



Functional categorization of gene regulatory variants that cause Mendelian conditions

Y. H. Hank Cheng^{1,2} · Stephanie C. Bohaczuk¹ · Andrew B. Stergachis^{1,2,3}

Received: 21 June 2023 / Accepted: 30 December 2023
© The Author(s) 2024

Abstract

Much of our current understanding of rare human diseases is driven by coding genetic variants. However, non-coding genetic variants play a pivotal role in numerous rare human diseases, resulting in diverse functional impacts ranging from altered gene regulation, splicing, and/or transcript stability. With the increasing use of genome sequencing in clinical practice, it is paramount to have a clear framework for understanding how non-coding genetic variants cause disease. To this end, we have synthesized the literature on hundreds of non-coding genetic variants that cause rare Mendelian conditions via the disruption of gene regulatory patterns and propose a functional classification system. Specifically, we have adapted the functional classification framework used for coding variants (i.e., loss-of-function, gain-of-function, and dominant-negative) to account for features unique to non-coding gene regulatory variants. We identify that non-coding gene regulatory variants can be split into three distinct categories by functional impact: (1) non-modular loss-of-expression (LOE) variants; (2) modular loss-of-expression (mLOE) variants; and (3) gain-of-ectopic-expression (GOE) variants. Whereas LOE variants have a direct corollary with coding loss-of-function variants, mLOE and GOE variants represent disease mechanisms that are largely unique to non-coding variants. These functional classifications aim to provide a unified terminology for categorizing the functional impact of non-coding variants that disrupt gene regulatory patterns in Mendelian conditions.

Introduction

The genetic basis for multiple Mendelian conditions was initially identified by studying individuals harboring chromosomal translocations, which provided a signpost for where in the genome a gene was disrupted. It quickly became apparent that many of these chromosomal translocations did not disrupt coding sequence, but rather disrupted the positioning of coding sequence relative to a distal regulatory element or gene promoter (Vortkamp et al. 1991; Wallis et al. 1999; Fang et al. 2000; Crisponi et al. 2001). These initial studies helped establish that non-coding genetic variation can cause numerous Mendelian conditions, and work over the past several decades has solidified the central role of

non-coding genetic variation in the pathogenesis of hundreds of Mendelian conditions.

In this review, we compiled hundreds of non-coding genetic variants from ClinVar and the literature that cause rare human diseases via the disruption of gene regulatory patterns. In doing so, we have recognized that there is no unified vocabulary for describing how this class of genetic variation contributes to Mendelian conditions. Specifically, as is detailed in the section “[Functional categorization of gene regulatory variants](#)”, the current functional classification system used for coding variants (i.e., loss-of-function, gain-of-function, and dominant-negative) is not well suited to non-coding variants, because it does not capture the diversity of functional consequences associated with this class of genetic variation. Furthermore, simply describing non-coding variants based on their distance to gene promoters is similarly inadequate. We present a new functional classification system for describing non-coding variants impacting regulatory elements (Table 1) and provide specific examples of variants that fall into each category of this new functional classification system (Tables 2, 3, 4).

Of note, non-coding variants can cause Mendelian conditions via a myriad of mechanisms, and this review

✉ Andrew B. Stergachis
absterga@uw.edu

¹ Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA

² Department of Genome Sciences, University of Washington, Seattle, WA, USA

³ Brotman Baty Institute for Precision Medicine, Seattle, WA, USA

Table 1 Regulatory variant functional classifications

Functional classification	Definition
Non-modular loss-of-expression (LOE) variants	Genetic variants impacting gene regulatory elements that diminish or completely abolish the intrinsic expression pattern of a gene
Modular loss-of-expression (mLOE) variants	Genetic variants impacting gene regulatory elements that diminish or completely abolish the intrinsic expression pattern of a gene only in a subset of cell types or developmental stages that express that gene
Gain-of-ectopic-expression (GOE) variants	Genetic variants impacting gene regulatory elements that ectopically increase the expression pattern of a gene

specifically focuses on rare variants that cause Mendelian conditions by disrupting gene regulatory elements. Other classes of non-coding genetic variants include intronic variants that disrupt transcript splicing, 5'UTR genetic variants that alter initiation codon usage, 3'UTR and/or 5'UTR genetic variants that impact transcript stability, localization, or signal response, genetic variants within non-coding RNAs (ncRNAs), and non-coding repeat expansions that form altered RNA products (Stenson et al. 2017; French and Edwards 2020). Additionally, CpG methylation alterations at imprinted loci cause a handful of Mendelian conditions without altering the underlying DNA sequence (Cerrato et al. 2020). Common non-coding variants associated with common disease risk are covered elsewhere (Zhang and Lupski 2015; Spielmann and Mundlos 2016; French and Edwards 2020).

Structure of gene regulatory elements

Although the DNA content of a gene is present in every cell of the body, each gene may only be expressed within certain cell types and/or developmental time windows (herein referred to as the 'intrinsic' expression pattern of a gene). The intrinsic expression pattern of a gene is governed by regulatory elements, which are short DNA segments (often less than 400 bp in size) containing short binding elements (less than 30 bp) that govern the occupancy of sequence-specific transcription factors (TFs). Regulatory elements are distinguished based on their location relative to the transcriptional start site (TSS) of a gene (i.e., promoter-proximal, vs distal), as well as their functional impact on transcription (i.e., enhancers vs insulators vs silencers).

Promoters overlap the TSS of a gene and contain binding elements (e.g., TATA, CAAT, GC, CACCC boxes, etc.) that modulate RNA polymerase II binding and transcription (Juven-Gershon et al. 2008). Distal regulatory elements have complex roles in regulating the activity of promoters. For example, enhancers upregulate the transcriptional activity of a gene, and can be located adjacent to or over a megabase away from their target gene (Panigrahi and O'Malley 2021). Expression of each gene can be dependent on multiple enhancers,

some of which may be common to several cell types. This combinatorial system allows different cell types to express the same gene through overlapping, but distinct, regulatory mechanisms. For example, there are nine different neural enhancers with overlapping spatial expression domains that drive *Sonic Hedgehog* (*SHH*) expression in the brain. Between different brain regions, the set of active enhancers is overlapping, but not identical. This mechanism allows for nuanced control of *SHH* expression in different parts of the brain (Amano 2020).

Distal regulatory elements can also serve as insulators, which function to compartmentalize adjacent gene regulatory domains along the genome (Gaszner and Felsenfeld 2006). For example, binding of the sequence-specific TF CTCF can create a barrier that limits an enhancer from regulating genes located on the opposite side of a CTCF boundary (Kim et al. 2015). CTCF-bound insulator sequences are often located at the boundaries of topologically associating domains (TADs), which are large chromatin loops (often > 100 kb in size) that enable the creation of transcriptionally independent chromatin domains wherein the activity of regulatory elements is primarily restricted to genes within the same TAD.

In addition, regulatory elements can exhibit context-specific activity. For example, some regulatory elements can act as either enhancers or silencers depending on their cellular context (Erceg et al. 2017; Huang and Ovcharenko 2022). Furthermore, the classification of regulatory elements is actively evolving as we learn how different regulatory elements influence gene expression in different cellular contexts, or in conjunction with neighboring regulatory elements (Ngan et al. 2020). Finally, although the overwhelming majority of variants that impact gene regulation are in the non-coding genome, coding variants can also impact gene regulatory elements (Lango Allen et al. 2014) as ~3% of all TF binding elements are located within coding sequences (Stergachis et al. 2013).

Functional categorization of gene regulatory variants

The functional consequence of coding variants is classified into three distinct categories, loss-of-function (LOF), gain-of-function (GOF), and dominant-negative (DN). LOF

Table 2 Variants causing LOE

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>ANK1</i>	Hereditary spherocytosis	182900	Moderate	TATA-binding protein (TBP) and TFIIID complex binding element	In promoter	NM_000037.4(ANK1):c.-73_-72del	508	Pathogenic	16037067 (Gallagher et al. 2005)	2005 Sep
<i>ARPC1B</i>	Primary immune disorder	617718	Not specified	Not specified	In promoter	9 kb promoter deletion	N/A	N/A	32581362 (Turro et al. 2020)	2020 Jul
<i>CHRNE</i>	Congenital myasthenic syndrome	608931	Classical disease	N-box-binding element	In promoter	NM_000080.4(CHRNE):c.-11-85C>T	18353	Likely pathogenic	10211467 (Nichols et al. 1999)	1999 Apr
<i>CHRNE</i>	Congenital myasthenic syndrome	608931	Classical disease	N-box-binding element	In promoter	NM_000080.3(CHRNE):c.-95G>A	465866	Pathogenic/likely pathogenic	11960891 (Abicht et al. 2002)	2002 May
<i>CHRNE</i>	Congenital myasthenic syndrome	608931	Classical disease	N-box-binding element	In promoter	NM_000080.3(CHRNE):c.-96G>A	N/A	N/A	10382905 (Ohno et al. 1999)	1999 May
<i>GALT</i>	Galactosemia	230400	Duarte variant	E-box-binding element	In promoter	NM_000155.2(GALT):c.-119_-116delGTCA	25111	Pathogenic/likely pathogenic; other	11286503 (Elsas et al. 2001)	2001 Apr
<i>GJB1</i>	Charcot-Marie-Tooth Neuropathy X	302800	Classical disease	Not specified	In promoter	NM_000166.6(GJB1):c.-103C>T	217166	Pathogenic	8757034 (Ionasescu et al. 1996)	1996 Aug
<i>GJB1</i>	Charcot-Marie-Tooth Neuropathy X	302800	Classical disease	SOX10-binding element	In promoter	NM_001097642.3(GJB1):c.-16-513 T>C	543921	Likely pathogenic	15470753 (Houlden et al. 2004)	2004 Nov
<i>GJB1</i>	Charcot-Marie-Tooth Neuropathy X	302800	Classical disease	SOX10-binding element	In promoter	NM_001097642.3(GJB1):c.-16-513 T>G	10445	Pathogenic	8757034 (Ionasescu et al. 1996)	1996 Aug
<i>GJB1</i>	Charcot-Marie-Tooth Neuropathy X	302800	Classical disease	SOX10-binding element	In promoter	NM_001097642.3(GJB1):c.-16-511G>C	10449	Pathogenic	15470753 (Houlden et al. 2004)	2004 Nov

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>GJC2</i>	Pelizaeus-Merzbacher-Like Disease	608804	Milder disease	SOX10-binding element	In promoter	NM_020435.3(<i>GJC2</i>):c.-167A>G	30759	Pathogenic	24374284 (Gotoh et al. 2014)	2014 Mar
<i>GLIS3</i>	Diabetes mellitus, neonatal, with congenital hypothyroidism	610199	Milder disease	Not specified	In promoter	426 kb deletion encompassing part of the <i>GLIS3</i> 5' UTR	N/A	N/A	16715098 (Senée et al. 2006)	2006 Jun
<i>GLIS3</i>	Diabetes mellitus, neonatal, with congenital hypothyroidism	610199	Milder disease	Not specified	In promoter	134 kb deletion encompassing part of the <i>GLIS3</i> 5' UTR	N/A	N/A	16715098 (Senée et al. 2006)	2006 Jun
<i>GP1BB</i>	Bernard-Soulier Syndrome	231200	Classical disease	GATA-binding element	In promoter	NR_037611.1(<i>SEPT5-GP1BB</i>):n.3581C>G	16041	Pathogenic	8703016 (Ludlow et al. 1996)	1996 Sep
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-136C>T	36284	Likely pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-136C>G	15465	Pathogenic/likely pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-137C>G	15464	Pathogenic/likely pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-137C>A	36285	Pathogenic/likely pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-137C>T	36287	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-138C>A	393701	Pathogenic/likely pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-138C>T	15460	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-140C>T	15514	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(<i>HBB</i>):c.-142C>T	15462	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	CACCC box	In promoter	NM_000518.5(HBB):c.-151C>T	15461	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-78A>C	15470	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-78A>G	15471	Pathogenic/likely pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-79A>C	869288	Pathogenic	11559932 (Sadiq et al. 2001)	2001 Sep
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-79A>G	15469	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-80 T>G	869289	Pathogenic	26635043 (Henderson et al. 2016)	2015 Dec
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-80 T>C	869290	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-80 T>A	15467	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-81A>G	15466	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBB</i>	Beta-thalassemia	613985	Beta-thalassemia intermedia	TATA box	In promoter	NM_000518.5(HBB):c.-82C>A	15492	Pathogenic	11857738; 28385923 (Hardison et al. 2002; Ropero et al. 2017)	2002 Mar 2017 Oct
<i>HBD</i>	Delta-thalassemia	N/A	delta(+)-thalassemia	GATA1 binding element	In promoter	NM_000519.3(HBD):c.-126A>T	15085	Pathogenic	12402333 (De Angioletti et al. 2002)	2002 Nov
<i>HBD</i>	Delta-thalassemia	N/A	delta(+)-thalassemia	GATA1 binding element	In promoter	NM_000519.3(HBD):c.-127 T>C	15072	Pathogenic	1515647 (Matsuda et al. 1992)	1992 Sep
<i>HBD</i>	Delta-thalassemia	N/A	delta(+)-thalassemia	TATA box	In promoter	NM_000519.3(HBD):c.-81A>G	15087	Pathogenic	15921167 (Frischknecht and Dutly 2005)	2004 Nov
<i>HCFC1</i>	X-linked intellectual developmental disorder-3	309541	Classical disease	YY1-binding element	In promoter	NM_005334.3(HCFC1):c.-970 T>C	39694	Pathogenic	23000143 (Huang et al. 2012)	2012 Oct
<i>HNF1A</i>	MODY3	600496	Not specified	Not specified	In promoter	NM_000545.8(HNF1A):c.-119del	14936	Pathogenic	10649494 (Godart et al. 2000)	2000 Jan
<i>LDLR</i>	Familial hypercholesterolemia	143890	Not specified	SP1-binding element	In promoter	NM_000527.4(LDLR):c.-138del	3745	Likely pathogenic	14616764 (Dedoussis et al. 2003)	2003 Nov

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>MLH1</i>	Lynch syndrome	609310	Classical disease	Not specified	In promoter	NM_000249.4(NMLH1):c.-27C>A	89589	Pathogenic	21840485 (Hitchins et al. 2011)	2011 Aug
<i>MPL</i>	Thrombocytopenia	604498	Moderate	Not specified	In promoter	GRCh37:chr1:43803414G>A	N/A	N/A	32581362 (Turro et al. 2020)	2020 Jul
<i>MSH2</i>	Lynch Syndrome	120435	Classical disease	E1A-F binding element	In promoter	NM_000251.2(MSH2):c.-78_-77del	90494	Likely pathogenic	17894833 (Yan et al. 2007)	2007 Dec
<i>NIPBL</i>	Cornelia de Lange syndrome	122470	Mild form	Not specified	In promoter	NM_133433.4(NIPBL):c.-321_-320delinsA	2151	Pathogenic	16799922 (Borck et al. 2006)	2006 Aug
<i>PIGY</i>	Hyperphosphatasia with impaired intellectual development syndrome-6	616809	Milder disease	SP1-binding element	In promoter	NM_001042616.2(PIGY):c.-540G>A	222025	Pathogenic	26293662 (Ilkovski et al. 2015)	2015 Nov
<i>PKLR</i>	Pyruvate kinase deficiency	266200	Severe disease	GATA1 binding element	In promoter	NM_000298.6(PKLR):c.-72A>G	N/A	N/A	11054094 (Manco et al. 2000)	2000 Sep
<i>PKLR</i>	Pyruvate kinase deficiency	266200	Severe disease	Not specified	In promoter	NM_000298.6(PKLR):c.-83G>C	N/A	N/A	12393511 (Van Wijk et al. 2003)	2003 Feb
<i>RBI</i>	Hereditary retinoblastoma	180200	Unilateral retinoblastoma	ATF binding element	In promoter	NM_000321.2(RBI):c.-189G>T	13085	Pathogenic	1881452 (Sakai et al. 1991)	1991 Sep
<i>RBI</i>	Hereditary retinoblastoma	180200	Bilateral and unilateral retinoblastoma	SP1-binding element	In promoter	NM_000321.2(RBI):c.-198G>A	13086	Pathogenic	1881452 (Sakai et al. 1991)	1991 Sep
<i>RBI</i>	Hereditary retinoblastoma	180200	Not specified	SP1-binding element	In promoter	NM_000321.3(RBI):c.-206_-189del	995907	Likely Pathogenic	28873162 (Mandelker et al. 2017)	2017 Sep

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>RBM8A</i>	Radial aplasia-thrombocytopenia syndrome (Thrombocytopenia Absent Radius Syndrome, TAR)	274000	Milder disease	Not specified	In promoter	NM_005105.5(RBM8A):c.-21G>A	30464	Pathogenic/Likely pathogenic; other	22366785 (Albers et al. 2012)	2012 Feb
<i>SH2D1A</i>	X-linked lymphoproliferative disease	308240	Not specified	CCAAT box	In promoter	NM_002351.5(SH2D1A):c.-10C>T	10906	Pathogenic	9771704 (Coffey et al. 1998)	1998 Oct
<i>SLC39A4</i>	Acrodermatitis enteropathica	201100	Classical disease	CCAAT box	In promoter	NM_130849.4(SLC39A4):c.-169A>G	N/A	N/A	https://doi.org/10.1101/2022.09.09.22279746 (Galey et al. 2022)	2022 Sep
<i>SPINK1</i>	Juvenile onset chronic pancreatitis	167800	Not specified	Not specified	In promoter	NM_003122.5(SPINK1):c.-191-24G>A	13762	Pathogenic	11355022 (Kaneko et al. 2001)	2001 Feb
<i>TXNL4A</i>	Burn-McKown syndrome	608572	Classical disease	Not specified	In promoter	NC_000018.9:g.77748581_77748614del34	162203	Conflicting interpretations of pathogenicity	25434003 (Wieczorek et al. 2014)	2014 Dec
<i>TXNL4A</i>	Burn-McKown syndrome	608572	Classical disease	Not specified	In promoter	NC_000018.9(TXNL4A):g.77748604_77748637del34	190413	Pathogenic	25434003 (Wieczorek et al. 2014)	2014 Dec
<i>UROS</i>	Congenital erythropoietic porphyria (CEP)	263700	Severe disease	GATA1 binding element	In promoter	NM_000375.3(UROS):c.-203 T>C	3762	Pathogenic	11254675 (Solis et al. 2001)	2001 Mar
<i>UROS</i>	Congenital Erythropoietic Porphyria (CEP)	263700	Mild disease	Not specified	In promoter	NM_000375.3(UROS):c.-26-183G>A	3763	Pathogenic	11254675 (Solis et al. 2001)	2001 Mar

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>UROS</i>	Congenital erythro-poietic porphyria (CEP)	263700	Mild disease	Not specified	In promoter	NM_000375.3(<i>UROS</i>):c.-26-193C>A	3764	Pathogenic	11254675 (Solis et al. 2001)	2001 Mar
<i>UROS</i>	Congenital erythro-poietic porphyria (CEP)	263700	Severe disease	CP2-binding element	In promoter	NM_000375.3(<i>UROS</i>):c.-26-197C>A	3765	Pathogenic	11254675 (Solis et al. 2001)	2001 Mar
<i>VHL</i>	von Hippel-Lindau disease	193300	Not specified	Not specified	In promoter	NM_000551.3(<i>VHL</i>):c.-75_-55del	166561	Likely Pathogenic	22357542 (Wu et al. 2012)	2012 Apr
Promoter-distal regulatory variants										
<i>DLX5</i>	Split hand-split foot malformation (SHFM)	183600	Reduced penetrance	Not specified	850 kb upstream	1 Mb deletion	N/A	N/A	26075025 (Delgado and Velinov 2015)	2015 Jun
<i>DLX5</i>	Split hand-split foot malformation (SHFM)	183600	Reduced penetrance	DYNC111 exonic enhancers	835 kb upstream	106 kb deletion	N/A	N/A	24459211 (Allen et al. 2014)	2014 Apr
<i>DLX5</i>	Split hand-split foot malformation (SHFM)	183600	Not specified	DYNC111 exonic enhancers	370 kb upstream	t(2;7)(p25.1;q22) translocation	N/A	N/A	24459211 (Allen et al. 2014)	2014 Apr
<i>FOXC2</i>	Lymphedema-distichiasis syndrome	153400	Not specified	Not specified	120 kb downstream	Translocation	N/A	N/A	11078474 (Fang et al. 2000)	2000 Dec
<i>FOXP1</i>	Ret syndrome, congenital variant	613454	Severe phenotype	Enhancer	265 kb downstream	Translocation	N/A	N/A	21441262 (Kortüm et al. 2011)	2011 Jun
<i>FOXL2</i>	Blepharophthalmos-ptosis-epicanthus inversus syndrome	110100	Not specified	Not specified	180 kb upstream	Translocation	N/A	N/A	11175783 (Crisponi et al. 2001)	2001 Feb

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>GATA2</i>	MonoMAC syndrome	614172	Classical phenotype	Enhancer	9.5 kb downstream	28 bp deletion	N/A	N/A	23502222; 22996659 (Johnson et al. 2012; Hsu et al. 2013)	2013 May 2012 Oct
<i>GATA2</i>	MonoMAC syndrome	614172	Classical phenotype	ETS binding element	9.5 kb downstream	NM_032638.5(GATA2):c.1017+572C>T	566562	Pathogenic/Likely pathogenic	23502222; 20040766 (Vinh et al. 2010; Hsu et al. 2013)	2013 May 2010 Feb
<i>GLL3</i>	Greig syndrome	175700	Classical phenotype	Not specified	10 kb downstream	Translocation	N/A	N/A	1650914 (Vortkamp et al. 1991)	1991 Aug
<i>HBA</i>	Alpha-thalassemia	604131	Not specified	Not specified	10 kb upstream	62 kb deletion	N/A	N/A	2364173 (Haton et al. 1990)	1990 Jul
<i>HBA</i>	Alpha-thalassemia	604131	Not specified	HS-40 element	60 kb upstream	Multiple deletions	N/A	N/A	18391781 (Higgs and Wood 2008)	2008 May
<i>HBB</i>	Beta-thalassemia	613985	Milder disease	Not specified	50 kb upstream	Translocation	N/A	N/A	6318113 (Kiousis et al. 1983)	1983 Dec
<i>IRF6</i>	Van der Woude syndrome (VWS)	119300	Classical phenotype	Disruption of p63 and E47-binding element, and creation of Lef1 binding element	9.7 kb upstream	MCS9.7-350dupA	N/A	N/A	24442519 (Fakhouri et al. 2014)	2014 May
<i>LRBA</i>	Autoantibody-mediated pancytopenia	614700	Not specified	CTCF-binding element	53 kb downstream	7.7 kb deletion	N/A	N/A	32581362 (Turro et al. 2020)	2020 Jul
<i>MAF</i>	Cataract, anterior segment dysgenesis and microphthalmia	610202	Severe phenotype (associated with unbalanced translocation)	Not specified	1 Mb upstream	Translocation	N/A	N/A	11772997 (Jamieson et al. 2002)	2002 Jan

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>PAH</i>	Phenylketonuria	261600	Not specified	Enhancer	407 bp upstream	NM_001354304.2(PAH):c.-95-4071_-95-313del, 3.7 kb deletion	638	Likely pathogenic	11935335 (Chen et al. 2002)	2002 Mar
<i>PAX6</i>	Aniridia	106210	Classical phenotype	Not specified	150 kb downstream	Translocation	N/A	N/A	7795596 (Fantes et al. 1995)	1995 Mar
<i>PAX6</i>	Aniridia	106210	Classical phenotype	PAX6 autoregulatory element	150 kb downstream	NM_019040.5(ELP4):c.1143+14176C>A	120328	Pathogenic	24290376 (Bhatta et al. 2013)	2013 Dec
<i>PITX2</i>	Axenfeld–Rieger syndrome	180500	Not specified	Not specified	10 kb upstream	7.6 Mb deletion	N/A	N/A	20881290 (Volkmann et al. 2011)	2011 Mar
<i>PITX2</i>	Axenfeld–Rieger syndrome	180500	Not specified	Not specified	18 kb upstream	Translocation	N/A	N/A	14991915 (Trembath et al. 2004)	2004 Feb
<i>PITX2</i>	Axenfeld–Rieger syndrome	180500	Not specified	Not specified	76 kb upstream	Translocation	N/A	N/A	14991915 (Trembath et al. 2004)	2004 Feb
<i>PLP1</i>	Spastic paraplegia type 2	312920	Milder disease	Not specified	136 kb downstream	150 kb duplication	N/A	N/A	16374829 (Lee et al. 2006)	2006 Feb
<i>POU3F4</i>	X-linked deafness type 3 (DFN3)	304400	Classical phenotype	Not specified	900 kb downstream	Common 8 kb deletion	N/A	N/A	8872461 (De Kok et al. 1996)	1996 Sep
<i>SERPINH1</i>	Osteogenesis imperfecta (OI)	613848	Moderate phenotype	Enhancer	2.37 kb upstream	5,274 bp deletion	N/A	N/A	31179625 (Schwarze et al. 2019)	2019 Aug
<i>SHH</i>	Holoprosencephaly	142945	Unspecified	Not specified	265 kb upstream	t(2,7)(q31;q36)	N/A	N/A	9254845 (Roessler et al. 1997)	1997 Aug
<i>SHH</i>	Holoprosencephaly	142945	Unspecified	SBE2 enhancer	460 kb upstream	C>T variant	N/A	N/A	18836447 (Jeong et al. 2008)	2008 Nov
<i>SHOX</i>	Leri–Weill dyschondrosteosis (LWD)	127300	Classical phenotype	Not specified	205 kb downstream	common 29 kb deletion	N/A	N/A	16175500 (Benito-Sanz et al. 2005)	2005 Oct

Table 2 (continued)

Gene	Disease	OMIM phenotype MIM number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SHOX</i>	Leri-Weill dyschondrosteosis (LWD) and idiopathic short stature (ISS)	127300, 300582	Milder disease	Enhancer	160 kb downstream	NC_000024.9:g.730550_778092del, 47 kb deletion	66087	Pathogenic	22791839 (Benito-Sanz et al. 2012a, b)	2012 Jul
<i>SIX3</i>	Holoprosencephaly	157170	Not specified	Not specified	10–200 kb upstream	Translocation	N/A	N/A	10369266 (Wallis et al. 1999)	1999 Jun
<i>SOX9</i>	46,XY male-to-female sex reversal and acampomelic form of campomelic dysplasia	114290	Classical phenotype	Not specified	380 kb upstream	1.5 Mb deletion	N/A	N/A	15060123 (Pop et al. 2004)	2004 Apr
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	917–855 kb upstream	t(7;17)(p13;q24)	N/A	N/A	23648064 (Fonseca et al. 2013)	2013 May
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	601–585 kb upstream	t(17;20)(q24.3;q11.2)	N/A	N/A	23648064 (Fonseca et al. 2013)	2013 May
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	50 kb upstream	t(7;17)(q34;q25.1)	N/A	N/A	23648064; 8348155; 8001137 (Tommerup et al. 1993; Wagner et al. 1994; Fonseca et al. 2013)	2013 May 1993 Jun 1994 Dec
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	74–88 kb upstream	t(12;17)(q21.32;q24.3-q25.1)	N/A	N/A	23648064; 7747782; 8789441 (Ninomiya et al. 1995, 1996; Fonseca et al. 2013)	2013 May 1995 Mar 1996 Jan

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SOX9</i>	Campomelic dysplasia	114290	Not specified	Not specified	88 kb upstream	t(2;17)(q35;q23-q24)	N/A	N/A	23648064; 7990924; 1583645 (Young et al. 1992; Foster et al. 1994; Fonseca et al. 2013)	2013 May 1994 Dec 1992 Apr
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	110–140 kb upstream	t(9;17)	N/A	N/A	23648064; 9724758 (Wunderle et al. 1998; Fonseca et al. 2013)	2013 May 1998 Sep
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	134–142 kb upstream	t(13;17)(q22;q25.1)	N/A	N/A	23648064; 10364523; 8348155; 8566951 (Tommerup et al. 1993; Wirth et al. 1996; Pfeifer et al. 1999; Fonseca et al. 2013)	2013 May 1999 Jul 1993 Jun 1996 Feb
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	173–179 kb upstream	t(1;17)(q42.13;q24.3-q25.1)	N/A	N/A	23648064; 10364523; 8348155; 8566951 (Tommerup et al. 1993; Wirth et al. 1996; Pfeifer et al. 1999; Fonseca et al. 2013)	2013 May 1999 Jul 1993 Jun 1996 Feb

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SOX9</i>	Campomelic dysplasia	114290	Not specified	Not specified	161 kb upstream	t(5;17)(q23.2;q24)	N/A	N/A	23648064; 21890680 (Sobreira et al. 2011; Fonseca et al. 2013)	2013 May 2011 Oct
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	212–224 kb upstream	t(6;17)(q14;q24)	N/A	N/A	23648064; 10364523; 8566951 (Wirth et al. 1996; Pfeifer et al. 1999; Fonseca et al. 2013)	2013 May 1999 Jul 1996 Feb
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	228–229 kb upstream	t(10;17)(q24;q23)	N/A	N/A	23648064; 10364523 (Pfeifer et al. 1999; Fonseca et al. 2013)	2013 May 1999 Jul
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	288–319 kb upstream	t(5;17)(q13.3;q24.2)	N/A	N/A	23648064; 10364523 (Pfeifer et al. 1999; Fonseca et al. 2013)	2013 May 1999 Jul
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	70–350 kb upstream	inv(17)(q11.2;q24.3-q25.1)	N/A	N/A	23648064; 7666392; 9724758 (Mansour et al. 1995; Wunderle et al. 1998; Fonseca et al. 2013)	2013 May 1995 Jun 1998 Sep

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	375 kb upstream	t(1;17)(q42.1;q24.3)	N/A	N/A	23648064; 17204049 (Leipoldt et al. 2007; Fonseca et al. 2013)	2013 May 2007 Jan
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	380–1869 kb upstream	del(17)(q24.3)	N/A	N/A	23648064; 15060123 (Pop et al. 2004; Fonseca et al. 2013)	2013 May 2004 Apr
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	517–1477 kb upstream	del(17)(q24.3)	39777	Pathogenic	23648064; 19449405 (Lecointre et al. 2009; Fonseca et al. 2013)	2013 May 2009 Jun
<i>SOX9</i>	Campomelic dysplasia	114290	Not specified	Not specified	500–4700 kb upstream	t(7;17)(q33;q24)del(17)(q24.2q24.3)	N/A	N/A	23648064; 20453475 (Jakubiczka et al. 2010; Fonseca et al. 2013)	2013 May 2010 May
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	789 kb upstream	t(Y;17)(q11.2;q24.3)	N/A	N/A	23648064; 17204049 (Leipoldt et al. 2007; Fonseca et al. 2013)	2013 May 2007 Jan
<i>SOX9</i>	Pierre Robin sequence (PRS) and campomelic dysplasia	261800, 114290	Milder disease	Not specified	899 kb upstream	t(4;17)(q28.3;q24.3)	N/A	N/A	23648064; 15726498 (Velagleti et al. 2005; Fonseca et al. 2013)	2013 May 2005 Apr

Table 2 (continued)

Gene	Disease	OMIM phenotype number	Disease severity	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SOX9</i>	Pierre Robin sequence (PRS) and campomelic dysplasia	261800, 114290	Not specified	Not specified	932 kb upstream	t(13;17)(q22.1;q22.3)	N/A	N/A	23648064; 15717285 (Hill-Harfe et al. 2005; Fonseca et al. 2013)	2013 May 2005 Apr
<i>SOX9</i>	Campomelic dysplasia	114290	Milder disease	Not specified	900 kb upstream	t(17;22)(q25.1;p11.2)	N/A	N/A	23648064; 10364523; 15717285 (Pfeifer et al. 1999; Hill-Harfe et al. 2005; Fonseca et al. 2013)	2013 May 1999 Jul 2005 Apr
<i>SRY</i>	46,XY sex reversal	400044	Classical phenotype	Not specified	1.8 kb upstream	25 kb deletion	N/A	N/A	1438307 (McElreavy et al. 1992)	1992 Nov
<i>TWIST1</i>	Saethre-Chotzen syndrome	101400	Classical phenotype	Not specified	260 kb downstream	Inversions and translocation	N/A	N/A	14513358 (Cai et al. 2003)	2003 Dec
<i>UNC13D</i>	Familial hemophagocytic lymphohistiocytosis 3	608898	Unspecified	ELF1-binding element	< 1 kb downstream	NM_199242.3(UNC13D):c.118-308C>T	533095	Pathogenic/Likely pathogenic	21931115 (Meeths et al. 2011)	2011 Nov

(Galey et al. 2022; Kioussis et al. 1983; Hatton et al. 1990; Sakai et al. 1991; Vortkamp et al. 1991; Matsuda et al. 1992; McElreavy et al. 1992; Young et al. 1992; Tommerup et al. 1993; Foster et al. 1994; Wagner et al. 1994; Mansour et al. 1995; Nimomiya et al. 1995, 1996; Fantes et al. 1995; Ionasescu et al. 1996; Ludlow et al. 1996; Wirth et al. 1996; De Kok et al. 1996; Roessler et al. 1997; Coffey et al. 1998; Wunderte et al. 1998; Nichols et al. 1999; Pfeifer et al. 1999; Ohno et al. 1999; Wallis et al. 1999; Fang et al. 2000; Godart et al. 2000; Manco et al. 2000; Elsas et al. 2001; Kaneko et al. 2001; Sadiq et al. 2001; Solis et al. 2001; Crisponi et al. 2002; De Angioletti et al. 2002; Hardison et al. 2002; Jamieson et al. 2002; Chen et al. 2002; Dedoussis et al. 2003; Van Wijk et al. 2003; Cai et al. 2003; Houlden et al. 2004; Pop et al. 2004; Trembath et al. 2004; Benito-Sanz et al. 2005, 2012b; Frischknecht and Dutly 2005; Hill-Harfe et al. 2005; Velagaleti et al. 2005; Gallagher et al. 2005; Lee et al. 2006; Senée et al. 2006; Boreck et al. 2006; Leipoldt et al. 2007; Yan et al. 2007; Higgs and Wood 2008; Jeong et al. 2008; Lecointre et al. 2009; Vinh et al. 2010; Jakubiczka et al. 2010; Hitchins et al. 2011; Volkman et al. 2011; Sobreira et al. 2011; Meeths et al. 2011; Albers et al. 2012; Wu et al. 2012; Johnson et al. 2012; Huang et al. 2013; Hsu et al. 2013; Bhatia et al. 2014; Fakhouri et al. 2014; Gotoh et al. 2014; Wieczorek et al. 2014; Delgado and Velinov 2015; Ilkowski et al. 2015; Henderson et al. 2016; Mandelker et al. 2017; Ropero et al. 2017; Schuler et al. 2018; Schwarze et al. 2019; Turro et al. 2020)

Table 3 Variants causing mLOE

Gene	Disease	OMIM phenotype MIM number	OMIM associated with gene LOF variants	OMIM phenotype MIM number for LOF	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>APC</i>	Gastric adenocarcinoma and proximal polyposis of the stomach	619182	Familial Adenomatous Polyposis	175100	YY1-binding element	APC exon 1B promoter	NM_001127511.3(APC):c.-195A>C	264670	Pathogenic	27087319 (Li et al. 2016)	2016 May
<i>APC</i>	Gastric adenocarcinoma and proximal polyposis of the stomach	619182	Familial Adenomatous Polyposis	175100	YY1-binding element	APC exon 1B promoter	NM_001127511.2(APC):c.-125delA;-195A>C]	243004	Pathogenic	27087319 (Li et al. 2016)	2016 May
<i>APC</i>	Gastric adenocarcinoma and proximal polyposis of the stomach	619182	Familial Adenomatous Polyposis	175100	YY1-binding element	APC exon 1B promoter	NM_001127511.3(APC):c.-192A>T	243007	Pathogenic	27087319 (Li et al. 2016)	2016 May
<i>APC</i>	Gastric adenocarcinoma and proximal polyposis of the stomach	619182	Familial Adenomatous Polyposis	175100	YY1-binding element	APC exon 1B promoter	NM_001127511.3(APC):c.-192A>G	243006	Likely pathogenic	27087319 (Li et al. 2016)	2016 May
<i>APC</i>	Gastric adenocarcinoma and proximal polyposis of the stomach	619182	Familial Adenomatous Polyposis	175100	YY1-binding element	APC exon 1B promoter	NM_001127511.3(APC):c.-191T>C	243005	Pathogenic	27087319 (Li et al. 2016)	2016 May
<i>APC</i>	Gastric adenocarcinoma and proximal polyposis of the stomach	619182	Familial Adenomatous Polyposis	175100	YY1-binding element	APC exon 1B promoter	NM_001127511.3(APC):c.-190G>A	243008	Pathogenic	27087319 (Li et al. 2016)	2016 May
<i>F9</i>	Hemophilia B Leyden (mild form)	306900	Hemophilia B	306900	C/EBP-binding element	F9 promoter	NM_000133.4(F9):c.-17del	10561	Pathogenic	2342576 (Crossley and Brownlee 1990)	1990 May

Table 3 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF variants	OMIM phenotype MIM number for LOF	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>F9</i>	Hemophilia B Leyden (mild form)	306900	Hemophilia B	306900	C/EBP-binding element	F9 promoter	NM_000133.4(F9):c.-22 T>C	10644	Pathogenic	2004020 (Royle et al., 1991)	1991 Feb
<i>F9</i>	Hemophilia B Leyden (mild form)	306900	Hemophilia B	306900	C/EBP-binding element	F9 promoter	NM_000133.4(F9):c.-17A>G	10646	Pathogenic	2342576 (Crossley & Brownlee 1990)	1990 May
<i>F9</i>	Hemophilia B Leyden (mild form)	306900	Hemophilia B	306900	Not specified	F9 promoter	NM_000133.3(F9):c.-35G>A	641767	Pathogenic	2388855 (Crossley et al. 1990)	1990 Aug
<i>GATA1</i>	Low platelet count and normal RBC parameters	N/A	Severe platelet and RBC abnormalities	300367	Enhancer	6 kb downstream	4 kb deletion	N/A	N/A	32581362 (Turro et al. 2020)	2020 Jul
<i>PIGM</i>	Glycosylphosphatidylinositol deficiency	610293	No gene LOF phenotype reported	N/A	SP1-binding element	PIGM promoter	NM_145167.2(PIGM):c.-270C>G	1288	Pathogenic	16767100 (Almeida et al. 2006)	2006 Jul
<i>PTF1A</i>	Isolated pancreatic agenesis	615935	Pancreatic and cerebellar agenesis	609069	Developmental pancreatic enhancer	19 kb downstream	7.6 kb deletion	N/A	N/A	24212882 (Weedon et al. 2014)	2014 Jan
<i>PTF1A</i>	Isolated pancreatic agenesis	615935	Pancreatic and cerebellar agenesis	609069	Developmental pancreatic enhancer	25 kb downstream	GRCh37:chr10:23508437A>G	N/A	N/A	24212882 (Weedon et al. 2014)	2014 Jan
<i>PTF1A</i>	Isolated pancreatic agenesis	615935	Pancreatic and cerebellar agenesis	609069	Developmental pancreatic enhancer	25 kb downstream	GRCh37:chr10:23508363A>G	N/A	N/A	24212882 (Weedon et al. 2014)	2014 Jan
<i>PTF1A</i>	Isolated pancreatic agenesis	615935	Pancreatic and cerebellar agenesis	609069	Developmental pancreatic enhancer	25 kb downstream	GRCh37:chr10:23508305A>G	N/A	N/A	24212882 (Weedon et al. 2014)	2014 Jan

Table 3 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF variants	OMIM phenotype MIM number for LOF	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>PTF1A</i>	Isolated pancreatic agenesis	615935	Pancreatic and cerebellar agenesis	609069	Developmental pancreatic enhancer	25 kb downstream	GRCh37:chr10:23508365A>G	N/A	N/A	24212882 (Weedon et al. 2014)	2014 Jan
<i>PTF1A</i>	Isolated pancreatic agenesis	615935	Pancreatic and cerebellar agenesis	609069	Developmental pancreatic enhancer	25 kb downstream	GRCh37:chr10:23508446A>C	N/A	N/A	24212882 (Weedon et al. 2014)	2014 Jan
<i>SHOX</i>	Idiopathic short stature (ISS)	300582	Leri-Weill dyschondrosteosis	127300	Enhancer	95 kb upstream	286 kb deletion	N/A	N/A	22071895 (Benito-Sanz et al. 2012a, b)	2012 Jan
<i>SOX9</i>	Pierre Robin sequence (PRS)	261800	Campomelic dysplasia	114290	Not specified	1.38 Mb upstream	75 kb deletion	N/A	N/A	19234473 (Benko et al. 2009)	2009 Mar
<i>SOX9</i>	Pierre Robin sequence (PRS)	261800	Campomelic dysplasia	114290	Craniofacial region enhancer	1.44 Mb upstream	T>C variant	N/A	N/A	19234473 (Benko et al. 2009)	2009 Mar
<i>SOX9</i>	Pierre Robin sequence (PRS)	261800	Campomelic dysplasia	114290	Not specified	1.58 Mb upstream	> 319 kb deletion	N/A	N/A	19234473 (Benko et al. 2009)	2009 Mar
<i>SOX9</i>	Pierre Robin sequence (PRS)	261800	Campomelic dysplasia	114290	Not specified	1.16 Mb upstream	t(2;17)(q32;q24)	N/A	N/A	19234473 (Benko et al. 2009)	2009 Mar
<i>SOX9</i>	Pierre Robin sequence (PRS)	261800	Campomelic dysplasia	114290	Not specified	1.03 Mb upstream	t(5;17)(q15;q24)	N/A	N/A	19234473 (Benko et al. 2009)	2009 Mar
<i>SOX9</i>	Pierre Robin sequence (PRS)	261800	Campomelic dysplasia	114290	Proximal mandibular mesenchyme enhancer	1.23 Mb upstream	t(2;17)(q24.1;q24.3)	N/A	N/A	19234473 (Benko et al. 2009)	2009 Mar

Table 3 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF variants	OMIM phenotype MIM number for LOF	Regulatory element(s) impacted	Distance to promoter	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SOX9</i>	Pierre Robin sequence (PRS)	261800	Campomelic dysplasia	114290	Not specified	1.56 Mb downstream	36 kb deletion	N/A	N/A	19234473 (Benko et al. 2009)	2009 Mar
<i>TBX5</i>	Isolated congenital heart disease	N/A	Holt-Oram Syndrome	142900	TAL1 binding element	90 kb downstream	GRCCh37:chr12:114704515G>T	N/A	N/A	22543974 (Smemo et al. 2012)	2012 Jul
<i>UNC13D</i>	Recurrent Macrophage Activation Syndrome and Systemic Juvenile Idiopathic Arthritis	N/A	Familial hemophagocytic lymphohistiocytosis 3	608898	NF-κB binding element	UNC13D promoter	NM_199242.3(UNC13D):c.117 + 143A>G	1299418	Not provided	29409136 (Schulert et al. 2018)	2018 Jun

(Crossley and Brownlee 1990; Crossley et al. 1990; Royle et al. 1991; Almeida et al. 2006; Benko et al. 2009; Benito-Sanz et al. 2012a; Smemo et al. 2012; Weedon et al. 2014; Li et al. 2016; Turro et al. 2020)

Table 4 Variants causing GOE

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-134G>A	30853	Pathogenic/Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-128G>C	626920	Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-128G>A	812727	Pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-128G>T	812728	Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-127A>G	626942	Pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-127A>T	626943	Pathogenic/Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-127A>C	1175767	Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-126 T>C	626941	Pathogenic/Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-126 T>G	1684447	Pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-118C>G	626940	Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-118C>T	627410	Pathogenic/Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-118C>A	812726	Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>ANKRD26</i>	Thrombocytopenia 2	188000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>ANKRD26</i> expression	NM_014915.3(ANKRD26):c.-116C>G	627189	Likely pathogenic	21211618 (Pipucci et al. 2011)	2011 Jan
<i>BMP2</i>	Autosomal-dominant brachydactyly type A2 (BDA2)	112600	LOF variants cause Short Stature, Facial Dysmorphism, and Skeletal Anomalies	617877	Disruption of limb-specific enhancer of <i>BMP2</i>	Not specified	4.6–5.9 kb duplications located 110 kb downstream of <i>BMP2</i> gene	29614	Pathogenic	19327734; 21357617 (Dathe et al. 2009; Su et al. 2011)	2009 Apr 2011 May
<i>CTSB</i>	Keratolytic winter erythema (KWE)	148370	No gene LOF phenotype reported	N/A	Duplicated <i>CTSB</i> enhancer	Increased <i>CTSB</i> expression	7.6 and 15.9 kb duplications located 8 kb upstream of <i>CTSB</i> gene	N/A	N/A	28457472 (Ngungcu et al. 2017)	2017 May
<i>CYP11B1</i>	Glucocorticoid-remediable aldosteronism (GRA)	103900	LOF variants cause congenital adrenal hyperplasia/congenital hypoaldosteronism	202010	Entire <i>CYP11B2</i> promoter disrupted	Misexpression of <i>CYP11B2</i> gene under control of <i>CYP11B1</i> promoter	<i>CYP11B1</i> , <i>CYP11B1/CYP11B2</i> ANTI-LEPORE-LIKE CHIMERA	1172	Pathogenic	1731223 (Lifton et al. 1992)	1992 Jan
<i>HBG1</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	CCAAT box	Persistence of gamma-globin expression into adulthood	NM_000559.2(HBG1):c.-170G>A	15030	Pathogenic	1379347 (Berry et al. 1992)	1992 Aug
<i>HBG1</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	CCAAT box	Persistence of gamma-globin expression into adulthood	NM_000559.2(HBG1):c.-167C>T	15035	Pathogenic	1704803; 1698280 (Fuchareon et al. 1990; Oner et al. 1991)	1991 Mar 1990 Sep

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>HBG1</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	ZBTB7A-binding element	Persistence of gamma-globin expression into adulthood via removal of ZBTB7A repression	NM_000559.2(HBG1):c.-53-195C>G	15034	Pathogenic	2224140 (Costa et al. 1990)	1990 Nov
<i>HBG1</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	ZBTB7A-binding element	Persistence of gamma-globin expression into adulthood via removal of ZBTB7A repression	NM_000559.2(HBG1):c.-53-196C>T	15033	Pathogenic	2417646; 2423160; 3033668; 1487421 (Gelinas et al. 1986; Loudianos et al. 1992; Pirastu et al. 1987; Weber et al. 1986)	1986 Feb 1986 Jun 1987 May 1992 Sep
<i>HBG1</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	ZBTB7A-binding element	Persistence of gamma-globin expression into adulthood via removal of ZBTB7A repression	NM_000559.2(HBG1):c.-53-198 T>C	15031	Pathogenic	2430647 (Tate et al. 1986)	1986 Dec
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	Not specified	Persistence of gamma-globin expression into adulthood	NM_000184.2(HBG2):c.-255C>G	14982	Pathogenic	6208955 (Collins et al. 1984)	1984 Dec
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	Not specified	Persistence of gamma-globin expression into adulthood	NM_000184.2(HBG2):c.-228 T>C	14983	Pathogenic	2441598; 7687855 (Craig et al. 1993; Huang et al. 1987)	1987 Aug 1993 May

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	CCAAT box	Persistence of gamma-globin expression into adulthood	NM_000184.2(HBG2):c.-167C>T	14990	Pathogenic	1698280 (Fuchareon et al. 1990)	1990 Sep
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	CCAAT box	Persistence of gamma-globin expression into adulthood	NM_000184.2(HBG2):c.-167C>A	15001	Pathogenic	10335983 (Zertal-Zidani et al. 1999)	1999 May
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	BCL11A-binding element	Persistence of gamma-globin expression into adulthood via removal of BCL11A repression	c.-117G>A	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	BCL11A-binding element	Persistence of gamma-globin expression into adulthood via removal of BCL11A repression	c.-114C>T	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	BCL11A-binding element	Persistence of gamma-globin expression into adulthood via removal of BCL11A repression	c.-114C>A	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	BCL11A-binding element	Persistence of gamma-globin expression into adulthood via removal of BCL11A repression	c.-114C>G	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	BCL11A-binding element	Persistence of gamma-globin expression into adulthood via removal of BCL11A repression	13 bp deletion ~115 bp upstream of TSS	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	ZBTB7A-binding element	Persistence of gamma-globin expression into adulthood via removal of ZBTB7A repression	c.-197C>T	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	ZBTB7A-binding element	Persistence of gamma-globin expression into adulthood via removal of ZBTB7A repression	c.-201C>T	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	ZBTB7A-binding element	Persistence of gamma-globin expression into adulthood via removal of ZBTB7A repression	c.-202C>T	N/A	N/A	29610478 (Martyn et al. 2018)	2018 Apr

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>HBG2</i>	Hereditary persistence of fetal hemoglobin	141749	No gene LOF phenotype reported	N/A	ZBTB7A-binding element	Persistence of gamma-globin expression into adulthood via removal of ZBTB7A repression	c.-202C>G	N/A	N/A	29610478 (Martyń et al. 2018)	2018 Apr
<i>IHH</i>	Polydactyly	N/A	No gene LOF phenotype reported	N/A	CTCF-associated TAD boundary	Duplication brings the <i>IHH/Ihh</i> gene in proximity to the centromeric portion of the EPHA4-containing TAD. This causes <i>IHH</i> ectopic expression in the developing limb, resembling endogenous <i>Epha4</i> expression	~900 kb duplication in chromosomal region 2q35	N/A	N/A	25959774 (Lupiáñez et al. 2015)	2015 May
<i>MMP19</i>	Cavitary optic disk anomaly (CODA)	611543	No gene LOF phenotype reported	N/A	Enhancer	Increased <i>MMP19</i> expression	GRCh38/hg38 12q13.2(chr12:55,845,043–55851177)×4, 6 kb triplication located 2.1 kb upstream of the <i>MMP19</i>	180748	Pathogenic	25581579 (Hazlewood et al. 2015)	2015 Mar
<i>MTX2/HOXD genes</i>	Mesomelic dysplasia and vertebral defects	N/A	<i>MTX2</i> LOF variants cause Mandibuloacral dysplasia progeroid syndrome	619,127	Disruption of distal <i>MTX2</i> regulatory elements	Not specified	t(2:8)(q31;p21) balanced translocation with breakpoint between <i>MTX2</i> gene and <i>HOXD</i> gene cluster	N/A	N/A	11944980 (Spitz et al. 2002)	2002 Apr

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>OVOL2</i>	Corneal dystrophy, posterior polymorphous, 1	122000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>OVOL2</i> expression	NM_001303461.1(OVOL2):c.-297+886 T>C	224838	Pathogenic	26749309 (Davidson et al. 2016)	2016 Jan
<i>OVOL2</i>	Corneal dystrophy, posterior polymorphous, 1	122000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>OVOL2</i> expression	NM_001303461.1(OVOL2):c.-297+895_-297+916dup	224837	Pathogenic	26749309 (Davidson et al. 2016)	2016 Jan
<i>OVOL2</i>	Corneal dystrophy, posterior polymorphous, 1	122000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>OVOL2</i> expression	NM_001303461.1(OVOL2):c.-297+949 T>C	224839	Pathogenic	26749309 (Davidson et al. 2016)	2016 Jan
<i>OVOL2</i>	Corneal dystrophy, posterior polymorphous, 1	122000	No gene LOF phenotype reported	N/A	Not specified	Increased <i>OVOL2</i> expression	NM_021220.4(OVOL2):c.-274 T>G	224840	Pathogenic	26749309 (Davidson et al. 2016)	2016 Jan
<i>PAX3</i>	Brachydactyly	N/A	PAX3 LOF variants cause Waardenburg syndrome	193500	CTCF-associated TAD boundary	Deletions remove the predicted boundary between the EPHA4 and PAX3 TADs. This causes PAX3 ectopic expression at the developing limb, in a pattern resembling endogenous EphA4 expression	Deletions of 1.75–1.9 Mb on chromosome 2q35–36	N/A	N/A	25959774 (Lupiáñez et al. 2015)	2015 May

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>PITX1</i>	Homeotic arm-to-leg transformation	186550	<i>PITX1</i> LOF variants cause congenital clubfoot with or without deficiency of long bones and/or mirror-image polydactyly	119800	Disruption of <i>PITX1</i> forelimb suppressor/insulator elements	Not specified	134 kb deletion located 269 kb upstream of the <i>PITX1</i>	N/A	N/A	23022097 (Spielmann et al. 2012)	2012 Oct
<i>PITX1</i>	Homeotic arm-to-leg transformation	186550	<i>PITX1</i> LOF variants cause congenital clubfoot with or without deficiency of long bones and/or mirror-image polydactyly	119800	Disruption of <i>PITX1</i> forelimb suppressor/insulator elements	Not specified	107 kb deletion located 269 kb upstream of the <i>PITX1</i>	N/A	N/A	23022097 (Spielmann et al. 2012)	2012 Oct
<i>PITX1</i>	Homeotic arm-to-leg transformation	186550	<i>PITX1</i> LOF variants cause congenital clubfoot with or without deficiency of long bones and/or mirror-image polydactyly	119800	Disruption of <i>PITX1</i> forelimb suppressor/insulator elements	Not specified	Translocation with breakpoint 224 kb upstream of the <i>PITX1</i>	N/A	N/A	23022097 (Spielmann et al. 2012)	2012 Oct

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>POMP</i>	Keratosis linearis with ichthyosis congenita and sclerosis keratoderma (KLICK) syndrome	601952	No gene LOF phenotype reported	N/A	Not specified	Switch in transcription start sites for <i>POMP</i>	NM_015932.5(<i>POMP</i>):c.-95del	116	Pathogenic	20226437 (Dahlqvist et al. 2010)	2010 Apr
<i>SHH</i>	Laurin-Sandrow syndrome (LSS)	135750	<i>SHH</i> LOF variants cause holoprosencephaly	142945	Sonic hedgehog (<i>SHH</i>) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic <i>SHH</i> expression	NC_000007.13:g.156563856_156610632dup, 46.8 kb duplication encompassing the <i>SHH</i> ZRS	157547	Pathogenic	24456159 (Lohan et al. 2014)	2014 Oct
<i>SHH</i>	Laurin-Sandrow syndrome (LSS)	135750	<i>SHH</i> LOF variants cause holoprosencephaly	142945	Sonic hedgehog (<i>SHH</i>) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic <i>SHH</i> expression	NC_000007.13:g.156570780_156646750dup, 76 kb duplication encompassing the <i>SHH</i> ZRS	157548	Pathogenic	24456159 (Lohan et al. 2014)	2014 Oct
<i>SHH</i>	Laurin-Sandrow syndrome (LSS)	135750	<i>SHH</i> LOF variants cause holoprosencephaly	142945	Sonic hedgehog (<i>SHH</i>) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic <i>SHH</i> expression	NM_022458.3(<i>LMBR1</i>):c.320-5565_423+10975dup, 16.6 kb duplication encompassing the <i>SHH</i> ZRS	157546	Pathogenic	24456159 (Lohan et al. 2014)	2014 Oct
<i>SHH</i>	Haas-type polysyndactyly (syndactyly type IV)	186200	<i>SHH</i> LOF variants cause holoprosencephaly	142945	Sonic hedgehog (<i>SHH</i>) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic <i>SHH</i> expression	GRCh37:chr7:156,437,229-156,692,706 duplication, 255 kb duplication encompassing the <i>SHH</i> ZRS	N/A	N/A	24456159 (Lohan et al. 2014)	2014 Oct

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SHH</i>	Haas-type polysyndactyly (syndactyly type IV)	186200	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	GRCh37:chr7:156,491,887–156,671,016 duplication, 179 kb duplication encompassing the SHH ZRS	N/A	N/A	24456159 (Lohan et al. 2014)	2014 Oct
<i>SHH</i>	Haas-type polysyndactyly (syndactyly type IV)	186200	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	LMBR1, 235-KB DUP, IVS5, duplication encompassing the SHH ZRS	4905	Pathogenic	18417549 (Sun et al. 2008)	2008 Sep
<i>SHH</i>	Triphalangeal thumb-polysyndactyly syndrome	174500	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	NG_009240.1:g.(71605_101850)_(134420_151298)dup, duplication encompassing the SHH ZRS	4901	Pathogenic	18417549 (Sun et al. 2008)	2008 Sep
<i>SHH</i>	Preaxial polydactyly	174500	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	t(5.7)(q11,q36), translocation breakpoint at the SHH ZRS	N/A	N/A	12032320 (Lettice et al. 2002)	2002 May
<i>SHH</i>	Preaxial polydactyly	174500	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	NM_022458.4(LMBR1):c.423+5252A>G	4902	Pathogenic	17152067 (Gunnnett et al. 2007)	2007 Jan

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	Variant	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SHH</i>	Preaxial polydactyly	174500	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	NM_022458.4(LMBR1):c.423+5134C>G	4903	4903	Pathogenic	17152067 (Gurnett et al. 2007)	2007 Jan
<i>SHH</i>	Triphalangeal thumb	174500	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	NM_022458.4(LMBR1):c.423+4808 T>C	4906	4906	Pathogenic	18463159 (Furniss et al. 2008)	2008 Aug
<i>SHH</i>	Werner meselic syndrome (WMS)	N/A	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	NM_022458.4(LMBR1):c.423+4917G>C	155921	155921	Pathogenic	19847792 (Wieczorek et al. 2010)	2010 Jan
<i>SHH</i>	Werner meselic syndrome (WMS)	N/A	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	NM_022458.4(LMBR1):c.423+4917G>A	4898	4898	Pathogenic	19847792 (Wieczorek et al. 2010)	2010 Jan
<i>SHH</i>	Haas-type polysyndactyly (syndactyly type IV)	186200	SHH LOF variants cause holoprosencephaly	142945	Sonic hedgehog (SHH) limb enhancer ZPA regulatory sequence (ZRS)	Ectopic SHH expression	LMBR1, 73-KB DUP, duplication encompassing the SHH ZRS	155922	155922	Pathogenic	19847792 (Wieczorek et al. 2010)	2010 Jan

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>SLC16A1</i>	Exercise-induced hyperinsulinism (EIH)	610021	<i>SLC16A1</i> LOF variants cause Monocarboxylate transporter 1 deficiency	616095	Not specified	Misexpression of <i>SLC16A1</i> in beta-cells	NM_003051.4(<i>SLC16A1</i>):c.-202G>A	8916	Pathogenic	17701893 (Otonkoski et al. 2007)	2007 Sep
<i>SLC16A1</i>	Exercise-induced hyperinsulinism (EIH)	610021	<i>SLC16A1</i> LOF variants cause Monocarboxylate transporter 1 deficiency	616095	Not specified	Misexpression of <i>SLC16A1</i> in beta-cells	NM_003051.3(<i>SLC16A1</i>):c.-391_-390insACG CCGGTCACGTGGCGGGGTGGGG	8917	Pathogenic	17701893 (Otonkoski et al. 2007)	2007 Sep
<i>SOX9</i>	46,XX female-to-male sex reversal	278850	<i>SOX9</i> LOF variants cause cam-pomelic dysplasia	114290	Disruption of distal <i>SOX9</i> regulatory element	Increased <i>SOX9</i> expression	178 kb duplication located 600 kb upstream of <i>SOX9</i>	30708	Pathogenic	21208124 (Cox et al. 2011)	2011 Jan
<i>SOX9</i>	46,XX female-to-male sex reversal	278850	<i>SOX9</i> LOF variants cause cam-pomelic dysplasia	114290	Disruption of distal <i>SOX9</i> regulatory element	Increased <i>SOX9</i> expression	96 kb triplication 500 kb upstream of <i>SOX9</i>	192386	Pathogenic	21653197 (Vetro et al. 2011)	2011 Oct
<i>SOX9</i>	Brachydaetyly-anonychia (Cooks syndrome)	106995	<i>SOX9</i> LOF variants cause cam-pomelic dysplasia	114290	Disruption of distal <i>SOX9</i> regulatory element	Not specified	1.21–1.96 mb duplications > 0.1 mb upstream of <i>SOX9</i>	N/A	N/A	19639023 (Kurth et al. 2009)	2009 Aug
<i>SOX9</i>	46,XX female-to-male sex reversal	278850	<i>SOX9</i> LOF variants cause cam-pomelic dysplasia	114290	Disruption of distal <i>SOX9</i> regulatory element	Increased <i>SOX9</i> expression	t(12;17)(q14.3;q24.3) with translocation breakpoint > 776 upstream of <i>SOX9</i>	N/A	N/A	20082466 (Refai et al. 2010)	2010 Feb

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>TERT</i>	Familial melanoma predisposition syndrome	615134	<i>TERT</i> LOF variants cause dyskeratosis congenita	613989	Ets/TCF binding element	Altered <i>TERT</i> expression	NM_198253.3(<i>TERT</i>):c.-57A>C	242210	Conflicting interpretations of pathogenicity Pathogenic(1); Uncertain significance(2)	23348503 (Horn et al. 2013)	2013 Feb
<i>WNT6</i>	F-syndrome	N/A	No gene LOF phenotype reported	N/A	CTCF-associated TAD boundary	An inversion and a duplication both result in bringing the centromeric portion of the <i>EPHA4</i> -containing TAD into closer proximity of the <i>WNT6</i> gene. This causes <i>WNT6</i> ectopic expression in the developing limb, in a pattern resembling endogenous <i>Epha4</i> expression	~ 1.1 Mb inversion with telomeric breakpoint located 1.4 Mb away from the <i>EPHA4</i> gene and centromeric breakpoint located telomeric of <i>WNT6</i>	N/A	N/A	25959774 (Lupiáñez et al. 2015)	2015 May

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>WNT6</i>	F-syndrome	N/A	No gene LOF phenotype reported	N/A	CTCF-associated TAD boundary	An inversion and a duplication both result in bringing the centromeric portion of the <i>EPHA4</i> -containing TAD into closer proximity of the <i>WNT6</i> gene. This causes <i>WNT6</i> ectopic expression in the developing limb, in a pattern resembling endogenous <i>Epha4</i> expression	~1.4 Mb duplication with telomeric breakpoint located 1.2 Mb away from the <i>EPHA4</i> gene and centromeric breakpoint located centromeric of <i>WNT6</i>	N/A	N/A	25959774 (Lupiáñez et al. 2015)	2015 May
Variants mimicking gene duplication events											
<i>FOXG1</i>	Rett syndrome congenital variant	613454	<i>FOXG1</i> gene duplications cause Rett syndrome, congenital variant	613454	Silencer	Increased <i>FOXG1</i> expression	0.4 to 2.1 Mb deletions located <100 kb to 637 kb downstream of <i>FOXG1</i>	N/A	N/A	22739344 (Allou et al. 2012)	2012 Jun

Table 4 (continued)

Gene	Disease	OMIM phenotype MIM number	Phenotype associated with gene LOF or duplication variants	OMIM phenotype MIM number associated with gene LOF or duplication variants	Regulatory element(s) impacted	Mechanism	Variant ID	ClinVar ID	ClinVar classification	PMID(s)	Publication date
<i>LMNB1</i>	Autosomal-dominant adult-onset demyelinating leukodystrophy (ADLD)	169500	LMNB1 duplications cause Autosomal-dominant adult-onset demyelinating leukodystrophy (ADLD)	169500	Not specified	Not specified	660 kb heterozygous deletion located 66 kb upstream of LMNB1	N/A	N/A	25701871 (Giorgio et al. 2014)	2015 Jun
<i>NR0B1</i>	46,XY male-to-female sex reversal	300018	NR0B1 gene duplications cause 46,XY male-to-female sex reversal	300018	Not specified	Not specified	257 kb deletion located 11 kb upstream of the NR0B1 gene	N/A	N/A	17503084 (Smyk et al. 2007)	2007 Aug
<i>PMP22</i>	Charcot-Marie-Tooth disease, type 1A	118220	PMP22 gene duplications cause Charcot-Marie-Tooth disease, type 1A	118220	Not specified	Not specified	194 kb duplication located 9 kb upstream of PMP22	N/A	N/A	20493460 (Zhang et al. 2010)	2010 Jun
<i>PMP22</i>	Charcot-Marie-Tooth disease, type 1A	118220	PMP22 gene duplications cause Charcot-Marie-Tooth disease, type 1A	118220	Not specified	Not specified	186 kb duplication located 34 kb upstream of PMP22	N/A	N/A	20493460 (Zhang et al. 2010)	2010 Jun

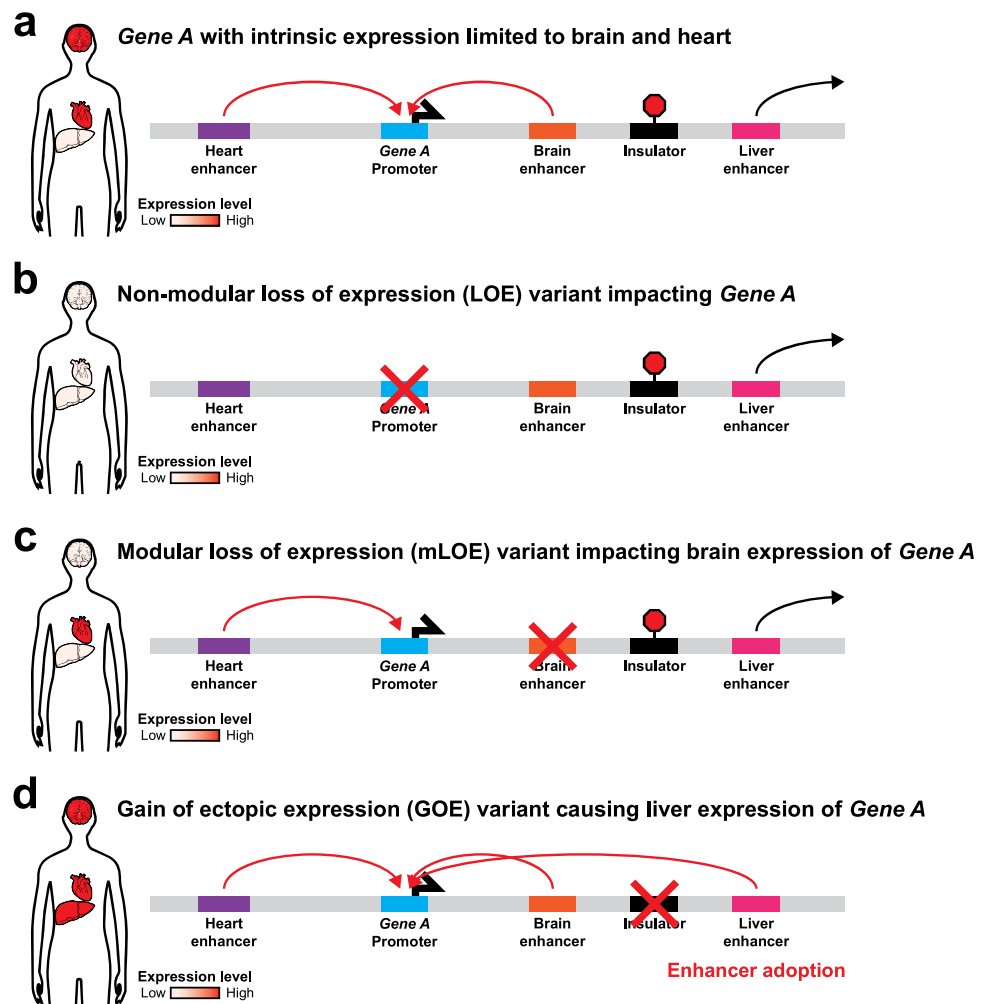
(Collins et al. 1984; Gelinas et al. 1986; Tate et al. 1986; Waber et al. 1986; Huang et al. 1987; Pirastu et al. 1987; Costa et al. 1990; Fucharoen et al. 1990; Oner et al. 1991; Berry et al. 1992; Lifton et al. 1992; Loudianos et al. 1992; Craig et al. 1993; Zertal-Zidani et al. 1999; Spitz et al. 2002; Lettice et al. 2002; Gurnett et al. 2007; Smyk et al. 2007; Otonkoski et al. 2007; Furniss et al. 2008; Sun et al. 2008; Dathe et al. 2009; Kurth et al. 2009; Wiczorek et al. 2010; Refai et al. 2010; Zhang et al. 2010; Cox et al. 2011; Pippucci et al. 2011; Su et al. 2011; Vetro et al. 2011; Spielmann et al. 2012; Allou et al. 2012; Horn et al. 2013; Lohan et al. 2014; Giorgio et al. 2014; Hazlewood et al. 2015; Lupiáñez et al. 2015; Davidson et al. 2016; Ngcungcu et al. 2017; Martyn et al. 2018)

variants result in the loss of the normal biological function of a protein via either complete (amorphic) or partial (hypomorphic) LOF. In contrast, GOF variants create a protein with a function distinct from that of the wild-type protein via increasing protein activity (hypermorphic) or creating a completely new function (neomorphic). DN variants create a protein that either directly or indirectly blocks the normal function of the remaining wild-type protein (antimorphic). Notably, both GOF and DN variants are largely defined based on their alterations to the protein product of a gene. In contrast, since gene regulatory variants do not alter the protein sequence of a gene, but rather modulate gene expression patterns, these coding-centric categorizations are often ill suited for gene regulatory variants and create confusion.

We propose an alternative framework designed to functionally categorize genetic variants that disrupt gene regulatory elements. Specifically, we identify three distinct classes based on their impact on gene regulation at the level of gene transcripts (Table 1): (1) non-modular loss-of-expression (LOE) variants; (2) modular loss-of-expression (mLOE) variants; and (3) gain-of-ectopic-expression (GOE)

variants. LOE variants are defined as variants that diminish or completely abolish the expression of a gene universally across all cell types that intrinsically express that gene. In contrast, mLOE variants are defined as variants that diminish or completely abolish the expression of a gene within a limited subset of the cell types or developmental windows that intrinsically express it (i.e., a modular loss of expression). GOE variants are defined as variants that result in the ectopic spatial and/or temporal expression of a gene (Fig. 1). Of note, unlike LOE variants, we chose not to further subdivide GOE variants into modular GOE variants, as it is quite challenging to obtain the appropriate clinical and molecular data that are necessary to firmly state that a GOE variant is truly limited to only a specific developmental window or cell type. For example, with mLOE variants, one can infer that there is a modular loss of expression based on a modular phenotype when compared to coding LOF variants in the same gene, which cause loss of protein function in all tissues or developmental windows. In contrast, with GOE variants, one cannot readily infer from clinical data that the ectopic gain of expression is limited to only a specific cell

Fig. 1 Schematic of the proposed gene regulatory variants functional classifications. **a** Schematic depicting the intrinsic expression pattern of a gene that is limited to the heart and brain, as well as the gene regulatory elements driving this intrinsic expression pattern. **b–d** Schematics depicting the impact of **b** LOE, **c** mLOE, and **d** GOE variants on the expression pattern of that gene across different tissues, as well as how various genetic variants can cause these changes



type. Specifically, a gene product may gain ectopic expression across all cell types, but only have a functional consequence in a select number of cell types—limiting the utility of tissue-selective phenotypes for inferring modular ectopic expression.

As opposed to the traditional LOF, GOF, and DN categories, this functional categorization more intuitively reflects the mechanisms by which disruptions in gene expression patterns cause Mendelian conditions. Of note, these functional classifications can be related to LOF or GOF variant types. For example, LOE variants can correspond to either amorphic or hypomorphic LOF variants. Although mLOE variants could be categorized as hypomorphic LOF variants, the mechanism by which mLOE variants cause disease is quite distinct from that of coding hypomorphic LOF variants, making ‘hypomorphic LOF’ an imprecise label for mLOE variants. In contrast, GOE variants can correspond to hypermorphic GOF, neomorphic GOF, and even LOF variants. Notably, there are no examples of DN gene regulatory variants causing Mendelian conditions in humans. However, this type of regulatory variant has been observed in other organisms and is termed “transvection” (Lewis 1954), which is a phenomenon where a regulatory element on one chromosome interacts with and enhances or silences its corresponding regulatory element on the homologous chromosome. More recently, this mechanism has been described in human cancers, wherein strong enhancers encoded on extrachromosomal circular DNA (ecDNA) can enhance the expression of autosomal genes (Zhu et al. 2021). It is possible that examples of transvection as a cause of Mendelian disease could be described in the future. The LOE, mLOE, and GOE functional categorizations represent the molecular consequences of regulatory variants more closely than the traditional LOF, DN, and GOF classification, and provide an improved framework for conceptualizing the putative role of novel gene regulatory variants in the pathogenesis of Mendelian conditions.

Non-modular loss-of-expression (LOE) variants

LOE variants diminish or abolish the expression of a gene across all cell types that intrinsically express that gene. Consequently, these variants often mirror the clinical manifestations of coding LOF variants for the same gene, as both LOE and LOF variants result in reduced/absent functional protein levels within the cell, albeit via distinct mechanisms. While some LOE variants cause complete loss of expression (analogous to amorphic LOF variants), others reduce the intrinsic expression level of a gene (analogous to hypomorphic LOF variants). The latter class of variants often results in a more attenuated clinical phenotype compared to variants

that result in complete LOE. We have provided examples of over a hundred LOE variants (Table 2), and detail below some key examples of diverse LOE variants.

Many LOE variants are located within the gene promoter, where they disrupt essential TF binding elements required for the intrinsic expression of a gene. For example, genetic variants that disrupt the TATA box and/or CACCC box within the *HBB* gene promoter decrease the intrinsic expression of *HBB* by abrogating the ability of TFs to bind these elements. Notably, these variants do not completely abolish *HBB* transcription, and consequently, individuals harboring these variants in trans with *HBB* LOF variants often still produce adult hemoglobin (HbA), resulting in a milder form of beta-thalassemia (i.e., beta-thalassemia intermedia) compared to individuals with biallelic *HBB* LOF coding variants (Ropero et al. 2017).

Of note, different variants within the same gene promoter can cause varying magnitudes of LOE. For example, variants within the *UROS* promoter that disrupt the GATA1 or CP2-binding elements significantly reduce *UROS* transcription and cause a severe form of congenital erythropoietic porphyria (CEP), whereas other *UROS* promoter variants that do not disrupt these elements only cause a modest reduction in *UROS* transcription and mild cutaneous manifestations (Solis et al. 2001).

LOE variants can also disrupt distal regulatory elements. For example, monocytopenia and mycobacterial infection (MonoMAC) syndrome is typically caused by LOF coding variants within the gene *GATA2*. However, MonoMAC syndrome can also be caused by small deletions or single-nucleotide variants (SNVs) in a *GATA2* intronic enhancer 9.5 kb downstream of the *GATA2* promoter. These variants result in the loss of *GATA2* expression via the disruption of enhancer TF binding elements that are essential for *GATA2* transcription (i.e., an E-box, GATA, and ETS binding element) (Johnson et al. 2012; Hsu et al. 2013).

Variants within regulatory elements that are quite distal to a gene promoter can also cause LOE. For example, most cases of hereditary aniridia are caused by heterozygous LOF coding variants within *PAX6*. However, hereditary aniridia can also be caused by SNVs within a *PAX6* enhancer located 150 kb downstream of the *PAX6* promoter that disrupt a *PAX6* autoregulatory element, causing loss of enhancer activity and subsequent loss of *PAX6* transcription (Bhatia et al. 2013). Furthermore, some patients with hereditary aniridia will have deletions or chromosomal translocations that disrupt this *PAX6* enhancer (Fantes et al. 1995), highlighting the diversity of genetic variant classes that can cause LOE.

Variants that disrupt insulators can also cause LOE. For example, a homozygous deletion of a CTCF-binding site within the first intron of *LRBA* has been reported to cause autoantibody-mediated pancytopenia, a phenotype

associated with biallelic coding LOF *LRBA* variants (Turro et al. 2020). It is presumed that loss of this CTCF insulator element alters a TAD boundary, permitting heterochromatin spreading to silence *LRBA* promoter activity.

In summary, LOE variants can be located within promoter or distal gene regulatory elements, can completely mimic coding LOF variants or cause an attenuated phenotype relative to complete LOF variants, and are caused by diverse classes of genetic variants.

Modular loss-of-expression (mLOE) variants

In contrast to non-modular LOE variants, mLOE variants reduce or abolish the expression of a gene only in a subset of cell types that intrinsically express that gene. mLOE variants represent a disease mechanism largely unique to gene regulatory variants, as coding LOF variants typically disrupt the function of a gene across all cell types that intrinsically express that gene, with the exception of coding LOF variants within exons that are alternatively spliced only within certain tissues or somatic coding LOF variants that only exist within certain tissues (Poduri et al. 2013; Biesecker and Spinner 2013; Jaiswal and Ebert 2019). As a result of their modular impact on gene expression, mLOE variants can produce a subset of features associated with coding LOF variants in that same gene (i.e., phenotype modularity) (Table 3). As gene expression patterns are not typically measured across multiple tissues or developmental stages in individuals with Mendelian conditions, the modular nature of these variants is often inferred based on their phenotypic spectrum relative to individuals with coding LOF variants.

To illustrate the functional impact of mLOE variants, it is helpful to compare the full phenotype associated with coding LOF variants to the modular phenotype associated with mLOE variants in a gene regulatory element for the same gene. For example, coding LOF variants in *GATA1* result in both severe platelet and red blood cell abnormalities, because *GATA1* expression is critical for both of these cell types (Gutiérrez et al. 2020). In contrast, a 4 kb deletion of a megakaryocyte-specific enhancer element for *GATA1* is associated with platelet abnormalities, but normal red blood cell parameters (Turro et al. 2020), as this enhancer is necessary for *GATA1* expression within megakaryocytes but not within red blood cells.

Similarly, whereas coding LOF variants in *PTF1A* cause both pancreatic and cerebellar agenesis (Sellick et al. 2004), deletions or single-nucleotide variants within a pancreas-specific enhancer located 25 kb downstream of *PTF1A* cause only isolated pancreatic agenesis, likely because *PTF1A* expression during cerebellar neurogenesis is maintained (Weedon et al. 2014).

mLOE variants can also be located in promoters. For example, LOF variants in *APC* cause familial adenomatous polyposis, a condition associated with adenocarcinoma and numerous polyps in the stomach and colon. However, *APC* has two distinct promoters termed 1A and 1B, and *APC* transcription within the stomach mucosa is selectively initiated via promoter 1B. Consequently, individuals with variants in *APC* promoter 1B are at risk for developing gastric adenocarcinoma and proximal polyposis isolated to the stomach (GAPPS) without colon polyposis as a comorbidity (Li et al. 2016).

By selectively disrupting the expression of a gene in only a particular cell type, mLOE variants have the potential to produce a disease phenotype mediated by genes associated with embryonic lethality in the context of coding LOF variants. For example, biallelic LOF variants in *PIGM* are embryonic lethal in mice. In contrast, biallelic variants within the *PIGM* promoter that disrupt an SP1-binding element cause an inherited glycosylphosphatidylinositol deficiency characterized by a propensity for venous thrombosis and seizures (Almeida et al. 2006). The modular phenotype associated with this promoter variant results from the differential importance of this SP1 element in *PIGM* expression across cell types (Costa et al. 2014).

In addition to cell type selectivity, mLOE variants can also cause loss of expression at particular developmental stages. For example, variants that disrupt a C/EBP or HNF4-binding element within the *F9* gene promoter cause Hemophilia B Leyden, which is characterized by severe factor IX deficiency at birth that ameliorates after puberty (Veltkamp et al. 1970). The affected C/EBP- and HNF4-binding elements are essential for *F9* transcription in early childhood. However, after puberty, androgen-responsive TFs bind to an androgen response element within the *F9* promoter, dramatically increasing *F9* transcription to levels that largely resolve the disease phenotype (Crossley et al. 1992). Consequently, Hemophilia B Leyden is caused by a modular loss of *F9* expression only within the prepubescent developmental stage.

In summary, mLOE variants are located within promoter or distal gene regulatory elements, can restrict the disease phenotype associated with coding LOF variants to only a specific tissue or developmental stage, and can result in a disease phenotype for genes wherein coding LOF variants would be embryonic lethal.

Gain-of-ectopic-expression (GOE) variants

GOE variants cause ectopic spatial and/or temporal expression patterns and represent a disease mechanism that is largely unique to regulatory variants (Table 4). Notably, some GOE variants can mimic Mendelian

conditions caused by duplications of the target gene. For example, autosomal-dominant adult-onset demyelinating leukodystrophy (ADLD) is caused by overexpression of *LMNB1* protein usually attributed to duplication of the *LMNB1* gene. However, an ADLD family was discovered to have a deletion that begins 66 kb upstream of the *LMNB1* promoter. This deletion encompasses a TAD boundary and results in overexpression of *LMNB1* protein via a mechanism termed ‘enhancer adoption’. Specifically, a strong enhancer that typically does not regulate *LMNB1* is now brought into the same TAD as the *LMNB1* promoter, resulting in *LMNB1* overexpression analogous to that seen with *LMNB1* duplication (Giorgio et al. 2014).

Enhancer adoption is a common mechanism through which structural variants can cause regulatory element GOE (Fig. 1D). For example, structural variants within the *WNT6/IHH/EPHA4/PAX3* locus can cause distinct phenotypes depending on where a strong cluster of limb enhancers for *EPHA4* is situated relative to the *WNT6*, *IHH*, or *PAX3* genes. Specifically, deletion of a TAD boundary between *EPHA4* and *PAX3* results in *PAX3* adopting this cluster of limb enhancers, resulting in ectopic *PAX3* expression and brachydactyly. In contrast, inversions or duplications involving *IHH* and the TAD boundary between *IHH* and *EPHA4* result in *WNT6* adopting this cluster of limb enhancers, resulting in ectopic *WNT6* expression and F-syndrome (Lupiañez et al. 2015).

SNVs within distal regulatory elements can also cause GOE. For example, the zone of polarizing activity regulatory sequence (ZRS), located in intron 5 of the *LMBR1* gene, regulates *SHH*. SNVs within the ZRS located ~1 Mb upstream of *SHH* cause preaxial polydactyly (Lettice et al. 2002; Gurnett et al. 2007; Furniss et al. 2008) via the introduction of novel ETV2-binding sites in the ZRS, resulting in ectopic *SHH* expression within the developing limb bud (Koyano-Nakagawa et al. 2022). However, it is important to recognize that for a given gene, not all SNVs within distal regulatory elements result in the same phenotype, as non-coding SNVs within the *SHH* brain enhancer-2 (SBE2) located 460 kb upstream of *SHH* cause holoprosencephaly via an LOE mechanism (Jeong et al. 2008).

In addition to distal regulatory elements, GOE variants can also affect promoters. For example, glucocorticoid-remediable aldosteronism (GRA) is caused by ‘promoter switching’ between the genes *CYP11B1* and *CYP11B2*, resulting in a chimeric gene wherein the adrenocorticotropic hormone (ACTH)-responsive promoter of the 11-beta-hydroxylase gene (*CYP11B1*) is fused with the coding region of the aldosterone synthase gene (*CYP11B2*) (Lifton et al. 1992). This results in ectopic expression of aldosterone synthase in zona fasciculata cells of the adrenal cortex, causing aldosterone synthase to be overexpressed and inducible by

ACTH, hence a hyperaldosteronism state that normalizes upon treatment with glucocorticoids.

GOE variants can also cause Mendelian conditions for which the target gene does not have a known human phenotype associated with coding LOF variants, such as when coding LOF variants would result in embryonic lethality. This is notable, because the clinical identification of non-coding variants that cause Mendelian conditions is often informed by comparison to known LOF phenotypes. For example, complete loss of *OVOL2* expression has been associated with embryonic lethality in mice, likely because *OVOL2* is a transcription factor critical for epithelial cell lineage determination and differentiation (Mackay et al. 2006). Meanwhile, in humans, *OVOL2* promoter variants that result in GOE can cause autosomal-dominant corneal endothelial dystrophies. These promoter variants result in the creation of binding elements for several activating TFs within the *OVOL2* promoter, resulting in the inappropriate ectopic expression of *OVOL2* in the developing or adult corneal endothelium (Davidson et al. 2016).

Promoter GOE variants can also disrupt the ability of transcriptional repressors to appropriately silence a gene at a particular developmental stage. For example, the gamma-globin genes *HBG1* and *HBG2* encode a component of fetal hemoglobin (HbF) and are normally expressed only during fetal erythropoiesis, as their promoters are silenced during adult erythropoiesis by the transcriptional repressors *BCL11A* and *ZBTB7A*. However, regulatory variants within the *HBG1* and *HBG2* promoters that disrupt *BCL11A*- and *ZBTB7A*-binding elements result in the hereditary persistence of fetal hemoglobin (HPFH) into adulthood (Martyn et al. 2018). As HbF is capable of preventing red blood cell sickling from sickle hemoglobin (HbS) and can compensate for deficient HbA as seen in beta-thalassemia, these HPFH variants can attenuate the phenotype of sickle cell disease and beta-thalassemia (Jackson et al. 1961; Cappellini et al. 1981; Labie et al. 1985; Weatherall 2001; Thein 2008; Thein et al. 2009). The discovery of HPFH variants has fortuitously enabled the development of gene editing therapies, which introduce these variants into adult erythroid progenitor cells to reactivate HbF as treatment for sickle cell disease and beta-thalassemia (Traxler et al. 2016; Li et al. 2021).

GOE variants and LOE variants impacting the same gene can lead to similar phenotypes. For example, a ‘Goldilocks’ level of *FOXG1* expression is likely required for normal brain development, because both *FOXG1* duplications and deletions are associated with Rett-like phenotypes (Florian et al. 2012). Thus, it is unsurprising that GOE variants that remove a silencer and LOE variants that remove an enhancer have both been reported to cause Rett-like phenotypes via increasing and decreasing *FOXG1* expression, respectively (Kortüm et al. 2011; Allou et al. 2012).

Finally, variants within gene regulatory elements can cause mixed effects. For example, the *POMP* gene typically has a short 5'UTR that originates from a TSS located at position c.-81. A single-nucleotide deletion in the *POMP* promoter at position c.-95 does not change the overall transcript levels of *POMP*, but results in decreased utilization of the canonical TSS and increased utilization of an upstream TSS located at position c.-181. This results in *POMP* transcripts that preferentially contain a long 5'UTR with reduced translational efficiency. Consequently, *POMP* expression within the granular layer of the epidermis is reduced, causing keratosis linearis with ichthyosis congenita and sclerosing keratoderma (KLICK) syndrome (Dahlqvist et al. 2010). This example illustrates how non-coding variants can have mixed effects, wherein they result in GOE of one transcript, LOE of a different transcript, and LOF at the protein level. In contrast, coding GOF variants in *POMP* result in proteasome-associated autoinflammatory syndrome 2 (PRAAS2) which has a quite distinct clinical presentation (Poli et al. 2018), demonstrating that gene regulatory GOE and coding GOF variants involving the same gene can cause completely different clinical phenotypes.

In summary, GOE can result from structural variants and SNVs located within promoters and distal gene regulatory elements. GOE variants often arise from the ectopic activity of enhancers (e.g., enhancer adoption) or promoters, or the disruption of normal repressive gene regulatory machinery. Furthermore, variants can cause complex gene regulatory outcomes wherein they cause GOE for one transcript, but LOE for a different one. Importantly, GOE variants often result in clinical phenotypes that markedly diverge from that of coding variants, complicating efforts to systematically identify this class of genetic variation using our current catalog of phenotypes associated with coding variants.

Concluding thoughts

In this review, we summarize the literature on gene regulatory variants that are known to cause Mendelian conditions and present a framework for categorizing these variants based on their proximate impact on gene expression patterns. We highlight that certain classes of gene regulatory variants can mimic coding LOF variants and gene duplication variants. However, gene regulatory variants can also create novel phenotypes. Specifically, the phenotypes associated with GOE and mLOE variants may markedly differ from those associated with LOE or LOF variants impacting the same gene. Consequently, extrapolating our knowledge of coding variants to the other 99% of the genome is insufficient for resolving how variants within gene regulatory elements cause Mendelian conditions. The current practice for identifying non-coding variants that cause Mendelian conditions

often relies upon the phenotypic similarity to known coding LOF phenotypes, delaying or missing the identification of non-coding genetic variants when the resulting phenotype differs substantially from coding LOF of the same target gene. A functional classification system tailored to the impact of non-coding variants can facilitate the organization of knowledge, so that novel non-coding variants are more readily identified. Additionally, this functional classification system has the potential to improve how we integrate results from regulatory element mutational scanning experiments with observed genetic variants in databases like ClinVar. Specifically, this functional framework can serve as a standardized framework to articulate the functional impact of non-coding variants in relation to different disease phenotypes.

Although it has been well established for several decades that gene regulatory variants cause numerous Mendelian conditions in a dominant, recessive, or X-linked inheritance pattern, our current catalog of disease-causing variants is overwhelmingly populated with coding variants. Specifically, whereas ClinVar contains over 150,000 pathogenic or likely pathogenic coding variants (Landrum et al. 2018), our non-systematic review of the literature identified only several hundred genetic variants known to disrupt gene regulatory elements (Fig. 2). It is possible that this imbalance accurately reflects the relative contributions of coding and gene regulatory variants to Mendelian conditions. However, it is notable that the rate of discovery of non-coding regulatory variants has only modestly increased since the transition in 2010 from family-based linkage analysis to exome sequencing as the predominant mode for gene discovery and clinical testing (Fig. 2). In contrast, the rate of discovery of pathogenic coding variants has substantially increased over the past 10 years (Landrum et al. 2018; Bamshad et al. 2019). Consequently, we hypothesize the current imbalance in the identification of pathogenic coding variants over gene

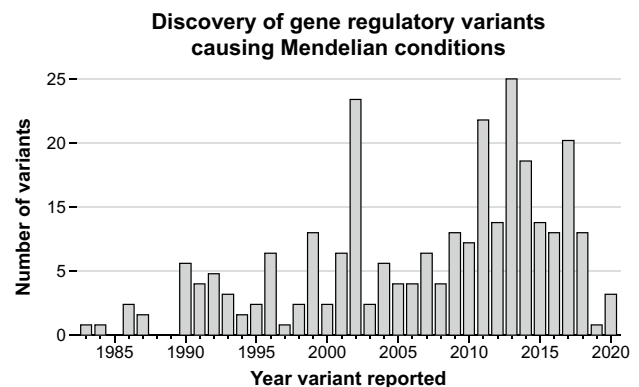


Fig. 2 Pace of discovery of gene regulatory variants causing Mendelian conditions. Histogram of the year of publication for all variants cited in this manuscript

regulatory variants more likely reflects the inadequacy of exome sequencing and current tools for analysis and interpretation to implicate this class of variation in disease. As the use of genome sequencing and epigenetic profiling becomes more common within clinical genomics, we anticipate that more examples of gene regulatory variation causing Mendelian conditions will emerge.

Acknowledgements The authors would like to thank Jessica Chong and Jacob Greene for their critical reading of this manuscript.

Author contributions All authors contributed to conceptualization of the manuscript. Y.H.H.C. and A.B.S. performed literature review. The first draft of the manuscript was written by A.B.S. and all authors provided critical revisions. All authors read and approved the final manuscript.

Funding A.B.S. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund and is a Pew Biomedical Scholar. This study was supported by National Institutes of Health (NIH) grants awarded to A.B.S. (1DP5OD029630 and OT2OD002748). S.C.B. was supported by a training grant (T32) from the NIH (T32GM007454). Y.H.H.C. is supported by NIGMS grant T32GM007266 (Univ. of WA Medical Scientist Training Program).

Data availability No new data were created or analyzed in this study.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abicht A, Stucka R, Schmidt C et al (2002) A newly identified chromosomal microdeletion and an N-box mutation of the AChRe gene cause a congenital myasthenic syndrome. *Brain* 125:1005–1013. <https://doi.org/10.1093/brain/awf095>
- Albers CA, Paul DS, Schulze H et al (2012) Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet* 44:435–439. <https://doi.org/10.1038/ng.1083>
- Allen HL, Caswell R, Xie W et al (2014) Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for DYNC1I1 exonic enhancers of DLX5/6 expression in humans. *J Med Genet* 51:264–267. <https://doi.org/10.1136/jmedgenet-2013-102142>
- Allou L, Lambert L, Amsellem D et al (2012) 14q12 and severe Rett-like phenotypes: New clinical insights and physical mapping of FOXP1-regulatory elements. *Eur J Hum Genet* 20:1216–1223. <https://doi.org/10.1038/EJHG.2012.127>
- Almeida AM, Murakami Y, Layton DM et al (2006) Hypomorphic promoter mutation in PIGM causes inherited glycosylphosphatidylinositol deficiency. *Nat Med* 12:846–851. <https://doi.org/10.1038/nm1410>
- Amano T (2020) Gene regulatory landscape of the sonic hedgehog locus in embryonic development. *Dev Growth Differ* 62:334–342. <https://doi.org/10.1111/dgd.12668>
- Bamshad MJ, Nickerson DA, Chong JX (2019) Mendelian Gene Discovery: Fast and Furious with No End in Sight. *Am J Hum Genet* 105:448–455. <https://doi.org/10.1016/j.ajhg.2019.07.011>
- Benito-Sanz S, Thomas NS, Huber C et al (2005) A novel class of pseudoautosomal region 1 deletions downstream of SHOX is associated with Léri-Weill dyschondrosteosis. *Am J Hum Genet* 77:533–544. <https://doi.org/10.1086/449313>
- Benito-Sanz S, Aza-Carmona M, Rodríguez-Estevéz A et al (2012a) Identification of the first PAR1 deletion encompassing upstream SHOX enhancers in a family with idiopathic short stature. *Eur J Hum Genet* 20:125–127. <https://doi.org/10.1038/ejhg.2011.210>
- Benito-Sanz S, Royo JL, Barroso E et al (2012b) Identification of the first recurrent PAR1 deletion in Léri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. *J Med Genet* 49:442–450. <https://doi.org/10.1136/jmedgenet-2011-100678>
- Benko S, Fantes JA, Amiel J et al (2009) Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat Genet* 41:359–364. <https://doi.org/10.1038/ng.329>
- Berry M, Grosveld F, Dillon N (1992) A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature* 358:499–502. <https://doi.org/10.1038/358499a0>
- Bhatia S, Bengani H, Fish M et al (2013) Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet* 93:1126–1134. <https://doi.org/10.1016/j.ajhg.2013.10.028>
- Biesscker LG, Spinner NB (2013) A genomic view of mosaicism and human disease. *Nat Rev Genet* 14:307–320. <https://doi.org/10.1038/nrg3424>
- Borck G, Zarhrate M, Cluzeau C et al (2006) Father-to-daughter transmission of Cornelia de Lange syndrome caused by a mutation in the 5' untranslated region of the NIPBL gene. *Hum Mutat* 27:731–735. <https://doi.org/10.1002/humu.20380>
- Cai J, Goodman BK, Patel AS et al (2003) Increased risk for developmental delay in Saethre-Chotzen syndrome is associated with TWIST deletions: An improved strategy for TWIST mutation screening. *Hum Genet* 114:68–76. <https://doi.org/10.1007/s00439-003-1012-7>
- Cappellini MD, Fiorelli G, Bernini LF (1981) Interaction between Homozygous β 0 Thalassaemia and the Swiss Type of Hereditary Persistence of Fetal Haemoglobin. *Br J Haematol* 48:561–572. <https://doi.org/10.1111/j.1365-2141.1981.tb02753.x>
- Cerrato F, Sparago A, Ariani F et al (2020) DNA Methylation in the Diagnosis of Monogenic Diseases. *Genes (basel)*. <https://doi.org/10.3390/genes11040355>
- Chen KJ, Chao HK, Hsiao KJ, Su TS (2002) Identification and characterization of a novel liver-specific enhancer of the human phenylalanine hydroxylase gene. *Hum Genet* 110:235–243. <https://doi.org/10.1007/s00439-002-0677-7>
- Coffey AJ, Brooksbank RA, Brandau O et al (1998) Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene. *Nat Genet* 20:129–135. <https://doi.org/10.1038/2424>

- Collins FS, Boehm CD, Waber PG et al (1984) Concordance of a point mutation 5' to the (G) γ globin gene with (G) $\gamma\beta$ + hereditary persistence of fetal hemoglobin in the black population. *Blood* 64:1292–1296. <https://doi.org/10.1182/blood.v64.6.1292.1292>
- Costa FF, Zago MA, Cheng G et al (1990) The Brazilian type of non-deletional (A) γ -fetal hemoglobin has a C \rightarrow G substitution at nucleotide -195 of the (A) γ -globin gene. *Blood* 76:1896–1897. <https://doi.org/10.1182/blood.v76.9.1896.1896>
- Costa JR, Caputo VS, Makarona K et al (2014) Cell-type-specific transcriptional regulation of PIGM underpins the divergent hematologic phenotype in inherited GPI deficiency. *Blood* 124:3151–3154. <https://doi.org/10.1182/blood-2014-09-598813>
- Cox JJ, Willatt L, Homfray T, Woods CG (2011) A SOX9 duplication and familial 46, XX developmental testicular disorder. *N Engl J Med* 364:91–93. <https://doi.org/10.1056/NEJMc1010311>
- Craig JE, Sheerin SM, Barnetson R, Thein SL (1993) The molecular basis of HPPFH in a British family identified by heteroduplex formation. *Br J Haematol* 84:106–110. <https://doi.org/10.1111/j.1365-2141.1993.tb03032.x>
- Crisponi L, Deiana M, Loi A et al (2001) The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat Genet* 27:159–166. <https://doi.org/10.1038/84781>
- Crossley M, Brownlee GG (1990) Disruption of a C/EBP-binding site in the factor IX promoter is associated with haemophilia B. *Nature* 345:444–446. <https://doi.org/10.1038/345444a0>
- Crossley M, Winship PR, Austen DEG et al (1990) A less severe form of haemophilia B leyden. *Nucleic Acids Res* 18:4633. <https://doi.org/10.1093/nar/18.15.4633>
- Crossley M, Ludwig M, Stowell KM et al (1992) Recovery from hemophilia B Leyden: an androgen responsive element in the factor IX promoter. *Science* 257:377–379. <https://doi.org/10.1126/science.1631558>
- Dahlqvist J, Klar J, Tiwari N et al (2010) A Single-Nucleotide Deletion in the POMP 5' UTR Causes a Transcriptional Switch and Altered Epidermal Proteasome Distribution in KLiCK Genodermatosis. *Am J Hum Genet* 86:596–603. <https://doi.org/10.1016/j.ajhg.2010.02.018>
- Dathe K, Kjaer KW, Brehm A et al (2009) Duplications Involving a Conserved Regulatory Element Downstream of BMP2 Are Associated with Brachydactyly Type A2. *Am J Hum Genet* 84:483–492. <https://doi.org/10.1016/j.ajhg.2009.03.001>
- Davidson AE, Liskova P, Evans CJ et al (2016) Autosomal-Dominant Corneal Endothelial Dystrophies CHED1 and PPCD1 Are Allelic Disorders Caused by Non-coding Mutations in the Promoter of OVOL2. *Am J Hum Genet* 98:75–89. <https://doi.org/10.1016/j.ajhg.2015.11.018>
- De Angioletti M, Lacerra G, Gaudiano C et al (2002) Epidemiology of the delta globin alleles in southern Italy shows complex molecular, genetic, and phenotypic features. *Hum Mutat* 20:358–367. <https://doi.org/10.1002/humu.10132>
- De Kok YJM, Vossenaar ER, Cremers CWRJ et al (1996) Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum Mol Genet* 5:1229–1235. <https://doi.org/10.1093/hmg/5.9.1229>
- Dedoussis GVZ, Pitsavos C, Kelberman D et al (2003) FH-Pyrgos: A novel mutation in the promoter (-45delT) of the low-density lipoprotein receptor gene associated with familial hypercholesterolemia. *Clin Genet* 64:414–419. <https://doi.org/10.1034/j.1399-0004.2003.00164.x>
- Delgado S, Velinov M (2015) 7q21.3 Deletion involving enhancer sequences within the gene DYNC111 presents with intellectual disability and split hand-split foot malformation with decreased penetrance. *Mol Cytogenet*. <https://doi.org/10.1186/s13039-015-0139-2>
- Elsas LJ, Lai K, Saunders CJ, Langley SD (2001) Functional analysis of the human galactose-1-phosphate uridylyltransferase promoter in Duarte and LA variant galactosemia. *Mol Genet Metab* 72:297–305. <https://doi.org/10.1006/mgme.2001.3157>
- Erceg J, Pakozdi T, Marco-Ferreres R et al (2017) Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. *Genes Dev* 31:590–602. <https://doi.org/10.1101/gad.292870.116>
- Fakhouri WD, Rahimov F, Attanasio C et al (2014) An etiologic regulatory mutation in IRF6 with loss- and gain-of-function effects. *Hum Mol Genet* 23:2711–2720. <https://doi.org/10.1093/hmg/ddt664>
- Fang J, Dagenais SL, Erickson RP et al (2000) Mutations in FOXC2 (MFH-1), a forkhead family transcription factor, are responsible for the hereditary lymphedema-distichiasis syndrome. *Am J Hum Genet* 67:1382–1388. <https://doi.org/10.1086/316915>
- Fantes J, Redeker B, Breen M et al (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum Mol Genet* 4:415–422. <https://doi.org/10.1093/hmg/4.3.415>
- Florian C, Bahi-Buisson N, Bienvenu T (2012) FOXG1-related disorders: From clinical description to molecular genetics. *Mol Syndromol* 2:153–163. <https://doi.org/10.1159/000327329>
- Fonseca ACS, Bonaldi A, Bertola DR et al (2013) The clinical impact of chromosomal rearrangements with breakpoints upstream of the SOX9 gene: Two novel de novo balanced translocations associated with acampomelic campomelic dysplasia. *BMC Med Genet*. <https://doi.org/10.1186/1471-2350-14-50>
- Foster JW, Dominguez-Steglich MA, Guioli S et al (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* 372:525–530. <https://doi.org/10.1038/372525a0>
- French JD, Edwards SL (2020) The Role of Noncoding Variants in Heritable Disease. *Trends Genet* 36:880–891. <https://doi.org/10.1016/j.tig.2020.07.004>
- Frischknecht H, Dutly F (2005) Two new δ -globin mutations: Hb A2-ninive [δ 133(H11)Val \rightarrow Ala] and a δ + δ -thalassemia mutation [-31 (A \rightarrow G)] in the tata box of the δ -globin gene. *Hemoglobin* 29:151–154. <https://doi.org/10.1081/HEM-200058593>
- Fuchareon S, Shimizu K, Fukumaki Y (1990) A novel C-T transition within the distal CCAAT motif of the G γ -globin gene in the Japanese HPPFH: Implication of factor binding in elevated fetal globin expression. *Nucleic Acids Res* 18:5245–5253. <https://doi.org/10.1093/nar/18.17.5245>
- Furniss D, Lettice LA, Taylor IB et al (2008) A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. *Hum Mol Genet* 17:2417–2423. <https://doi.org/10.1093/hmg/ddn141>
- Galey M, Reed P, Wenger T et al (2022) 3-hour genome sequencing and targeted analysis to rapidly assess genetic risk. <https://doi.org/10.1101/2022.09.09.22279746>
- Gallagher PG, Nilson DG, Wong C et al (2005) A dinucleotide deletion in the ankyrin promoter alters gene expression, transcription initiation and TFIID complex formation in hereditary spherocytosis. *Hum Mol Genet* 14:2501–2509. <https://doi.org/10.1093/hmg/ddi254>
- Gaszner M, Felsenfeld G (2006) Insulators: Exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 7:703–713. <https://doi.org/10.1038/nrg1925>
- Gelinas R, Bender M, Lotshaw C et al (1986) Chinese (A) γ fetal hemoglobin: C to T substitution at position -196 of the (A) γ gene

- promoter. *Blood* 67:1777–1779. <https://doi.org/10.1182/blood.v67.6.1777.bloodjournal6761777>
- Giorgio E, Robyr D, Spielmann M et al (2014) A large genomic deletion leads to enhancer adoption by the lamin B1 gene: A second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum Mol Genet* 24:3143–3154. <https://doi.org/10.1093/hmg/ddv065>
- Godart F, Bellanné-Chantelot C, Clauin S et al (2000) Identification of seven novel nucleotide variants in the hepatocyte nuclear factor-1 α (TCF1) promoter region in MODY patients. *Hum Mutat* 15:173–180
- Gotoh L, Inoue K, Helman G et al (2014) GJC2 promoter mutations causing Pelizaeus-Merzbacher-like disease. *Mol Genet Metab* 111:393–398. <https://doi.org/10.1016/j.ymgme.2013.12.001>
- Gurnett CA, Bowcock AM, Dietz FR et al (2007) Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *Am J Med Genet A* 143:27–32. <https://doi.org/10.1002/ajmg.a.31563>
- Gutiérrez L, Caballero N, Fernández-Calleja L et al (2020) Regulation of GATA1 levels in erythropoiesis. *IUBMB Life* 72:89–105. <https://doi.org/10.1002/iub.2192>
- Hardison RC, Chui DHK, Giardine B et al (2002) HbVar. A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Hum Mutat* 19:225–233. <https://doi.org/10.1002/humu.10044>
- Hatton CSR, Wilkie AOM, Drysdale HC et al (1990) α -Thalassemia caused by a large (62 kb) deletion upstream of the human α globin gene cluster. *Blood* 76:221–227. <https://doi.org/10.1182/blood.v76.1.221.bloodjournal761221>
- Hazlewood RJ, Roos BR, Solivan-Timpe F et al (2015) Heterozygous triplication of upstream regulatory sequences leads to dysregulation of matrix metalloproteinase 19 in patients with cavitory optic disc anomaly. *Hum Mutat* 36:369–378. <https://doi.org/10.1002/humu.22754>
- Henderson SJ, Timbs AT, McCarthy J et al (2016) Ten Years of Routine α - and β -Globin Gene Sequencing in UK Hemoglobinopathy Referrals Reveals 60 Novel Mutations. *Hemoglobin* 40:75–84. <https://doi.org/10.3109/03630269.2015.1113990>
- Higgs DR, Wood WG (2008) Long-range regulation of α globin gene expression during erythropoiesis. *Curr Opin Hematol* 15:176–183. <https://doi.org/10.1097/MOH.0b013e3282f734c4>
- Hill-Harfe KL, Kaplan L, Stalker HJ et al (2005) Fine mapping of chromosome 17 translocation breakpoints ≥ 900 Kb upstream of SOX9 in acampomelic campomelic dysplasia and a mild, familial skeletal dysplasia. *Am J Hum Genet* 76:663–671. <https://doi.org/10.1086/429254>
- Hitchins MP, Rapkins RW, Kwok CT et al (2011) Dominantly Inherited Constitutional Epigenetic Silencing of MLH1 in a Cancer-Affected Family Is Linked to a Single Nucleotide Variant within the 5'UTR. *Cancer Cell* 20:200–213. <https://doi.org/10.1016/j.ccr.2011.07.003>
- Horn S, Figl A, Rachakonda PS et al (2013) TERT promoter mutations in familial and sporadic melanoma. *Sci* 339:959–961. <https://doi.org/10.1126/science.1230062>
- Houlden H, Girard M, Cockerell C et al (2004) Connexin 32 promoter P2 mutations: A mechanism of peripheral nerve dysfunction. *Ann Neurol* 56:730–734. <https://doi.org/10.1002/ana.20267>
- Hsu AP, Johnson KD, Falcone EL et al (2013) GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome. *Blood* 121:3830–3837. <https://doi.org/10.1182/blood-2012-08-452763>
- Huang D, Ovcharenko I (2022) Enhancer-silencer transitions in the human genome. *Genome Res* 32:437–448. <https://doi.org/10.1101/gr.275992.121>
- Huang HJ, Stoming TA, Harris HF et al (1987) The greekkey β + β h β observed in a large black family. *Am J Hematol* 25:401–408. <https://doi.org/10.1002/ajh.2830250406>
- Huang L, Jolly LA, Willis-Owen S et al (2012) A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *Am J Hum Genet* 91:694–702. <https://doi.org/10.1016/j.ajhg.2012.08.011>
- Ilkovski B, Pagnamenta AT, O'Grady GL et al (2015) Mutations in PIGY: Expanding the phenotype of inherited glycosylphosphatidylinositol deficiencies. *Hum Mol Genet* 24:6146–6159. <https://doi.org/10.1093/hmg/ddv331>
- Ionasescu VV, Searby C, Ionasescu R et al (1996) Mutations of the noncoding region of the connexin32 gene in X-linked dominant Charcot-Marie-Tooth neuropathy. *Neurology* 47:541–544. <https://doi.org/10.1212/WNL.47.2.541>
- Jackson JF, Odom JL, Bell WN (1961) Amelioration of Sickle Cell Disease by Persistent Fetal Hemoglobin. *JAMA J Am Med Assoc* 177:867–869
- Jaiswal S, Ebert BL (2019) Clonal hematopoiesis in human aging and disease. *Science* (1979). <https://doi.org/10.1126/science.aan4673>
- Jakubiczka S, Schröder C, Ullmann R et al (2010) Translocation and deletion around SOX9 in a patient with acampomelic campomelic dysplasia and sex reversal. *Sexual Development* 4:143–149. <https://doi.org/10.1159/000302403>
- Jamieson RV, Perveen R, Kerr B et al (2002) Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma. *Hum Mol Genet* 11:33–42. <https://doi.org/10.1093/hmg/11.1.33>
- Jeong Y, Leskow FC, El-Jaick K et al (2008) Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat Genet* 40:1348–1353. <https://doi.org/10.1038/ng.230>
- Johnson KD, Hsu AP, Ryu MJ et al (2012) Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J Clin Invest* 122:3692–3704. <https://doi.org/10.1172/JCI61623>
- Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT (2008) The RNA polymerase II core promoter: the gateway to transcription. *Curr Opin Cell Biol* 20:253–259. <https://doi.org/10.1016/j.ceb.2008.03.003>
- Kaneko K, Nagasaki Y, Furukawa T et al (2001) Analysis of the human pancreatic secretory trypsin inhibitor (PSTI) gene mutations in Japanese patients with chronic pancreatitis. *J Hum Genet* 46:293–297. <https://doi.org/10.1007/s100380170082>
- Kim S, Yu NK, Kaang BK (2015) CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med*. <https://doi.org/10.1038/EMM.2015.33>
- Kiouis D, Vanin E, Delange T et al (1983) β -Globin gene inactivation by DNA translocation in $\gamma\beta$ -thalassaemi. *Nature* 306:662–666. <https://doi.org/10.1038/306662a0>
- Kortüm F, Das S, Flindt M et al (2011) The core FOXG1 syndrome phenotype consists of postnatal microcephaly, severe mental retardation, absent language, dyskinesia, and corpus callosum hypogenesis. *J Med Genet* 48:396–406. <https://doi.org/10.1136/JMG.2010.087528>
- Koyano-Nakagawa N, Gong W, Das S et al (2022) Etv2 regulates enhancer chromatin status to initiate Shh expression in the limb bud. *Nat Commun*. <https://doi.org/10.1038/s41467-022-31848-6>
- Kurth I, Klopocki E, Stricker S et al (2009) Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. *Nat Genet* 41:862–863. <https://doi.org/10.1038/ng0809-862>
- Labie D, Pagnier J, Lapoumeroulie C et al (1985) Common haplotype dependency of high (G) γ -globin gene expression and high Hb F levels in β -thalassemia and sickle cell anemia patients. *Proc Natl*

- Acad Sci U S A 82:2111–2114. <https://doi.org/10.1073/PNAS.82.7.2111>
- Landrum MJ, Lee JM, Benson M et al (2018) ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 46:D1062–D1067. <https://doi.org/10.1093/NAR/GKX1153>
- Lango Allen H, Caswell R, Xie W et al (2014) Next generation sequencing of chromosomal rearrangements in patients with split-hand/split-foot malformation provides evidence for DYNC111 exonic enhancers of DLX5/6 expression in humans. *J Med Genet* 51:264–267. <https://doi.org/10.1136/jmedgenet-2013-102142>
- Lecointre C, Pichon O, Hamel A et al (2009) Familial acampomelic form of campomelic dysplasia caused by a 960 kb deletion upstream of SOX9. *Am J Med Genet A* 149:1183–1189. <https://doi.org/10.1002/ajmg.a.32830>
- Lee JA, Madrid RE, Sperle K et al (2006) Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Ann Neurol* 59:398–403. <https://doi.org/10.1002/ana.20732>
- Leipoldt M, Erdel M, Bien-Willner GA et al (2007) Two novel translocation breakpoints upstream of SOX9 define borders of the proximal and distal breakpoint cluster region in campomelic dysplasia. *Clin Genet* 71:67–75. <https://doi.org/10.1111/j.1399-0004.2007.00736.x>
- Lettice LA, Horikoshi T, Heaney SJH et al (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* 99:7548–7553. <https://doi.org/10.1073/pnas.112212199>
- Lewis EB (1954) The theory and application of a new method of detecting chromosomal rearrangements in drosophila melanogaster. *Am Nat* 88:225–239. <https://doi.org/10.1086/281833>
- Li J, Woods SL, Healey S et al (2016) Point Mutations in Exon 1B of APC Reveal Gastric Adenocarcinoma and Proximal Polyposis of the Stomach as a Familial Adenomatous Polyposis Variant. *Am J Hum Genet* 98:830–842. <https://doi.org/10.1016/j.ajhg.2016.03.001>
- Li C, Georgakopoulou A, Mishra A et al (2021) In vivo HSPC gene therapy with base editors allows for efficient reactivation of fetal γ -globin in b-YAC mice. *Blood Adv* 5:1122–1135. <https://doi.org/10.1182/BLOODADVANCES.2020003702>
- Lifton RP, Dluhy RG, Powers M et al (1992) A chimeric 11 β -hydroxylase/aldosterone synthase gene causes glucocorticoid-remediable aldosteronism and human hypertension. *Nature* 355:262–265. <https://doi.org/10.1038/355262a0>
- Lohan S, Spielmann M, Doelken SC et al (2014) Microduplications encompassing the sonic hedgehog limb enhancer ZRS are associated with haas-type polysyndactyly and Laurin–Sandrow syndrome. *Clin Genet* 86:318–325. <https://doi.org/10.1111/cge.12352>
- Loudianos G, Lavinha PM, Galanello R et al (1992) Normal δ globin gene sequences in sardinian nondeletional δ beta;thalassemia. *Hemoglobin* 16:503–509. <https://doi.org/10.3109/03630269208993118>
- Ludlow LB, Schick BP, Budarf ML et al (1996) Identification of a mutation in a GATA binding site of the platelet glycoprotein Ib β promoter resulting in the Bernard–Soulier Syndrome. *J Biol Chem* 271:22076–22080. <https://doi.org/10.1074/jbc.271.36.22076>
- Lupiáñez DG, Kraft K, Heinrich V et al (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161:1012–1025. <https://doi.org/10.1016/j.cell.2015.04.004>
- Mackay DR, Hu M, Li B et al (2006) The mouse *Ovo12* gene is required for cranial neural tube development. *Dev Biol* 291:38–52. <https://doi.org/10.1016/j.ydbio.2005.12.003>
- Manco L, Ribeiro ML, Maximo V et al (2000) A new PKLR gene mutation in the R-type promoter region affects the gene transcription causing pyruvate kinase deficiency. *Br J Haematol* 110:993–997. <https://doi.org/10.1046/j.1365-2141.2000.02283.x>
- Mandelker D, Zhang L, Kemel Y et al (2017) Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. *JAMA J Am Med Assoc* 318:825–835. <https://doi.org/10.1001/jama.2017.11137>
- Mansour S, Hall CM, Pembrey ME, Young ID (1995) A clinical and genetic study of campomelic dysplasia. *J Med Genet* 32:415–420. <https://doi.org/10.1136/jmg.32.6.415>
- Martyn GE, Wienert B, Yang L et al (2018) Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A-binding. *Nat Genet* 50:498–503. <https://doi.org/10.1038/s41588-018-0085-0>
- Matsuda M, Sakamoto N, Fukumaki Y (1992) δ -Thalassemia caused by disruption of the site for an erythroid-specific transcription factor, GATA-1, in the δ -globin gene promoter. *Blood* 80:1347–1351. <https://doi.org/10.1182/blood.v80.5.1347.1347>
- McElreavy K, Vilain E, Abbas N et al (1992) XY sex reversal associated with a deletion 5' to the SRY “HMG box” in the testis-determining region. *Proc Natl Acad Sci U S A* 89:11016–11020. <https://doi.org/10.1073/pnas.89.22.11016>
- Meeths M, Chiang SCC, Wood SM et al (2011) Familial hemophagocytic lymphohistiocytosis type 3 (FHL3) caused by deep intronic mutation and inversion in UNC13D. *Blood* 118:5783–5793. <https://doi.org/10.1182/blood-2011-07-369090>
- Ngan CY, Wong CH, Tjong H et al (2020) Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nat Genet* 52:264–272. <https://doi.org/10.1038/s41588-020-0581-x>
- Ngcungcu T, Oti M, Sitek JC et al (2017) Duplicated Enhancer Region Increases Expression of CTSB and Segregates with Keratolytic Winter Erythema in South African and Norwegian Families. *Am J Hum Genet* 100:737–750. <https://doi.org/10.1016/j.ajhg.2017.03.012>
- Nichols P, Croxson R, Vincent A et al (1999) Mutation of the acetylcholine receptor ϵ -subunit promoter in congenital myasthenic syndrome. *Ann Neurol* 45:439–443
- Ninomiya S, Narahara K, Tsuji K et al (1995) Acampomelic campomelic syndrome and sex reversal associated with de novo t(12;17) translocation. *Am J Med Genet* 56:31–34. <https://doi.org/10.1002/ajmg.1320560109>
- Ninomiya S, Isomura M, Narahara K et al (1996) Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12;17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. *Hum Mol Genet* 5:69–72. <https://doi.org/10.1093/hmg/5.1.69>
- Ohno K, Anlar B, Engel AG (1999) Congenital myasthenic syndrome caused by a mutation in the Ets-binding site of the promoter region of the acetylcholine receptor ϵ subunit gene. *Neuromuscul Disord* 9:131–135. [https://doi.org/10.1016/S0960-8966\(99\)00007-3](https://doi.org/10.1016/S0960-8966(99)00007-3)
- Oner R, Kutlar F, Gu LH, Huisman THJ (1991) The Georgia type of nondeletional hereditary persistence of fetal hemoglobin has a C \rightarrow T mutation at nucleotide -114 of the (A) γ -globin gene [1]. *Blood* 77:1124–1125. <https://doi.org/10.1182/blood.v77.5.1124.1124>
- Otonkoski T, Jiao H, Kaminen-Ahola N et al (2007) Physical exercise-induced hypoglycemia caused by failed silencing of monocarboxylate transporter 1 in pancreatic beta cells. *Am J Hum Genet* 81:467–474. <https://doi.org/10.1086/520960>

- Panigrahi A, O'Malley BW (2021) Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* <https://doi.org/10.1186/s13059-021-02322-1>
- Pfeifer D, Kist R, Dewar K et al (1999) Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: Evidence for an extended control region. *Am J Hum Genet* 65:111–124. <https://doi.org/10.1086/302455>
- Pippucci T, Savoia A, Perrotta S et al (2011) Mutations in the 5' UTR of ANKRD26, the ankirin repeat domain 26 gene, cause an autosomal-dominant form of inherited thrombocytopenia, THC2. *Am J Hum Genet* 88:115–120. <https://doi.org/10.1016/j.ajhg.2010.12.006>
- Pirastu M, Galanello R, Doherty MA et al (1987) The same β -globin gene mutation is present on nine different β -thalassemia chromosomes in a Sardinian population. *Proc Natl Acad Sci U S A* 84:2882–2885. <https://doi.org/10.1073/pnas.84.9.2882>
- Poduri A, Evrony GD, Cai X, Walsh CA (2013) Somatic mutation, genomic variation, and neurological disease. *Science* (1979). <https://doi.org/10.1126/science.1237758>
- Poli MC, Ebstein F, Nicholas SK et al (2018) Heterozygous Truncating Variants in POMP Escape Nonsense-Mediated Decay and Cause a Unique Immune Dysregulatory Syndrome. *Am J Hum Genet* 102:1126–1142. <https://doi.org/10.1016/j.ajhg.2018.04.010>
- Pop R, Conz C, Lindenberg KS et al (2004) Screening of the 1 Mb SOX9 5' control region by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. *J Med Genet.* <https://doi.org/10.1136/jmg.2003.013185>
- Refai O, Friedman A, Terry L et al (2010) De Novo 12;17 translocation upstream of sox9 resulting in 46, xx testicular disorder of sex development. *Am J Med Genet A* 152:422–426. <https://doi.org/10.1002/ajmg.a.33201>
- Roessler E, Ward DE, Gaudenz K et al (1997) Cytogenetic rearrangements involving the loss of the Sonic Hedgehog gene at 7q36 cause holoprosencephaly. *Hum Genet* 100:172–181. <https://doi.org/10.1007/s004390050486>
- Ropero P, Erquiaga S, Arizabalaga B et al (2017) Phenotype of mutations in the promoter region of the β -globin gene. *J Clin Pathol* 70:874–878. <https://doi.org/10.1136/jclinpath-2017-204378>
- Royle G, Van de Water NS, Berry E, et al (1991) Haemophilia B Leyden arising de novo by point mutation in the putative factor IX promoter region. *Br J Haematol* 77:191–194.
- Sadiq MF, Eigel A, Horst J (2001) Spectrum of β -thalassemia in Jordan: Identification of two novel mutations. *Am J Hematol* 68:16–22. <https://doi.org/10.1002/ajh.1143>
- Sakai T, Ohtani N, McGee TL et al (1991) Oncogenic germ-line mutations in Sp1 and ATF sites in the human retinoblastoma gene. *Nature* 353:83–86. <https://doi.org/10.1038/353083a0>
- Schulert GS, Zhang M, Husami A et al (2018) Brief Report: Novel UNC13D Intronic Variant Disrupting an NF- κ B Enhancer in a Patient With Recurrent Macrophage Activation Syndrome and Systemic Juvenile Idiopathic Arthritis. *Arthritis and Rheumatology* 70:963–970. <https://doi.org/10.1002/art.40438>
- Schwarze U, Cundy T, Liu YJ et al (2019) Compound heterozygosity for a frameshift mutation and an upstream deletion that reduces expression of SERPINH1 in siblings with a moderate form of osteogenesis imperfecta. *Am J Med Genet A* 179:1466–1475. <https://doi.org/10.1002/ajmg.a.61170>
- Sellick GS, Barker KT, Stolte-Dijkstra I et al (2004) Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat Genet* 36:1301–1305. <https://doi.org/10.1038/ng1475>
- Senée V, Chelala C, Duchatelet S et al (2006) Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet* 38:682–687. <https://doi.org/10.1038/ng1802>
- Smemo S, Campos LC, Moskowitz IP et al (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum Mol Genet* 21:3255–3263. <https://doi.org/10.1093/hmg/dds165>
- Smyk M, Berg JS, Pursley A et al (2007) Male-to-female sex reversal associated with an ~250 kb deletion upstream of NR0B1 (DAX1). *Hum Genet* 122:63–70. <https://doi.org/10.1007/s00439-007-0373-8>
- Sobreira NLM, Gnanakkan V, Walsh M et al (2011) Characterization of complex chromosomal rearrangements by targeted capture and next-generation sequencing. *Genome Res* 21:1720–1727. <https://doi.org/10.1101/gr.122986.111>
- Solis C, Aizencang GI, Astrin KH et al (2001) Uroporphyrinogen III synthase erythroid promoter mutations in adjacent GATA1 and CP2 elements cause congenital erythropoietic porphyria. *J Clin Invest* 107:753–762. <https://doi.org/10.1172/JCI10642>
- Spielmann M, Mundlos S (2016) Looking beyond the genes: the role of non-coding variants in human disease. *Hum Mol Genet* 25:R157–R165. <https://doi.org/10.1093/hmg/ddw205>
- Spielmann M, Brancati F, Krawitz PM et al (2012) Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. *Am J Hum Genet* 91:629–635. <https://doi.org/10.1016/j.ajhg.2012.08.014>
- Spitz F, Montavon T, Monso-Hinard C et al (2002) A t(2;8) balanced translocation with breakpoints near the human HOXD complex causes mesomelic dysplasia and vertebral defects. *Genomics* 79:493–498. <https://doi.org/10.1006/geno.2002.6735>
- Stenson PD, Mort M, Ball EV et al (2017) The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 136:665–677. <https://doi.org/10.1007/s00439-017-1779-6>
- Stergachis AB, Haugen E, Shafer A et al (2013) Exonic transcription factor binding directs codon choice and affects protein evolution. *Sci* 342:1367–1372. <https://doi.org/10.1126/science.1243490>
- Su P, Ding H, Huang D et al (2011) A 4.6 kb genomic duplication on 20p12.2-12.3 is associated with brachydactyly type A2 in a Chinese family. *J Med Genet* 48:312–316. <https://doi.org/10.1136/jmg.2010.084814>
- Sun M, Ma F, Zeng X et al (2008) Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. *J Med Genet* 45:589–595. <https://doi.org/10.1136/jmg.2008.057646>
- Tate VE, Wood WG, Weatherall DJ (1986) The British form of hereditary persistence of fetal hemoglobin results from a single base mutation adjacent to an S1 hypersensitive site 5' to the (A) γ globin gene. *Blood* 68:1389–1393. <https://doi.org/10.1182/blood.v68.6.1389.1389>
- Thein SL (2008) Genetic modifiers of the β -haemoglobinopathies. *Br J Haematol* 141:357–366. <https://doi.org/10.1111/J.1365-2141.2008.07084.X>
- Thein SL, Menzel S, Lathrop M, Garner C (2009) Control of fetal hemoglobin: New insights emerging from genomics and clinical implications. *Hum Mol Genet.* <https://doi.org/10.1093/HMG/DDP401>
- Tommerup N, Schempp W, Meinecke P et al (1993) Assignment of an autosomal sex reversal–locus (SRA1) and campomelic dysplasia (CMPD1) to 17q24.3–q25.1. *Nat Genet* 4:170–174. <https://doi.org/10.1038/ng0693-170>
- Traxler EA, Yao Y, Wang Y-D et al (2016) A genome-editing strategy to treat β -hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat Med* 22:987–990. <https://doi.org/10.1038/nm.4170>

- Trembath DG, Semina EV, Jones DH et al (2004) Analysis of Two Translocation Breakpoints and Identification of a Negative Regulatory Element in Patients with Rieger's Syndrome. *Birth Defects Res A Clin Mol Teratol* 70:82–91. <https://doi.org/10.1002/bdra.10154>
- Turro E, Astle WJ, Megy K et al (2020) Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* 583:96–102. <https://doi.org/10.1038/s41586-020-2434-2>
- Van Wijk R, Van Solinge WW, Nerlov C et al (2003) Disruption of a novel regulatory element in the erythroid-specific promoter of the human PKLR gene causes severe pyruvate kinase deficiency. *Blood* 101:1596–1602. <https://doi.org/10.1182/blood-2002-07-2321>
- Velagaleti GVN, Bien-Willner GA, Northup JK et al (2005) Position effects due to chromosome breakpoints that map ~900 Kb upstream and ~1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am J Hum Genet* 76:652–662. <https://doi.org/10.1086/429252>
- Veltkamp JJ, Meilof J, Remmelts HG et al (1970) Another Genetic Variant of Haemophilia B: Haemophilia B Leyden. *Scand J Haematol* 7:82–90. <https://doi.org/10.1111/j.1600-0609.1970.tb01873.x>
- Vetro A, Ciccone R, Giorda R et al (2011) XX males SRY negative: A confirmed cause of infertility. *J Med Genet* 48:710–712. <https://doi.org/10.1136/jmedgenet-2011-100036>
- Vinh DC, Patel SY, Uzel G et al (2010) Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood* 115:1519–1529. <https://doi.org/10.1182/blood-2009-03-208629>
- Volkman BA, Zinkevich NS, Mustonen A et al (2011) Potential novel mechanism for Axenfeld–Rieger syndrome: Deletion of a distant region containing regulatory elements of PITX2. *Invest Ophthalmol vis Sci* 52:1450–1459. <https://doi.org/10.1167/iovs.10-6060>
- Vortkamp A, Gessler M, Grzeschik KH (1991) GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* 352:539–540. <https://doi.org/10.1038/352539a0>
- Waber PG, Bender MA, Gelinis RE et al (1986) Concordance of a point mutation 5' to the (A) γ -globin gene with (A) $\gamma\beta$ + hereditary persistence of fetal hemoglobin in Greeks. *Blood* 67:551–554. <https://doi.org/10.1182/blood.v67.2.551.bloodjournal672551>
- Wagner T, Wirth J, Meyer J et al (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* 79:1111–1120. [https://doi.org/10.1016/0092-8674\(94\)90041-8](https://doi.org/10.1016/0092-8674(94)90041-8)
- Wallis DE, Roessler E, Hehr U et al (1999) Mutations in the homeo-domain of the human SIX3 gene cause holoprosencephaly. *Nat Genet* 22:196–198. <https://doi.org/10.1038/9718>
- Weatherall DJ (2001) Phenotype-genotype relationships in monogenic disease: Lessons from the thalassaemias. *Nat Rev Genet* 2:245–255. <https://doi.org/10.1038/35066048>
- Weedon MN, Cebola I, Patch A-M et al (2014) Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat Genet* 46:61–64. <https://doi.org/10.1038/ng.2826>
- Wieczorek D, Pawlik B, Li Y et al (2010) A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner Mesomelic Syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Hum Mutat* 31:81–89. <https://doi.org/10.1002/humu.21142>
- Wieczorek D, Newman WG, Wieland T et al (2014) Compound Heterozygosity of Low-Frequency Promoter Deletions and Rare Loss-of-Function Mutations in TXNL4A Causes Burn–McKewon Syndrome. *Am J Hum Genet* 95:698–707. <https://doi.org/10.1016/j.ajhg.2014.10.014>
- Wirth J, Wagner T, Meyer J et al (1996) Translocation breakpoints in three patients with campomelic dysplasia and autosomal sex reversal map more than 130 kb from SOX9. *Hum Genet* 97:186–193. <https://doi.org/10.1007/BF02265263>
- Wu P, Zhang N, Wang X et al (2012) Family history of von Hippel–Lindau disease was uncommon in Chinese patients: Suggesting the higher frequency of de novo mutations in VHL gene in these patients. *J Hum Genet* 57:238–243. <https://doi.org/10.1038/jhg.2012.10>
- Wunderle VM, Critcher R, Hastie N et al (1998) Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc Natl Acad Sci U S A* 95:10649–10654. <https://doi.org/10.1073/pnas.95.18.10649>
- Yan H, Jin H, Xue G et al (2007) Germline hMSH2 promoter mutation in a Chinese HNPCC kindred: Evidence for dual role of LOH. *Clin Genet* 72:556–561. <https://doi.org/10.1111/j.1399-0004.2007.00911.x>
- Young ID, Zuccollo JM, Maltby EL, Broderick NJ (1992) Campomelic dysplasia associated with a de novo 2q:17q reciprocal translocation. *J Med Genet* 29:251–252. <https://doi.org/10.1136/jmg.29.4.251>
- Zertal-Zidani S, Merghoub T, Ducrocq R et al (1999) A novel C→A transversion within the distal CCAAT motif of the (G) γ -globin gene in the Algerian (G) $\gamma\beta$ + hereditary persistence of fetal hemoglobin. *Hemoglobin* 23:159–169. <https://doi.org/10.3109/03630269908996160>
- Zhang F, Lupski JR (2015) Non-coding genetic variants in human disease. *Hum Mol Genet* 24:R102–R110. <https://doi.org/10.1093/hmg/ddv259>
- Zhang F, Seeman P, Liu P et al (2010) Mechanisms for Nonrecurrent Genomic Rearrangements Associated with CMT1A or HNPP: Rare CNVs as a Cause for Missing Heritability. *Am J Hum Genet* 86:892–903. <https://doi.org/10.1016/j.ajhg.2010.05.001>
- Zhu Y, Gujar AD, Wong C-H et al (2021) Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell* 39:694–707.e7. <https://doi.org/10.1016/j.ccell.2021.03.006>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.