PHENCODE: LINKING HUMAN MUTATIONS AND PHENOTYPE

Belinda Giardine[1], Riemer C.[1], Kent W.J.[2], Miller W.[1], Hardison R.C.[1]
[1]Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, PA;
[2]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA

URLS
Browser: genome.ucsc.edu
Documentation: www.bx.psu.edu/phencode

Sources

Cystic Fibrosis Mutation Database · HbVar · IDbases · IRSA · PAHdb · UniProt · BGMUT · ARdb · LMDp · srd5a2

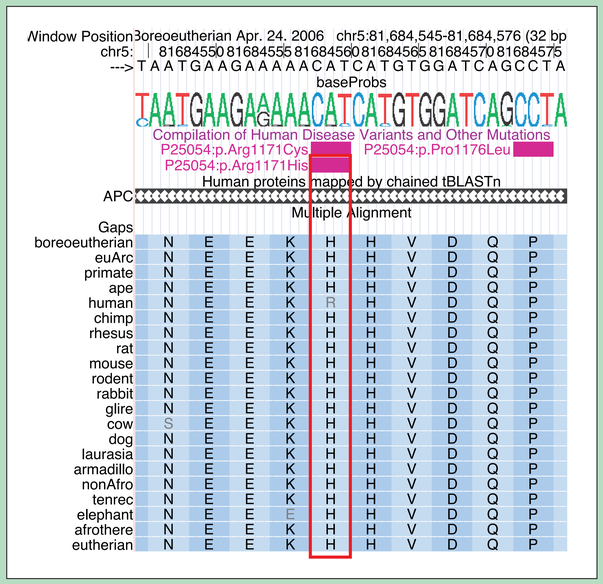
↓

UCSC Genome Browser
Locus Variants track

What is PhenCode?

Human mutations are stored in Locus Specific Databases (LSDBs) all over the world, as well as several genome wide collections such as Swiss-Prot variants. The mutations and phenotypes from cooperating databases are mapped onto genome coordinates and displayed at the UCSC Genome Browser in a new Locus Variants track. Swiss-Prot has 24,065 variants in 5,806 RefSeq genes, which averages out to 4 variants per gene. The LSDBs included so far have 10,915 variants in 360 genes, which is 30 variants per gene on average. The track also provides links back to the source database for more in-depth data.

Example 1: There are 25,590 human amino acid substitutions recorded in the Locus Variants track. Of these, 776 variants have a mutant allele matching the wild-type sequence in chimp, rhesus, the human-chimp ancestor, and the human-rhesus ancestor. Some of these variants are linked to disease in humans. In the example below, one of the human mutations is identical to the wild-type in many species.



Example 1, part 2:
Clicking on a mutation in the track takes you to a page with more details. This variant (His replacing Arg) is associated with gastric cancer in humans, but His is normal in other animals. Could His at this position be advantageous to animals in the wild but increase risk of cancer in contemporary human lifestyles?

Common name:
VAR_005043
External links:
UniProt variant - VAR_005043
Disease association:
phenotype-associated
Phenotype common name:
Gastric cancer
Variation and Disease information related to gene locus:
UniProt disease - Defects in APC are a cause of familial adenomatous polyposis (FAP) [MIM:175100]; which includes also Gardner syndrome (GS). FAP and GS contribute to tumor development in patients with uninherited forms of colorectal cancer. FAP is characterized by adenomatous polyps of the colon and rectum, but also of upper gastrointestinal tract (ampullary, duodenal and gastric adenomas). This is a viciously premalignant disease with one or more polyps progressing through dysplasia to malignancy in untreated gene carriers with a median age at diagnosis of 40 years.
UniProt disease - APC mutations have led to some interesting observations. (1) the great majority of the mutations found to date would result in truncation of the APC product. (2) almost all the mutations have occurred within the first half of the coding sequence, and somatic mutations in colorectal tumors are further clustered in a particular region, called MCR (mutation cluster region). (3) most identified point mutations in the APC gene are transitions from cytosine to other nucleotides. (4) the location of germ-line mutations tends to correlate with the number of colorectal polyps in FAP patients. Inactivation of both alleles of the APC gene seems to be required as an early event to develop most adenomas and carcinomas in the colon and rectum as well as some of those in the stomach.
UniProt disease - Defects in APC are a cause of hereditary desmoid disease (HDD) [MIM:135290]; also called familial infiltrative fibromatosis (FIF). It is an autosomal dominant trait with 100% penetrance and possible variable expression among affected relatives. HDD patients show multifocal fibromatosis of the paraspinal muscles, breast, occiput, arms, lower ribs, abdominal wall, and mesentery. Desmoid tumors appears also as a complication of familial adenomatous polyposis.
UniProt disease - Defects in APC are a cause of medulloblastoma (MDB) [MIM:155255]. MDB is a malignant, invasive embryonal tumor of the cerebellum with a preferential manifestation in children. Although the majority of medulloblastomas occur sporadically, some manifest within familial cancer syndromes such as Turcot syndrome and basal cell nevus syndrome (Gorlin syndrome).
UniProt disease - Defects in APC are a cause of Turcot syndrome [MIM:276300]. Turcot syndrome is an autosomal dominant disorder characterized by malignant tumors of the brain associated with multiple colorectal adenomas. Skin features include sebaceous cysts, hyperpigmented and cafe au lait spots.
OMIM title - DESMOID DISEASE, HEREDITARY - 135290
OMIM title - MEDULLOBLASTOMA - 155255
OMIM title - ADENOMATOUS POLYPOSIS OF THE COLON - 175100
OMIM title - TURCOT SYNDROME - 276300
Comments:
in gastric cancer

Example 1, part 3:
Following the "OMIM title" link gives more data and references for further reading.

+175100 GeneTests, Links
ADENOMATOUS POLYPOSIS OF THE COLON; APC

Alternative titles; symbols

FAMILIAL POLYPOSIS OF THE COLON; FPC
POLYPOSIS, ADENOMATOUS INTESTINAL
FAMILIAL ADENOMATOUS POLYPOSIS; FAP
APC GENE, INCLUDED
GARDNER SYNDROME, INCLUDED; GS, INCLUDED
ADENOMATOUS POLYPOSIS COLI, ATTENUATED, INCLUDED; AAPC, INCLUDED
DELETED IN POLYPOSIS 2.5, INCLUDED; DP2.5, INCLUDED

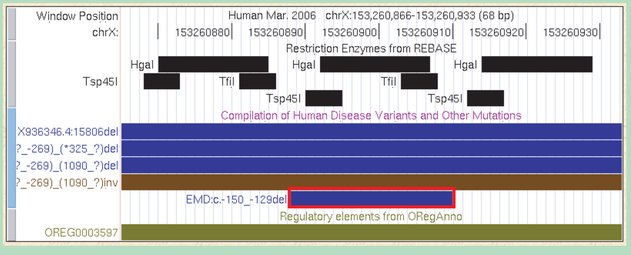
Gene map locus 5q21-q22

TEXT

DESCRIPTION

Familial adenomatous polyposis (FAP) is an autosomal dominant disorder which typically presents with colorectal cancer (CRC; 114500) in early adult life secondary to extensive adenomatous polyps of the colon. Polyps also develop in the upper gastrointestinal tract and malignancies may occur in other sites including the brain and the thyroid. Helpful diagnostic features include pigmented retinal lesions known as congenital hypertrophy of the retinal pigment, jaw cysts, sebaceous cysts, and osteomata. The APC gene at chromosome 5q21 is mutant in FAP.

Example 2: The UCSC Table Browser was used to find 710 variants that are upstream of UCSC genes. Of these, 101 intersect regions involved in regulation of gene expression, as recorded in the ORegAnno track (Ref. Montgomery SB, et al., ORegAnno: an open access database and curation system for literature-derived promoters, transcription factor binding sites and regulatory variation. Bioinformatics. 2006 Mar 1; 22(5):637-40.). This example shows a 22 base-pair deletion that takes out a portion of an ORegAnno regulatory element including 3 restriction sites.



Example 2, part 2:
This time clicking on the mutation in the track brings up a link to a LSDB, Leiden Muscular Dystrophy pages.

source: LSDB; Leiden Muscular Dystrophy pages
location: not within known transcription unit
type: deletion

Common name:
c.-150_-129del
External links:
LOVD - EMD_00001

Example 2, part 3:
At LMDp you find a summary from 2 references with patient data. EDMD stands for Emery-Dreifuss muscular dystrophy. Following the second Reference link takes you to the original paper where you learn about a young patient who is very ill. The grandmother, mother, and 2 aunts have the same mutation but are not ill. (Ref. Manilal S, et al., Mutations in Emery-Dreifuss muscular dystrophy and their effects on emerin protein expression. Hum. Mol. Genet. 7: 855-864. 1998)

EMD @ www.dmd.nl

Back
Sequence variant tables - Search sequence variants
Total results: 2.

Exon	DNA: Allele 1	RE-Site	DNA: Allele 2	RNA	Protein	Frequency	Disease	Reference	DNA/RNA	Technique	EMDdb-ID	Remarks
01	c.-150_-129del		-	?	-		?EDMD	Wulff	DNA	SEQ	EMD_00001	sporadic patient Deutschland; EMD normal in WBC
01	c.-150_-129del		-	(=)	-		?EDMD	Manilal	RNA	RT-PCR+SEQ	EMD_00001	sporadic patient Deutschland (maternal grandmother carrier); EMD normal in WBC

Total results: 2. Showing results 1 to 2.

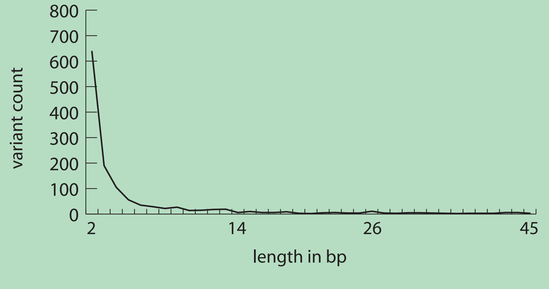
Summary of data in Locus Variants track

Substitutions comprise the majority of the variants within the LSDBs and Swiss-Prot.

Value histogram for hgFixed.gv.baseChangeType

value	count	graph
substitution	30302	**********
deletion	2076	****
insertion	322	*
complex	302	*
duplication	181	

Examining just the LSDB data, the length distribution of the variants reflects the dominance of substitutions, with 9,396 single bp variants (not shown in graph). The larger variants vary greatly in size, with some as large as 187, 143 bp. These larger mutations are deletions, insertions, and duplications, some of which remove genes while others remove regulatory regions.



Distribution of variants by chromosome