GeneReviews Educational Materials: Comprehensive Genome Sequencing and Multi-Gene Panels

Supplementary Material: Table 3. Genes and Related Disorders with Highly Homologous Gene Family Members or a Pseudogene(s)

Posted: 14 March 2017

Gene families are genes of similar sequence and function that arose through duplication of an ancestral gene.

A pseudogene is a sequence of DNA that has some homology with a coding gene. Although most pseudogenes have the same structural elements (promotors, splice sites, and introns) found in coding genes, they do not encode proteins as they are often disrupted by multiple pathogenic variants. Processed pseudogenes are mRNA sequences copied and inserted into the genome and do not contain promotors or introns. The human genome contains approximately 20,000 pseudogenes.

The presence of non-unique sequence within the genome interferes with molecular genetic testing for many genetic disorders. The genes listed in Table 3 are common examples of genes with non-unique sequence and their associated disorders. Target enrichment (by PCR amplification or pull-down methods) can simultaneously amplify or capture sequence from a gene and other homologous regions. In addition, presence of homologous next-generation sequence reads can lead to loss of data (reads mapping to more than one location are discarded), false negative, or false positive results. In some cases, the length and degree of homology do not interfere with sequence analysis.

Although specific assays have been developed to distinguish the sequence of medically important genes from homologous sequences, these complex techniques are not easily implemented across the entire exome. Therefore, laboratories performing exome sequencing often exclude the analysis of highly homologous exons to avoid errors.

Many additional homologous sequences that may or may not (depending on specific assay design) interfere with sequence analysis are known. Assay performance must be assessed by laboratories performing testing.

Table 3. Genes and Related Disorders with Highly Homologous Gene Family Members or a Pseudogene(s)

Gene [1]	Homologous Genes and Pseudogenes	Disorder Caused by Pathogenic Variants in this Gene [2]	Proportion of Affected Individuals with Pathogenic Variants in this Gene
ABCC6 [3]	*ABCC6P1* *ABCC6P2*	Pseudoxanthoma elasticum	100%
ABCD1	≥5 pseudogenes	X-linked adrenoleukodystrophy	100%
ACTB	≥18 pseudogenes	Baraitser-Winter syndrome cerebrofrontofacial syndrome	>60%
AK2	*AK2P1* *AK2P2*	Reticular dysgenesis (see X-Linked Severe Combined Immunodeficiency)	100%

Gene [1]	Homologous Genes and Pseudogenes	Disorder Caused by Pathogenic Variants in this Gene [2]	Proportion of Affected Individuals with Pathogenic Variants in this Gene
ALG1	≥8 pseudogenes	*ALG1*-CDG (CDG-type Ik) (see Congenital Disorders of N-linked Glycosylation Pathway Overview)	100%
ANKRD11	*LOC100419906* *LOC100287912*	KBG syndrome (OMIM 148050)	100%
ASL	*ASLP1*	Argininosuccinate lyase deficiency	100%
ASS1	≥14 pseudogenes	Citrullinemia type I	100%
BANF1	≥5 pseudogenes	Nestor-Guillermo progeria syndrome (OMIM 614008)	100%
BMPR1A	*BMPR1APS1* *BMPR1APS2*	Juvenile polyposis syndrome	20%-25%
C4A	*CYP21A*	C4A deficiency (OMIM 614380)	100%
C4B	*CYP21A*	C4B deficiency (OMIM 614379)	100%
CA5A	*CA5AP1*	Carbonic anhydrase VA deficiency	100%
CDC42	≥9 pseudogenes	Takenouchi-Kosaki syndrome (OMIM 616737)	2 individuals
CEL	*CELP*	Diabetes-pancreatic exocrine dysfunction syndrome (OMIM 609812)	2 families
CFTR [4]	*CFTRP1* *CFTRP3*	*CFTR*-related disorders	100%
CHEK2	≥5 pseudogenes	Breast cancer susceptibility (OMIM 114480) Prostate cancer susceptibility (OMIM 176807)	<5%
CORO1A	*LOC606724*	Immunodeficiency 8 (OMIM 615401)	100%
CRYBB2 [5]	*CRYBB2P1*	Cataract 3, multiple types (OMIM 601547)	100%
CYCS	≥37 pseudogenes	Thrombocytopenia, AD, nonsyndromic, type IV (OMIM 612004)	100%
CYP21A2 [6]	*CYP21A1P*	21-hydroxylase-deficient congenital adrenal hyperplasia	100%
DCLRE1C [7]	*DCLRE1CP1*	Omenn syndrome (OMIM 603554)	>70 individuals
		SCID Athabaskan (OMIM 602450)	100%
DDX11	≥6 pseudogenes	Warsaw breakage syndrome (OMIM 613398)	100%
DHFR	≥4 pseudogenes	Megaloblastic anemia due to dihydrofolate reductase deficiency (OMIM 613839)	100%
DIS3L2	*DIS3L2P1*	Perlman syndrome (OMIM 267000)	100%
DPY19L2	≥5 pseudogenes	Globozoospermia, spermatogenic failure (OMIM 613958)	100%
EIF4E	≥5 pseudogenes	Autism (OMIM 615091)	3 families
FANCD2	*FANCD2P1*	Fanconi anemia	~3%
FLNC	*LOC392787*	Hypertrophic cardiomyopathy (OMIM 102565)	8 families
		Restrictive cardiomyopathy (OMIM 617047)	2 families
		Distal ABD-filaminopathy	100%
		Myofibrillar myopathy	3%

Gene [1]	Homologous Genes and Pseudogenes	Disorder Caused by Pathogenic Variants in this Gene [2]	Proportion of Affected Individuals with Pathogenic Variants in this Gene
GBA [8]	*GBAP*	Gaucher disease	100%
GCSH	≥8 pseudogenes	Glycine encephalopathy	<1%
GJA1	*GJA6P* *GJA1P1*	Oculodentodigital dysplasia (OMIM 164200)	100%
		Craniometaphyseal dysplasia, AR (OMIM 218400)	100%
		Palmoplantar keratoderma and congenital alopecia type 1 (OMIM 104100)	100%
		Hypoplastic left heart syndrome (OMIM 241550)	8 individuals
GK	*GK3P* *GK6P* *GK4P*	Glycerol kinase deficiency (OMIM 307030)	100%
GLDC [9]	*GLDCP1*	Glycine encephalopathy	70%-75%
GLUD1	≥6 pseudogenes	Familial hyperinsulinism	5%
GNAQ	*GNAQP1*	Sturge-Weber syndrome (OMIM 185300)	88%
HBA1 [10] *HBA2*	*HBAP1* *HBZP*	Alpha-thalassemia	100%
HBB	*HBBP1*	Sickle cell disease Beta-thalassemia	100%
HCN4	≥2 pseudogenes	Brugada syndrome	<1%
HPS1 [11]	*LOC100500719*	Hermansky-Pudlak syndrome	75% of Puerto Ricans 44% of non-Puerto Ricans
HSPD1	≥23 pseudogenes	Hypomyelinating leukodystrophy (OMIM 612233)	2 individuals
		Hereditary spastic paraplegia (OMIM 605280)	1 family
HYDIN	*HYDIN2*	Primary ciliary dyskinesia	Founder variant in Faroe Islands; 1 additional family
IDS [12]	*IDSP1*	Mucopolysaccharidosis Type II	100%
IFT122	*LOC653712*	Cranioectodermal dysplasia	~10%
IKBKG	*IKBKGP1*	Incontinentia pigmenti	~75%
KAL1	*KALP*	Kallmann syndrome (see Isolated Gonadotropin-Releasing Hormone (GnRH) Deficiency)	5%-10%
KRT16	≥5 pseudogenes	Pachyonychia congenita	29%
		Palmoplantar keratoderma, nonepidermolytic (focal) (OMIM 613000)	≥5 families
KRT6A	≥4 pseudogenes	Pachyonychia congenita	42%
KRT86	*KRT87P* *KRT88P*	Monilethrix (OMIM 158000)	≥4 families
LEFTY2	*LEFTY3*	Left-right axis malformations (OMIM 601877)	2 individuals
MATR3	*LOC100499497* *LOC100499496* *LOC401957*	Amyotrophic lateral sclerosis type 21 (OMIM 606070)	5 families
NCF1	*NCF1B* *NCF1C*	Chronic granulomatous disease	20%
NF1	≥11 pseudogenes	Neurofibromatosis type 1	100%

Gene [1]	Homologous Genes and Pseudogenes	Disorder Caused by Pathogenic Variants in this Gene [2]	Proportion of Affected Individuals with Pathogenic Variants in this Gene
NLRP7	*LOC100421039*	Recurrent hydatidiform mole (OMIM 231090)	~75%
NOTCH2	*NOTCH2P1*	Alagile syndrome	1%-2%
		Hajdu-Cheney syndrome (OMIM 102500)	100%
OCLN	*LOC647859*	Band-like calcification with simplified gyration and polymicrogyria (OMIM 251290)	100%
OPHN1	*ARHGAP42P3*	X-linked intellectual disability with cerebellar hypoplasia and distinctive facial appearance (OMIM 300486)	8 families
OTOA	*LOC653786*	Deafness, AR, type 22 (OMIM 607038)	100%
PHKA1	*PHKA1P1*	Phosphorylase kinase deficiency	~17%
PIK3CA	*LOC100422375*	*PIK3CA*-related segmental overgrowth	100%
PKD1	≥7 pseudogenes	Polycystic kidney disease, AD	85%
PLEKHM1	*PLEKHM1P1*	Osteopetrosis, AR type 6 (OMIM 611497)	100%
PMM2	*PMM2P1*	PMM2-CDG (*CDG-Ia*)	100%
PMS2 [13]	≥14 pseudogenes	Lynch syndrome	<5%
PRODH	*LOC440792*	Hyperprolinemia, type 1 (OMIM 239500)	100%
PROS1	*PROS2P*	Thrombophilia due to protein S deficiency (OMIM 176880)	100%
PRSS1 [14]	*PRSS3P2* *PRSS3P1*	Hereditary pancreatitis (see Pancreatitis Overview)	60%-100%
PTEN	*PTENP1*	*PTEN* hamartoma tumor syndrome	100%
RBM8A	*RBM8B*	Thrombocytopenia absent radius syndrome	100%
RPS17	≥16 pseudogenes	Diamond-Blackfan anemia	~1%
RPS19	≥7 pseudogenes	Diamond-Blackfan anemia	25%
SALL1	*SALL1P1*	Townes-Brocks syndrome	75%
SBDS [15]	*SBDSP*	Shwachman-Diamond syndrome	100%
SDHA	≥4 pseudogenes	Hereditary paraganglioma-pheochromocytoma syndromes	1%-3%
SDHC	≥5 pseudogenes	Hereditary paraganglioma-pheochromocytoma syndromes	4%-8%
SDHD	≥7 pseudogenes	Hereditary paraganglioma-pheochromocytoma syndromes	~30%
SFTPA2	*SFTPA3P*	Idiopathic pulmonary fibrosis (OMIM 178500)	2 families
SLC25A15	≥5 pseudogenes	Hyperornithinemia-hyperammonemia-homocitrullinuria syndrome	100%
SLC6A8	*SCL6A10P* *SCL6A10PB*	Creatine deficiency syndromes	56%
SMAD4 [16]	≥1 pseudogene (not named)	Juvenile polyposis syndrome ± Hereditary hemorrhagic telangiectasia	20%
SMN1 *SMN2*	*SMNP* *LOC100132090*	Spinal Muscular Atrophy	100%

Gene [1]	Homologous Genes and Pseudogenes	Disorder Caused by Pathogenic Variants in this Gene [2]	Proportion of Affected Individuals with Pathogenic Variants in this Gene
STRC [17]	*STRCP1*	Non-syndrome Deafness, AR (see Deafness and Hereditary Hearing Loss Overview)	6%-11%
TARDBP	*LOC643387* *TARDBPP1* *TARDBPP2*	TARDP-related amyotrophiclateral sclerosis	100%
		Frontotemporal lobar degeneration (OMIM 612069)	1 individual
TBX20	*LOC100418730*	ASD type 4 (OMIM 611363)	3 families
TDGF1	≥7 pseudogenes	Holoprosencephaly	2 individuals
TIMM8A	*TIMM8AP1*	Deafness-dystonia-optic neuronopathy syndrome	13 individuals
TMEM231	*LOC100420067*	Joubert syndrome and related disorders	2 families and 2 individuals
TNXB	*TNXA*	Ehlers-Danlos syndrome, hypermobility type	~3%
TPI1	≥4 pseudogenes	Hemolytic anemia due to triosephosphate isomerase deficiency (OMIM 615512)	100%
TUBA1A	*TUBA3GP* *LOC100129818*	Lissencephaly and other complex cortical malformations (see Tubulinopathies Overview)	37% of classic lissencephaly
TUBB2B	*TUBB8P4* *TUBB4BP2* *TUBB2BP1*	Polymicrogyria-like cortical dysplasia (see Tubulinopathies Overview)	87.5%
TYR	*TYRL*	Oculocutaneous albinism, type I	100%
UBE3A	*UBE3AP2* *UBE3AP1*	Angelman syndrome	~11%
VWF	*VWFP1*	von Willebrand disease	100%

AD = autosomal dominant
AR = autosomal recessive
ASD = atrial septal defect

1. Included in this table are genes that have: (1) one or more identified pseudogenes; and (2) pathogenic variants identified in more than one individual or family. Genes from the same gene family are listed together.
2. For more information see hyperlinked *GeneReview*. An OMIM phenotype entry is provided if a *GeneReview* is not available.
3. Two pseudogenes are almost identical to *ABCC6* [Pfendner et al 2008].
4. Duplication of exon 10 [Rozmahel et al 1997]
5. Gene conversion between *CRYBB2* and *CRYBB2P1* was reported by Vanita et al [2001].
6. Testing the proband and parents may be required to clarify results [Hong et al 2015].
7. The most common *DCLRE1C* pathogenic variant is a deletion resulting from homologous recombination of *DCLRE1C* and the pseudogene [Pannicke et al 2010].
8. *GBAP* is 96% homologous to *GBA* [Basgalupp et al 2016].
9. The processed pseudogene has 97.5% homology to the coding sequence of *GLDC* [Takayanagi et al 2000].
10. *HBA1* and *HBA2* have identical coding regions. This gene family also includes the embryonically expressed *HBZ*, *HBD,* and *HBQ1.*
11. The pseudogene has 95% homology to *HPS1*; exon 6 is identical [Huizing et al 2000].
12. 9% of pathogenic variants are complex rearrangements with the pseudogene. The pseudogene is 96% homologous to *IDS* [Bondeson et al 1995].
13. Ongoing evolutionary sequence exchange between *PMS2* and one pseudogene (*PMS2CL)* has led to unreliable reference sequences and false positive and false negative results on sequencing [Hayward et al 2007] (see also Vaughn et al [2011]).
14. Regardless of the sequencing method employed, primers must be carefully chosen and validated to amplify the fragment for the correct gene and transcript. Thus a multi-step method is required to verify the presence of a pathogenic variant in *PRSS1* [Masson et al 2008].
15. Interpretations may be difficult as the extent of variation in *SBDSP* is not known. *SBDSP* is 97% homologous to *SBDS*.
16. The presence of a processed pseudogene led to false positive MLPA results in some individuals [Millson et al 2015].

17. *STRCP1* is 99.6% homologous to *STRC* [Vona et al 2015].

References

Basgalupp SP, Siebert M, Vairo FP, Chami AM, Pinto LL, Carvalho GD, Schwartz IV. Use of a multiplex ligation-dependent probe amplification method for the detection of deletions/duplications in the GBA1 gene in Gaucher disease patients. Blood Cells Mol Dis. 2016 Oct 20. [Epub ahead of print]

Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tönnesen T, Carlberg BM, Pettersson U. Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. Hum Mol Genet. 1995;4:615-21.

Hayward BE, De Vos M, Valleley EM, Charlton RS, Taylor GR, Sheridan E, Bonthron DT. Extensive gene conversion at the PMS2 DNA mismatch repair locus. Hum Mutat. 2007;28:424-30.

Hong G, Park HD, Choi R, Jin DK, Kim JH, Ki CS, Lee SY, Song J, Kim JW. CYP21A2 mutation analysis in Korean patients with congenital adrenal hyperplasia using complementary methods: sequencing after long-range PCR and restriction fragment length polymorphism analysis with multiple ligation-dependent probe amplification assay. Ann Lab Med. 2015;35:535-9.

Huizing M, Anikster Y, Gahl WA. Characterization of a partial pseudogene homologous to the Hermansky-Pudlak syndrome gene HPS-1; relevance for mutation detection. Hum Genet. 2000;106:370-3.

Masson E, Le Maréchal C, Delcenserie R, Chen JM, Férec C. Hereditary pancreatitis caused by a double gain-of-function trypsinogen mutation. Hum Genet. 2008;123:521-9.

Millson A, Lewis T, Pesaran T, Salvador D, Gillespie K, Gau CL, Pont-Kingdon G, Lyon E, Bayrak-Toydemir P. Processed pseudogene confounding deletion/duplication assays for SMAD4. J Mol Diagn. 2015;17:576-82.

Pannicke U, Hönig M, Schulze I, Rohr J, Heinz GA, Braun S, Janz I, Rump EM, Seidel MG, Matthes-Martin S, Soerensen J, Greil J, Stachel DK, Belohradsky BH, Albert MH, Schulz A, Ehl S, Friedrich W, Schwarz K. The most frequent DCLRE1C (ARTEMIS) mutations are based on homologous recombination events. Hum Mutat. 2010;31:197-207.

Pfendner EG, Uitto J, Gerard GF, Terry SF. Pseudoxanthoma elasticum: genetic diagnostic markers. Expert Opin Med Diagn. 2008;2:63-79.

Rozmahel R, Heng HH, Duncan AM, Shi XM, Rommens JM, Tsui LC. Amplification of CFTR exon 9 sequences to multiple locations in the human genome. Genomics. 1997;45:554-61.

Takayanagi M, Kure S, Sakata Y, Kurihara Y, Ohya Y, Kajita M, Tada K, Matsubara Y, Narisawa K. Human glycine decarboxylase gene (GLDC) and its highly conserved processed pseudogene (psiGLDC): their structure and expression, and the identification of a large deletion in a family with nonketotic hyperglycinemia. Hum Genet. 2000;106:298-305.

Vanita, Singh JR, Sarhadi VK, Singh D, Reis A, Rueschendorf F, Becker-Follmann J, Jung M, Sperling K. A novel form of "central pouchlike" cataract, with sutural opacities, maps to chromosome 15q21-22. Am J Hum Genet. 2001;68:509-14.

Vaughn CP, Hart KJ, Samowitz WS, Swensen JJ. Avoidance of pseudogene interference in the detection of 3' deletions in PMS2. Hum Mutat. 2011;32:1063-71.

Vona B, Hofrichter MA, Neuner C, Schröder J, Gehrig A, Hennermann JB, Kraus F, Shehata-Dieler W, Klopocki E, Nanda I, Haaf T. DFNB16 is a frequent cause of congenital hearing impairment: implementation of STRC mutation analysis in routine diagnostics. Clin Genet. 2015;87:49-55.

Suggested Reading

Brodehl A, Ferrier RA, Hamilton SJ, Greenway SC, Brundler MA, Yu W, Gibson WT, McKinnon ML, McGillivray B, Alvarez N, Giuffre M, Schwartzentruber J; FORGE Canada Consortium, Gerull B. Mutations in FLNC are associated with familial restrictive cardiomyopathy. Hum Mutat. 2016;37:269-79.

Mandelker D. Towards comprehensive clinical exome sequencing: Inclusive testing of disease genes with high homology. 2014. Available online.

Valdés-Mas R, Gutiérrez-Fernández A, Gómez J, Coto E, Astudillo A, Puente DA, Reguero JR, Álvarez V, Morís C, León D, Martín M, Puente XS, López-Otín C. Mutations in filamin C cause a new form of familial hypertrophic cardiomyopathy. Nat Commun. 2014;5:5326.