

Supplemental Material Table of Contents

CONSORTIUM NAME

Consortium members of Genomics England who did not directly contribute to this project are

⁵Ambrose, J. C. ; ⁵Arumugam, P. ; ⁵Baple, E. L. ; ⁵Bleda, M. ; ^{5,6}Boardman-Pretty, F.; ⁵Boissiere, J. M. ; ⁵Boustred, C. R. ; ⁵Brittain, H. ; ^{5,6}Caulfield, M. J.; ⁵Chan, G. C. ; ⁵Craig, C. E. H. ; ⁵Daugherty, L. C. ; ⁵de Burca, A. ; ⁵Devereau, A. ; ^{5,6}Elgar, G. 1,2 ; ⁵Foulger, R. E. ; ⁵Fowler, T. ; ⁵Furió-Tarí, P. ; ⁵Giess A. ; ⁵Hackett, J. M. ; ⁵Halai, D. ; ⁵Hamblin, A.; ^{5,6}Henderson, S.; ⁵Holman, J. E. ; ⁵Hubbard, T. J. P. ; ^{5,6}Ibáñez, K.; ⁵Jackson, R. ; ^{5,6}Jones, L. J.; ^{5,6}Kasperaviciute, D.; ⁵Kayikci, M. ; ⁵Kousathanas, A. ; ⁵Lahnstein, L.; ⁵Lawson, K. ; ⁵Leigh, S. E. A. ; ⁵Leong, I. U. S. ; ⁵Lopez, F. J. ; ⁵Maleady-Crowe, F. ; ⁵Mason, J. ; ^{5,6}McDonagh, E. M.; ^{5,6}Moutsianas, L.; ^{5,6}Mueller, M.; ⁵Murugaesu, N.; ^{5,6}Need, A. C.; ⁵Odhams, C. A.; ⁵Orioli A.; ^{5,6}Patch, C.; ⁵Perez-Gil, D.; ⁵Pereira, M. B.; ⁵Polychronopoulos, D.; ⁵Pullinger, J. ; ⁵Rahim, T.; ⁵Rendon, A.; ⁵Riesgo-Ferreiro, P. ; ⁵Rogers, T. ; ⁵Ryten, M. ; ⁵Savage, K. ; ⁵Sawant, K. ; ⁵Scott, R. H.; ⁵Siddiq, A. ; ⁵Sieghart, A.; ^{5,6}Smedley, D. ; ^{5,6}Smith, K. R.; ⁵Smith, S. C. ; ^{5,6}Sosinsky, A.; ⁵Spooner, W. ; ⁵Stevens, H. E.; ⁵Stuckey, A.; ⁵Sultana, R. ; ⁵Tanguy M.; ^{5,6}Thomas, E. R. A.; ⁵Thompson, S. R. ; ⁵Tregidgo, C. ; ^{5,6}Tucci, A. ; ⁵Walsh, E. ; ⁵Watters, S. A. ; ⁵Welland, M. J.; ⁵Williams, E. ; ^{5,6}Witkowska, K.; ^{5,6} Wood, S. M.; ⁵Zarowiecki, M.; ⁵Genomics England, London, UK, ⁶William Harvey Research Institute, Queen Mary University of London, London, EC1M 6BQ, UK.

Supplemental Table 1. NC1 domain locations within the six human collagen IV α -chains.

Supplemental Table 2. Splicing predictions for all exonic single-nucleotide substitutions located within 3 bases of a canonical splice site.

Supplemental Table 3. Categorisation of all possible unique collagen IV α 5 NC1 missense variants.

Supplemental Figure 1. Variant inclusion flowchart for the theoretical cohort comprising all possible collagen IV α 5 NC1 missense variants.

Supplemental Figure 2. Multiple sequence alignment of the six human collagen IV α -chains.

Supplemental Table 1. NC1 domain locations within the six human collagen IV α -chains.

Chain	Isoform	NC1 domain (residues)	NCBI transcript reference sequence	NCBI protein reference sequence
α 1	1	1439 – 1669	NM_001845.6	NP_001836.3
α 2	1	1484 – 1712	NM_001846.4	NP_001837.2
α 3	1	1437 – 1670	NM_000091.5	NP_000082.2
α 4	1	1458 – 1690	NM_000092.5	NP_000083.3
α 5	2	1461 – 1691	NM_033380.3	NP_203699.1
α 6	A	1462 – 1691	NM_001847.4	NP_001838.2

Supplemental Table 2. Splicing predictions for all exonic single-nucleotide substitutions located within 3 bases of a canonical splice site.

Gene	Nucleotide change	Expected protein change	Type	MES score (Mutant)	MES score (WT)	MES score change (%)	Predicted to affect splicing	Database
<i>COL4A5</i>	c.4528G>C	p.Gly1510Arg	Splice donor site	2.24	6.04	-62.91	Yes	ClinVar
<i>COL4A5</i>	c.4708T>C	p.Cys1570Arg	Splice acceptor site	9.72	10.91	-10.91	No	LOVD
<i>COL4A5</i>	c.4709G>T	p.Cys1570Phe	Splice acceptor site	10.19	10.91	-6.60	No	ClinVar
<i>COL4A5</i>	c.4709G>C	p.Cys1570Ser	Splice acceptor site	10.93	10.91	0.18	No	LOVD, ClinVar
<i>COL4A5</i>	c.4709G>A	p.Cys1570Tyr	Splice acceptor site	10.45	10.91	-4.22	No	LOVD
<i>COL4A5</i>	c.4819A>G	p.Met1607Val	Splice donor site	3.10	3.85	-19.48	Yes	LOVD
<i>COL4A5</i>	c.4823A>G	p.His1608Arg	Splice acceptor site	11.66	11.30	3.19	No	gnomAD
<i>COL4A5</i>	c.4993A>T	p.Ser1665Cys	Splice donor site	5.75	9.65	-40.41	Yes	LOVD

Grey rows indicate variants predicted to affect normal splicing. Known exonic splicing variants are not shown. MES, MaxEntScan; WT, Wild type.

Supplemental Table 3. Categorisation of all possible unique collagen IV $\alpha 5$ NC1 missense variants.

a) All variants

		Residue:		
		Conserved ^a	Not conserved	Total
Variant:	Any structural damage	196	78	274
	No structural damage	404	638	1042
	Total	600	716	1316

p<0.001, OR (95% CI)=3.96 (2.94, 5.38)

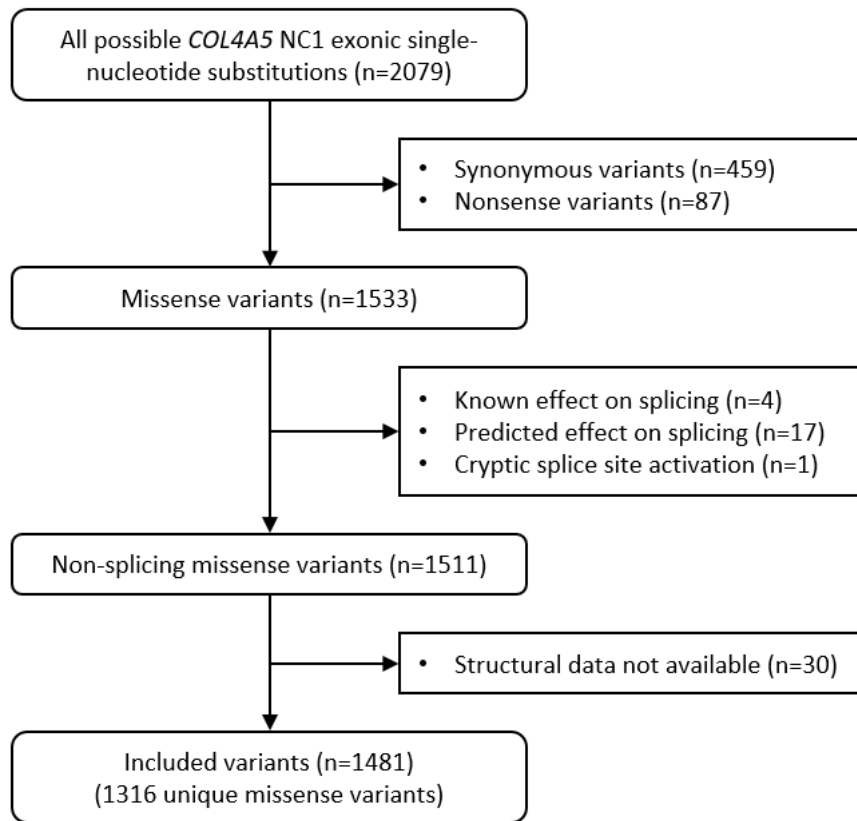
b) Excluding Cys variants

		Residue:		
		Conserved ^a	Not conserved	Total
Variant:	Any structural damage	124	78	202
	No structural damage	404	638	1042
	Total	528	716	1244

p<0.001, OR (95% CI)=2.51 (1.82, 3.47)

OR (95% CI), Odds ratio (95% confidence interval).

^a Residues fully conserved in all six human collagen IV α -chains.



Supplemental Figure 1. Variant inclusion flowchart for the theoretical cohort comprising all possible collagen IV $\alpha 5$ NC1 missense variants. This was used to calculate the expected frequencies of structural changes for variants reported in LOVD.

α1	--TPSVDHGFLVTRHSQTIDDPQCPSGTKILYHGYSLLVYQGNERRAHGQDLGTAGSCLRK	1496
α3	SPATWTTRGFVFTRHSQTTAIPSCPEGTVPLYSGFSFLFVQGNQRAHGQDLGTLGSCLR	1496
α5	--TSSVAHGFLITRHSQTTDAPQCPQGTQVYEGFSLLVYQGNKRAHGQDLGTAGSCLRR	1518
α2	---RSVSIYLLVKHSQTDQEPMPVGMNKLWSGYSLLYFEGQEKAHNQDLGLAGSCLAR	1540
α4	-FGPGYLGGLLVLHSQTDQEPTCLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPV	1516
α6	---QSMRVGYTLVKHSQSEQVPPCPIGMSQLWVGYSLLFVEGQEKAHNQDLGFAGSCLPR	1518
	*: .. ***: * * * * :: *: *: *: *: *: *: *: *: *: *: *	
α1	FSTMPFLFCNINNVCNFASTRNDYSYWLSTPEPMPMSMAPITGENIRPFISRCVCEAPAM	1556
α3	FTTMPFLFCNNDVCNFASTRNDYSYWLSTPALMPMNMAPITGRALPEYISRCTVCEGPAI	1556
α5	FSTMPFMFCNINNVCNFASTRNDYSYWLSTPEPMPMSMQPLKGQSIPFISRCVCEAPAV	1578
α2	FSTMPFLYCNPGDVCYYASRNDKSYWLSTTAPLPM--MPVAEDEIKPYISRCSVCEAPAI	1598
α4	FSTLPFAYCNIHQVCHYAQRNDRSYWLASAAPLPM--MPLSEEAIRPYVSRCAVCEAPAQ	1574
α6	FSTMPFIYCNINEVCHYARRNDKSYWLSTTAPIPM--MPVSTQTIQPYISRCSVCEAPSQ	1576
	*: *	
α1	VMAVHSQTIQIPPCPSGWSSLWIGYSFVMHTSAGAEGSGQALASPGSCLEEFRSAPFIEC	1616
α3	AIHAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTGQALASPGSCLEEFRASPFLEC	1616
α5	VIAVHSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSGQALASPGSCLEEFRSAPFIEC	1638
α2	AIHAVHSQDVSIIPHCPAGWRSWLWIGYSFLMHTAAGDEGGGQSLVSPGSCLEDFRATPFIEC	1658
α4	AVAVHSQDQSIPPCPQTWRSWLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLEC	1634
α6	AIHAVHSQDITIPQCPLGWRSWLWIGYSFLMHTAAGAEGGGQSLVSPGSCLEDFRATPFIEC	1636
	: *: ***** ** * * * * *: *: *: *: *: *: *: *: *: *: *: *: *	
α1	HG-RGTCNYYANAYSFWLATIERSEMF-KKPTPSTLKAGEL-RTHVSRQVCMRRT-	1669
α3	HG-RGTCNYYNSYSFWLASLNPERMF-RKPIPTSVKAGEL-EKIISRCQVCMKKRH	1670
α5	HG-RGTCNYYANSYSFWLATVDVSDMF-SKPQSETLKAGDL-RTRISRCQVCMKRT-	1691
α2	NGGRGTCHYYANKYSFWLTTIPEQ-SFQGSPADTLKAGLI-RTHISRCQVCMKNL-	1712
α4	QGRQGTCHFFANKYSFWLTTVKADLQFSSAPAPDTLKESQAQRKISRQVQVCKYS-	1690
α6	SGARGTCHYFANKYSFWLTTVEERQQFGLPVSETLKAGQL-HTRVSRQVCMKSL-	1691
	*: *****: *: * *: *: *: *: *: *: *: *: *: *: *: *: *: *: *	

Supplemental Figure 2. Multiple sequence alignment of the six human collagen IV α -chains. Chains are grouped into $\alpha 1$ -like chains ($\alpha 1$, $\alpha 3$, $\alpha 5$; above) and $\alpha 2$ -like chains ($\alpha 2$, $\alpha 4$, $\alpha 6$; below). Positions with a single, fully conserved residue are indicated with an asterisk (*). Alignment was performed using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).