laminin EGF domain	IPR002049	PF00053	1687	1731	2.80E-07
Laminin EGF domain	IPR002049	PF00053	1734	1784	7.40E-07
Laminin Domain I	IPR009254	PF06008	1847	2105	7.20E-92
Laminin Domain II	IPR010307	PF06009	2288	2416	2.10E-38
Laminin G domain	IPR001791	PF00054	2435	2570	3.80E-05
Laminin G domain	IPR001791	PF02210	2627	2742	7.80E-14
Laminin G domain	IPR001791	PF02210	2797	2901	7.70E-07
Laminin G domain	IPR001791	PF02210	3017	3135	2.60E-20
Laminin G domain	IPR001791	PF02210	3186	3310	2.10E-21
LAMC2					
Laminin EGF domain	IPR002049	PF00053	28	81	8.00E-05
Laminin EGF domain	IPR002049	PF00053	84	128	7.00E-10
Laminin EGF domain	IPR002049	PF00053	139	184	2.80E-08
Laminin B (Domain IV)	IPR000034	PF00052	250	380	8.50E-28
Laminin EGF domain	IPR002049	PF00053	381	405	0.048
Laminin EGF domain	IPR002049	PF00053	462	504	2.70E-04
Laminin EGF domain	IPR002049	PF00053	517	570	1.20E-09
Laminin EGF domain	IPR002049	PF00053	573	604	0.0049
COL17A1					
Collagen triple helix repeat (20 copies)	IPR008160	PF01391	567	624	3.20E-08
Collagen triple helix repeat (20 copies)	IPR008160	PF01391	749	807	1.80E-09
Collagen triple helix repeat (20 copies)	IPR008160	PF01391	1438	1482	2.00E-07

Table A3: Predicted domains by InterPro. Significance scores greater than E-05 are highlighted in red and excluded from schematics.

7. MasonMD results

```
> Mutation 1 (LAMA3 c.2038_2039dupAA)
> classify.nmd(gene_id = 3909, ref = 37, mut_start = 21487749, mut_end
= 21487750,
+ ref_nt = "AA", mut_nt = "AAAA")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "6997"           "TRUE"        "10004"
last_exon_exon_junction          n.exon
        "9855"           "75"
>
> Mutation 2 (LAMA3 c.4338delG)
> classify.nmd(gene_id = 3909, ref = 37, mut_start = 21523890, mut_end
= 21523890,
+ ref_nt = "G", mut_nt = "")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "9259"           "TRUE"        "10001"
last_exon_exon_junction          n.exon
        "9856"           "75"
>
> Mutation 3 (LAMA3 c.151insG)
> classify.nmd(gene_id = 3909, ref = 37, mut_start = 21453159, mut_end
= 21453159,
+ ref_nt = "G", mut_nt = "GG")
        mut_nmd          note          wt_nmd
        NA   "mutation in intron "     NA
        PTC              have.ptc      mutseq_length
        NA               "FALSE"       NA
last_exon_exon_junction          n.exon
        NA               "75"
>
> Mutation 4a (LAMB3 c.727C>T)
> classify.nmd(gene_id = 3914, ref = 37, mut_start = 209806023,
mut_end = 209806023,
+ ref_nt = "G", mut_nt = "A")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "727"            "TRUE"        "3519"
last_exon_exon_junction          n.exon
        "3382"           "22"
>
> Mutation 4b (LAMB3 c.1903C>T)
> classify.nmd(gene_id = 3914, ref = 37, mut_start = 209799066,
mut_end = 209799066,
+ ref_nt = "G", mut_nt = "A")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "1903"           "TRUE"        "3519"
last_exon_exon_junction          n.exon
        "3382"           "22"
>
> Mutation 5 (LAMB3 c.1702C>T)
> classify.nmd(gene_id = 3914, ref = 37, mut_start = 209799267,
mut_end = 209799267,
+ ref_nt = "G", mut_nt = "A")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "1702"           "TRUE"        "3519"
last_exon_exon_junction          n.exon
        "3382"           "22"
>
> Mutation 6 (LAMB3 c.1186_1196delACCGGGCAGTG)
> classify.nmd(gene_id = 3914, ref = 37, mut_start = 209801472,
mut_end = 209801482,
+ ref_nt = "CACTGCCCGGT", mut_nt = "")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "1219"           "TRUE"        "3508"
```

```
last_exon_exon_junction          n.exon
        "3372"           "22"
>
> Mutation 8a (LAMB3 c.1705C>T)
> classify.nmd(gene_id = 3914, ref = 37, mut_start = 209799264,
mut_end = 209799264,
+ ref_nt = "G", mut_nt = "A")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "1705"           "TRUE"        "3519"
last_exon_exon_junction          n.exon
        "3382"           "22"
>
> Mutation 10a (LAMB3 c.2914C>T)
> classify.nmd(gene_id = 3914, ref = 37, mut_start = 209791389,
mut_end = 209791389,
+ ref_nt = "G", mut_nt = "A")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "2914"           "TRUE"        "3519"
last_exon_exon_junction          n.exon
        "3382"           "22"
>
> Mutation 11a (LAMB3 c.3119G>A)
> classify.nmd(gene_id = 3914, ref = 37, mut_start = 209790864,
mut_end = 209790864,
+ ref_nt = "C", mut_nt = "T")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "3118"           "TRUE"        "3519"
last_exon_exon_junction          n.exon
        "3382"           "22"
>
> Mutation 12 and 13 (LAMC2 c.132_135delCAGA)
> classify.nmd(gene_id = 3918, ref = 37, mut_start = 183177068,
mut_end = 183177071,
+ ref_nt = "CAGA", mut_nt = "")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "319"            "TRUE"        "3578"
last_exon_exon_junction          n.exon
        "3325"           "23"
>
> Mutation 14 and 15 (COL17A1 c.2910insT)
> classify.nmd(gene_id = 1308, ref = 37, mut_start = 105798866,
mut_end = 105798866,
+ ref_nt = "A", mut_nt = "AA")
        mut_nmd          note          wt_nmd
        "TRUE"           " "           "FALSE"
        PTC.Stop         have.ptc      mutseq_length
        "2932"           "TRUE"        "4495"
last_exon_exon_junction          n.exon
        "4438"           "55"
```

6.1 Splice site mutation data in .vcf format

```
##contig=<ID=1,length=249250621>
##contig=<ID=2,length=243199373>
##contig=<ID=3,length=198022430>
##contig=<ID=4,length=191154276>
##contig=<ID=5,length=180915260>
##contig=<ID=6,length=171115067>
##contig=<ID=7,length=159138663>
##contig=<ID=8,length=146364022>
##contig=<ID=9,length=141213431>
##contig=<ID=10,length=135534747>
##contig=<ID=11,length=135006516>
##contig=<ID=12,length=133851895>
##contig=<ID=13,length=115169878>
##contig=<ID=14,length=107349540>
##contig=<ID=15,length=102531392>
##contig=<ID=16,length=90354753>
##contig=<ID=17,length=81195210>
##contig=<ID=18,length=78077248>
##contig=<ID=19,length=59128983>
##contig=<ID=20,length=63025520>
##contig=<ID=21,length=48129895>
##contig=<ID=22,length=51304566>
##contig=<ID=X,length=155270560>
##contig=<ID=Y,length=59373566>
##contig=<ID=MT,length=16569>
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
1	209795880	.	C	T	.	.	
1	209803958	.	A	G	.	.	
1	209806133	.	A	T	.	.	
1	209806480	.	T	C	.	.	
1	209811874	.	C	G	.	.	

6.2 Cases 5 and 6 mutation data in .vcf format

#CHROM	POS	ID	REF	ALT
1	209801471	**Case 6**	ACACTGCCCGGT	A
1	209799267	**Case 5**	G	A

6.3 Unformatted MMsplice output

Case	ID	exons	exon_id	gene_id	gene_name	transcript_id	delta_logit_psi
7	1:209795880:C>T	1:209795880-209796025:-	ENSE00002310168.1_1	ENSG00000196878.15_4	LAMB3	ENST00000356082.9_2	-4.1587118
7	1:209795880:C>T	1:209795880-209796025:-	ENSE00002310168.1_1	ENSG00000196878.15_4	LAMB3	ENST00000391911.5_1	-4.1587118
7	1:209795880:C>T	1:209795880-209796025:-	ENSE00002310168.1_1	ENSG00000196878.15_4	LAMB3	ENST00000367030.7_1	-4.1587118
8	1:209803958:A>G	1:209803959-209804080:-	ENSE00000792101.1_1	ENSG00000196878.15_4	LAMB3	ENST00000356082.9_2	-1.0718014
8	1:209803958:A>G	1:209803959-209804080:-	ENSE00000792101.1_1	ENSG00000196878.15_4	LAMB3	ENST00000391911.5_1	-1.0718014
8	1:209803958:A>G	1:209803959-209804080:-	ENSE00000792101.1_1	ENSG00000196878.15_4	LAMB3	ENST00000367030.7_1	-1.0718014
9	1:209811874:C>G	1:209811878-209811993:-	ENSE00000792106.1_1	ENSG00000196878.15_4	LAMB3	ENST00000356082.9_2	-4.0515497
9	1:209811874:C>G	1:209811878-209811993:-	ENSE00000792106.1_1	ENSG00000196878.15_4	LAMB3	ENST00000391911.5_1	-4.0515497
9	1:209811874:C>G	1:209811878-209811993:-	ENSE00000792106.1_1	ENSG00000196878.15_4	LAMB3	ENST00000367030.7_1	-4.0515497
9	1:209811874:C>G	1:209811878-209811993:-	ENSE00000792106.1_1	ENSG00000196878.15_4	LAMB3	ENST00000415782.1_2	-4.0515497
10	1:209806480:T>C	1:209806414-209806478:-	ENSE00000792103.1_1	ENSG00000196878.15_4	LAMB3	ENST00000356082.9_2	-3.1108091
10	1:209806480:T>C	1:209806414-209806478:-	ENSE00000792103.1_1	ENSG00000196878.15_4	LAMB3	ENST00000391911.5_1	-3.1108091
10	1:209806480:T>C	1:209806414-209806478:-	ENSE00000792103.1_1	ENSG00000196878.15_4	LAMB3	ENST00000367030.7_1	-3.1108091
11	1:209806133:A>T	1:209805927-209806121:-	ENSE00000792102.1_1	ENSG00000196878.15_4	LAMB3	ENST00000356082.9_2	-3.8153212
11	1:209806133:A>T	1:209805927-209806121:-	ENSE00000792102.1_1	ENSG00000196878.15_4	LAMB3	ENST00000391911.5_1	-3.8153212
11	1:209806133:A>T	1:209805927-209806121:-	ENSE00000792102.1_1	ENSG00000196878.15_4	LAMB3	ENST00000367030.7_1	-3.8153212

Case	ref_acceptorIntron	ref_acceptor	ref_exon	ref_donor	ref_donorIntron	alt_acceptorIntron	alt_acceptor	alt_exon	alt_donor	alt_donorIntron	pathogenicity	efficiency
7	-3.6467907	0.6808083	-2.7579508	5.361782	0.33129492	-3.6467907	0.6808083	-2.7579508	-0.183179	0.33129492	0.99999839	-5.9138386
7	-3.6467907	0.6808083	-2.7579508	5.361782	0.33129492	-3.6467907	0.6808083	-2.7579508	-0.183179	0.33129492	0.99999839	-5.9138386
7	-3.6467907	0.6808083	-2.7579508	5.361782	0.33129492	-3.6467907	0.6808083	-2.7579508	-0.183179	0.33129492	0.99999839	-5.9138386
8	-3.7807019	3.241016	-3.0677443	0.75753886	0.12461016	-3.7807019	3.241016	-3.0677443	-0.6721739	0.12461016	0.98637605	-1.6157299
8	-3.7807019	3.241016	-3.0677443	0.75753886	0.12461016	-3.7807019	3.241016	-3.0677443	-0.6721739	0.12461016	0.98637605	-1.6157299
8	-3.7807019	3.241016	-3.0677443	0.75753886	0.12461016	-3.7807019	3.241016	-3.0677443	-0.6721739	0.12461016	0.98637605	-1.6157299
9	-3.6589541	3.9429288	-2.581183	4.0741	0.14342465	-3.6589541	3.9429288	-2.581183	-1.328	0.14342465	0.9999977	-5.7646297
9	-3.6589541	3.9429288	-2.581183	4.0741	0.14342465	-3.6589541	3.9429288	-2.581183	-1.328	0.14342465	0.9999977	-5.7646297

Table A4: Unformatted MMsplice output