BIOL 419/519 Bioinformatics Research – July 16, 2013
Human Genome Annotation – Paul Szauter

I. Introduction to the Personal Genome Project (PGP)

The Personal Genome Project (PGP) is a group of researchers and volunteers who seek to make individual human genomes and health information freely available for public use by anyone. The PGP is an important complement to the biomedical research literature, because the consequences of individual genetic variants can be investigated in the context of an entire genome in a population of individuals who have not been singled out by their disease status.

Go to the PGP website:

http://www.personalgenomes.org

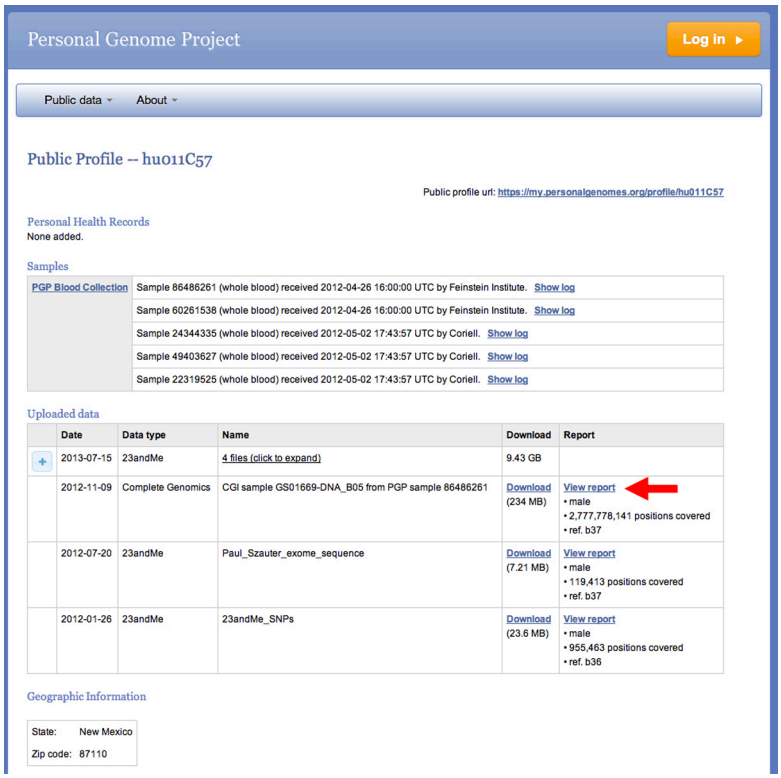
Click **PGP Community** in the top navigation menu to go to:

http://www.personalgenomes.org/community.html

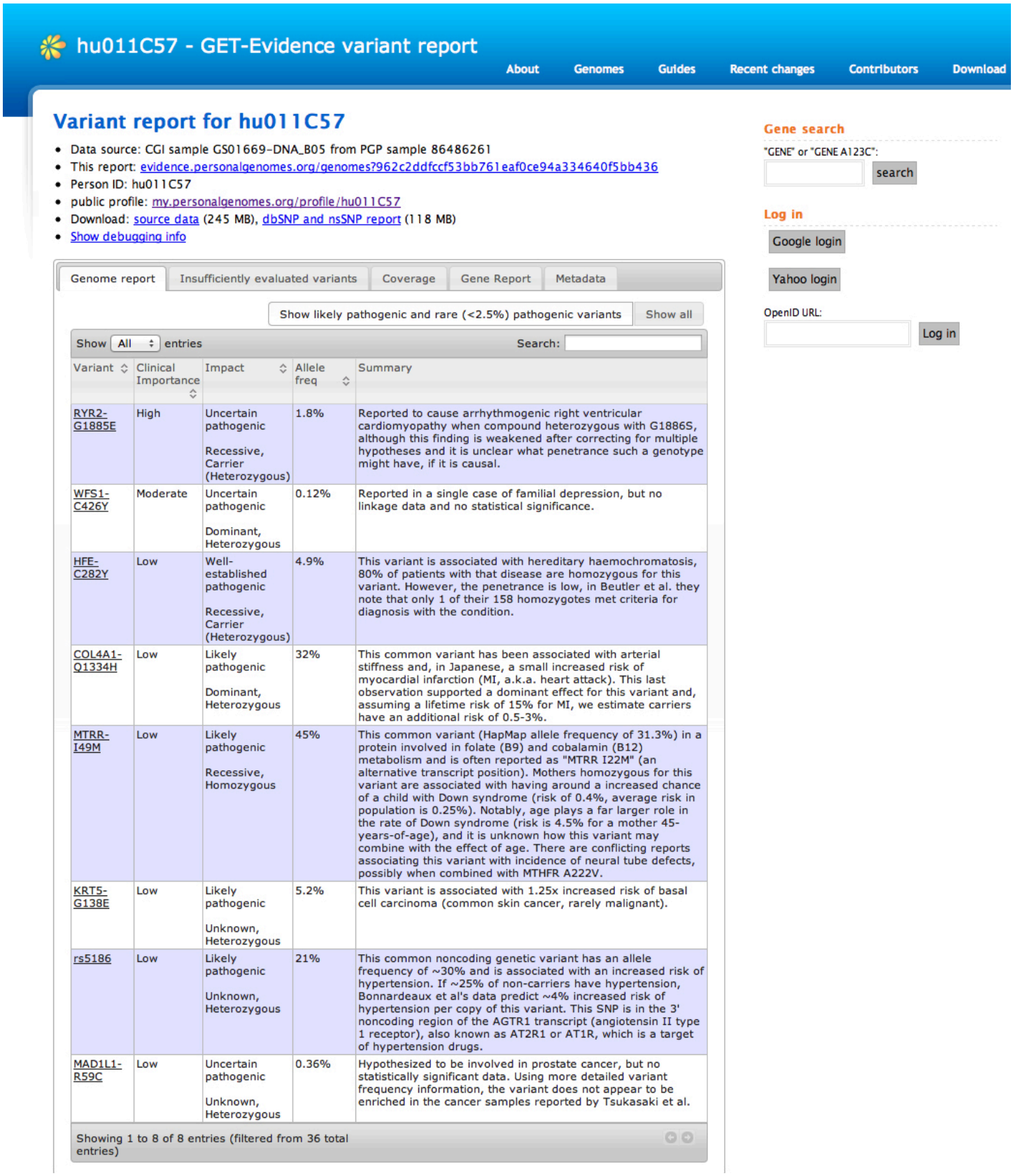
Click **View public profiles** to go to:

https://my.personalgenomes.org/users

Use the selector to **Show 100 entries**. Scroll down to **PGP89** and click **hu011C57**. In the **Complete Genomics** line of the table, click **View report**.



The page that appears displays the **Genome report** by default. These are likely pathogenic and rare variants found in the genome of that individual. Click any allele in the **Variant** column for more information.



hu011C57 - GET-Evidence variant report

About Genomes Guides Recent changes Contributors Download

Variant report for hu011C57

- Data source: CGI sample GS01669-DNA_B05 from PGP sample 86486261
- This report: evidence.personalgenomes.org/genomes?962c2ddfccf53bb761eaf0ce94a334640f5bb436
- Person ID: hu011C57
- public profile: my.personalgenomes.org/profile/hu011C57
- Download: source data (245 MB), dbSNP and nsSNP report (118 MB)
- Show debugging info

Gene search

"GENE" or "GENE A123C":

search

Log in

Google login

Yahoo login

OpenID URL:

Log in

Genome report | Insufficiently evaluated variants | Coverage | Gene Report | Metadata

Show likely pathogenic and rare (<2.5%) pathogenic variants | Show all

Show All entries Search:

Variant	Clinical Importance	Impact	Allele freq	Summary
RYR2-G1885E	High	Uncertain pathogenic Recessive, Carrier (Heterozygous)	1.8%	Reported to cause arrhythmogenic right ventricular cardiomyopathy when compound heterozygous with G1886S, although this finding is weakened after correcting for multiple hypotheses and it is unclear what penetrance such a genotype might have, if it is causal.
WFS1-C426Y	Moderate	Uncertain pathogenic Dominant, Heterozygous	0.12%	Reported in a single case of familial depression, but no linkage data and no statistical significance.
HFE-C282Y	Low	Well-established pathogenic Recessive, Carrier (Heterozygous)	4.9%	This variant is associated with hereditary haemochromatosis, 80% of patients with that disease are homozygous for this variant. However, the penetrance is low, in Beutler et al. they note that only 1 of their 158 homozygotes met criteria for diagnosis with the condition.
COL4A1-Q1334H	Low	Likely pathogenic Dominant, Heterozygous	32%	This common variant has been associated with arterial stiffness and, in Japanese, a small increased risk of myocardial infarction (MI, a.k.a. heart attack). This last observation supported a dominant effect for this variant and, assuming a lifetime risk of 15% for MI, we estimate carriers have an additional risk of 0.5-3%.
MTRR-I49M	Low	Likely pathogenic Recessive, Homozygous	45%	This common variant (HapMap allele frequency of 31.3%) in a protein involved in folate (B9) and cobalamin (B12) metabolism and is often reported as "MTRR I22M" (an alternative transcript position). Mothers homozygous for this variant are associated with having around a increased chance of a child with Down syndrome (risk of 0.4%, average risk in population is 0.25%). Notably, age plays a far larger role in the rate of Down syndrome (risk is 4.5% for a mother 45-years-of-age), and it is unknown how this variant may combine with the effect of age. There are conflicting reports associating this variant with incidence of neural tube defects, possibly when combined with MTHFR A222V.
KRT5-G138E	Low	Likely pathogenic Unknown, Heterozygous	5.2%	This variant is associated with 1.25x increased risk of basal cell carcinoma (common skin cancer, rarely malignant).
rs5186	Low	Likely pathogenic Unknown, Heterozygous	21%	This common noncoding genetic variant has an allele frequency of ~30% and is associated with an increased risk of hypertension. If ~25% of non-carriers have hypertension, Bonnardeaux et al's data predict ~4% increased risk of hypertension per copy of this variant. This SNP is in the 3' noncoding region of the AGTR1 transcript (angiotensin II type 1 receptor), also known as AT2R1 or AT1R, which is a target of hypertension drugs.
MAD1L1-R59C	Low	Uncertain pathogenic Unknown, Heterozygous	0.36%	Hypothesized to be involved in prostate cancer, but no statistically significant data. Using more detailed variant frequency information, the variant does not appear to be enriched in the cancer samples reported by Tsukasaki et al.

Showing 1 to 8 of 8 entries (filtered from 36 total entries)

The three most important tabs in the top Navigation tabs show you:

1. Genome report. Likely pathogenic and rare variants found in the genome of that individual.

2. Insufficiently evaluated variants. These are variants that need further annotation and evaluation.

3. Gene Report. This shows all genes in which variants have been identified.

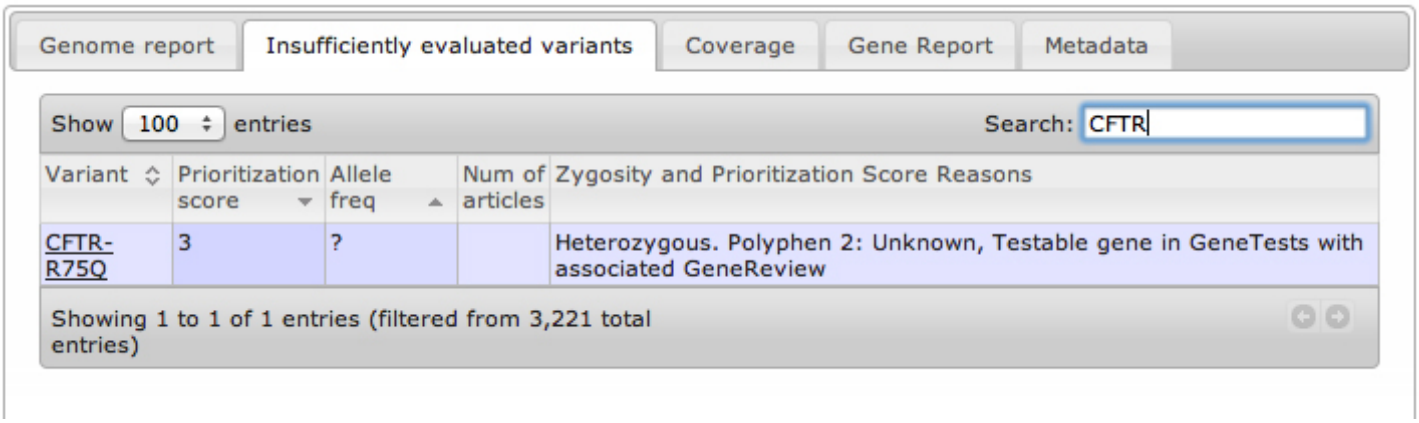
In the **Genome Report** table is a **Search** box that lets you search for variants of a specific gene in that individual.

There is a **Gene search box** in the upper right that lets you look for variants of a specific gene among all PGP individuals.

Immediately below the **Gene search box** are **Log in** buttons. If you have a Google or Yahoo email account, you can log in to improve the annotation.

II. Improving annotation for a specific variant

Click the tab for **Insufficiently evaluated variants**. In the Search box in the Genome report table, type **CFTR**. A single variant appears. Click the allele.

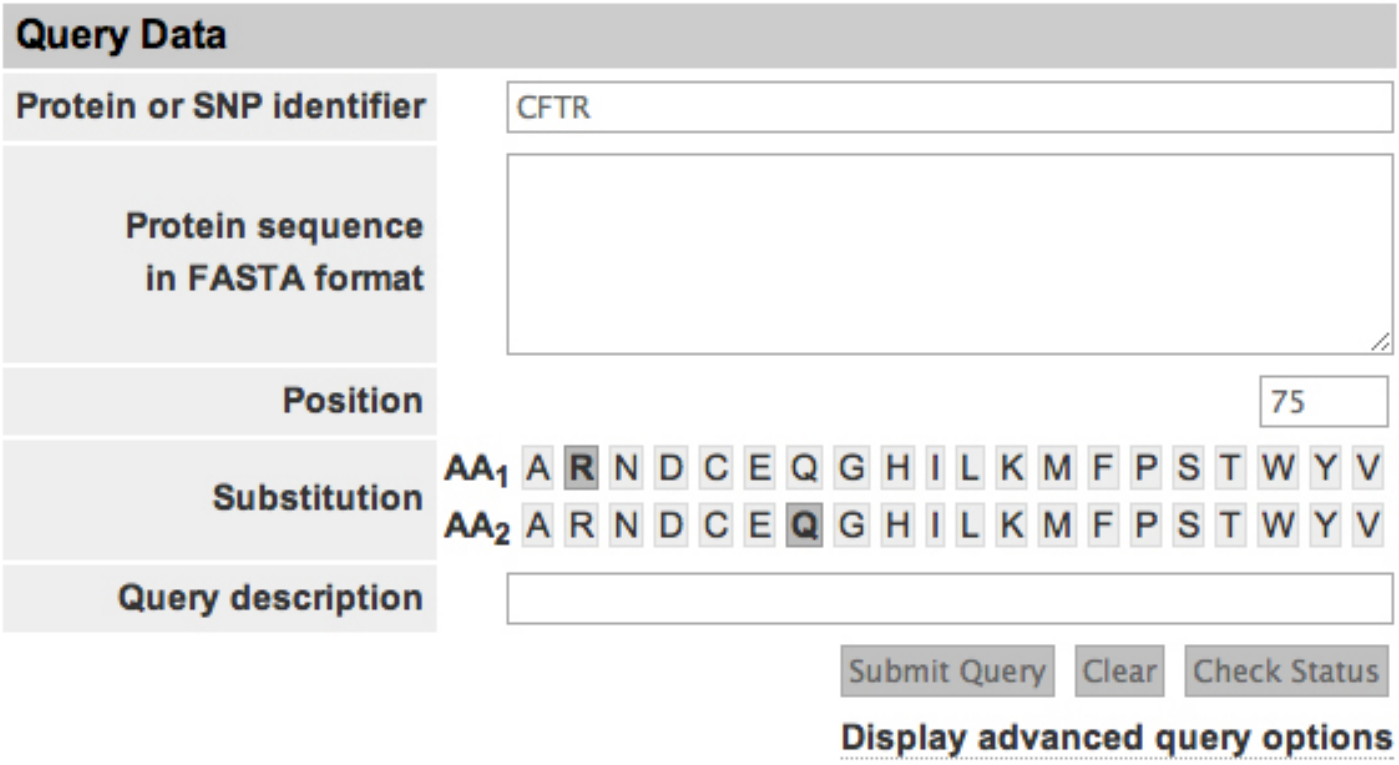


The allele is CFTR-R75Q. This is a missense allele in which the 75th amino acid of the CFTR protein is changed from Arginine (R) to Glutamine (Q).

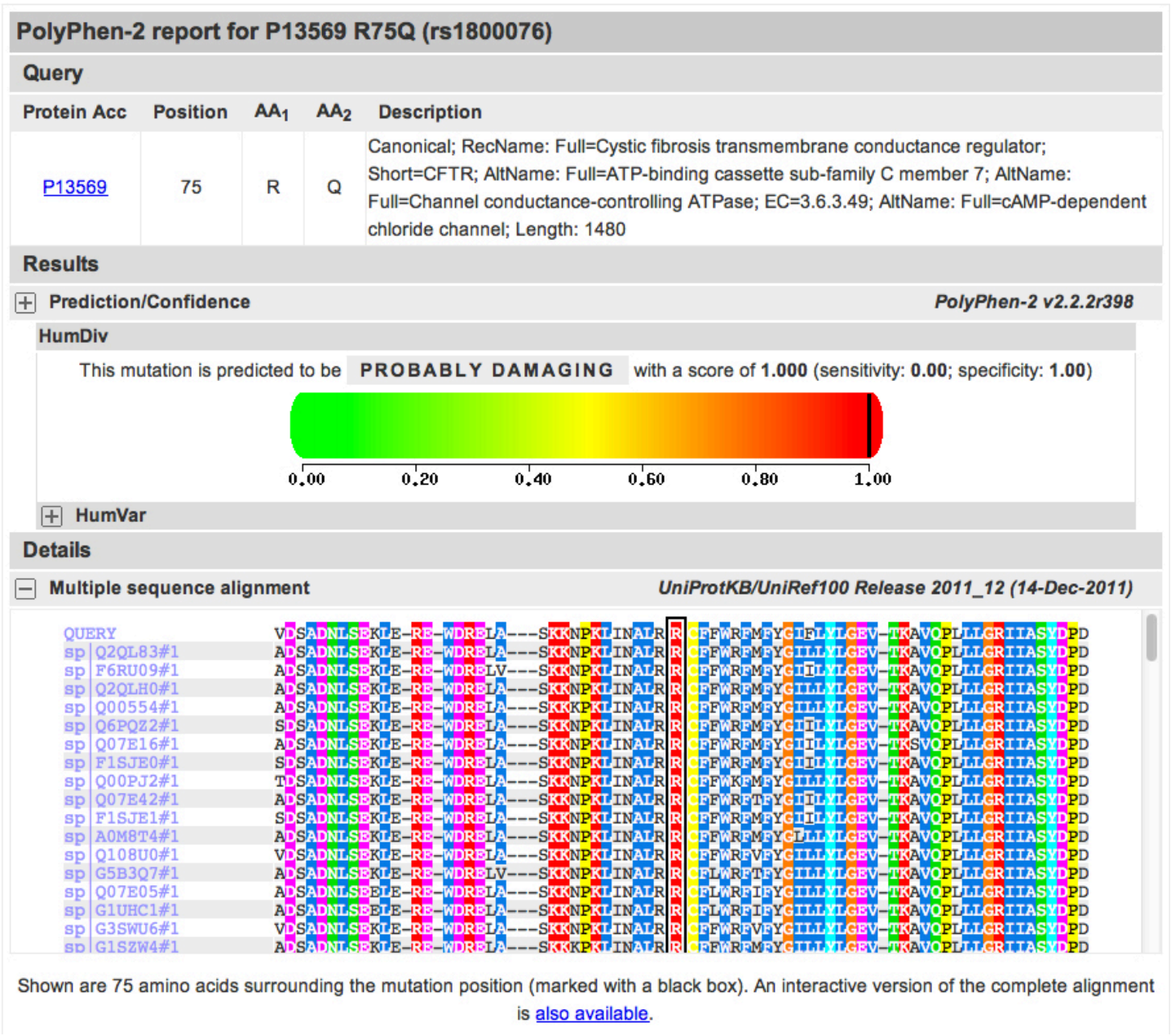
To evaluate whether this allele is likely to be harmful, we use PolyPhen-2. Open a new browser window to the PolyPhen-2 site from the Tools page on the course website:

http://genetics.bwh.harvard.edu/pph2/

Enter **`CFTR`** as the Protein, **`75`** as the position, and select **R** as AA1 and **Q** as AA2. Click **Submit Query**.



When the job is done running, click the link to go to the screen shown below. Click the plus sign to expand the **Multiple sequence alignment**.



This allele is rated as **Probably damaging**. Each line in the multiple alignment represents a different species. Click any line to see the entry for that species.

We can use OMIM to see if this variant has already been described. Open a new window to OMIM using the link from the Tools page of the course website:

http://www.omim.org

Enter **`CFTR`** in the search box and click **Search**.

The top link takes you to the gene entry for CFTR. At the top of the page are links to specific disease entries associated with the CFTR gene.

Scroll down to the **Allelic Variants** section or just search for **`ARG75GLN`**. This search fails, so try links in the **Variation** menu in the right margin. The **Locus Specific DBs** link works; search for ARG75GLN in the database there.

There is a citation listed, which you can locate through PubMed (PMID: 1710599). This paper is cited by a more recent paper (PMID: 20977904). You can use **summarize the information in this paper** on the page for the CFTR-R75Q allele (click the evaluating evidence link for more information).

III. Searching for all variants of a specific gene in the PGP

In the **Gene Search box**, enter **`CFTR`**.

This returns a multiple-page table of CFTR variants. The **Genomes** column contains links to individual genomes with that variant. Note that some of these alleles have summaries written by annotators.

IV. Approaches to improving annotation at the PGP

There are several approaches to improving annotation at the PGP, outlined below.

A. Genomic Checkup. The American College of Medical Genetics (ACMG) has released a list of genes to be investigated as incidental findings when a person's genome is analyzed. Use this gene list to perform a checkup on an individual in the PGP. Do you have any reportable findings?

B. Newborn Screening. Newborns are screened for treatable metabolic disorders. The genes associated with these disorders are known. You can screen individuals in the PGP to see if they carry variant alleles of any of these genes.

C. Insufficiently Evaluated Variants. For any individual, look at insufficiently evaluated variants. Pick variants whose frequency is known, but below 1% (or 5% if you are ambitious).

D. Literature Based. Use OMIM or the primary literature to find genes that are interesting. Look for variants among individuals in the PGP.

V. Report Interesting findings

Keep track of the genes and variants that you have evaluated, even those that you have found uninteresting. If you find something interesting, share it with the class. If you are not confident about annotating the variant yourself, collect and submit your information to pzauter@unm.edu for further evaluation.