University of Helsinki
Biometry and Bioinformatics II, Fall 2013
In silico prediction of protein-damaging single nucleotide variants

Practical work guidelines
Case study 1

Brief patient history

Girl with infantile cardiomyopathy, born healthy but somewhat small to non-consanguineous parents after an uneventful pregnancy. At 3.5 months she was admitted to hospital because of poor feeding, failure to thrive, delayed motor development, and severe generalized muscle weakness. The disorder progressed despite intensive medication for heart failure. The patient died at the age of 10 months of cardiac insufficiency.

Data

Current exome sequencing technologies identify approximately 30 000 SNVs in a human exome. The file cs1PatientExomeVariantDataSample.txt contains a small, simplified sample of variants found in the exome of the above patient. First and foremost, look at the data sample and try to understand it. This is an important step of the exercise.

Task

Seeking a genetic diagnosis, your lab has decided to perform costly functional studies on variants detected in the patient's sequenced exome. Your task is to use SIFT and PolyPhen-2 to identify, among the given data, the best gene variant candidate(s) for such studies.

Hints

- Take notes on your observations and results so that you are able to comment on them.
- On the SIFT web site, you will find several tools for DNA and protein sequences. The one to be used here is SIFT Human SNPs.
- For both SIFT and PolyPhen-2, format your input data as required by the tools.

- In the patient data, the counts of bases detected at a certain genomic position determine the 'Call base'. Observe this in the data. The 'Call base', in turn, indicates whether the variant is homozygous or heterozygous. Homozygous variants have an A, C, G, or T as 'Call base', whereas heterozygous variants have mixed base codes, according to the standard in the table below.

Mixed base code	Base
R	A or G
Y	C or T
K	G or T
M	A or C
S	G or C
W	A or T