

GET-Evidence - Feature #529

Interpret phased (haplotype) data

02/17/2011 07:59 PM - Madeleine Ball

Status:	New	Start date:	02/17/2011
Priority:	Normal	Due date:	
Assigned To:	Evan Maxwell	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:		Totalhours:	
Billable:		Resolution:	
Estimatedhours:			
Hours:			

Description

The sequencing companies aren't generating haplotype information from individual genomes, but someday they might, and there's no reason we can't start working on this now. Some haplotype information can be generated by combining data from trios: child and two parents. For example we can use NA19240, NA19238, NA19239 genome data. Even if sequencing companies never develop haplotypes for an individual genome, researchers/clinicians may sequence all three individuals and determine some haplotype information on their own.

We'd like to create a report that uses haplotype data to report which copy of a gene a variant is on and combines these in an intelligent manner. This is very important for interpretation -- many genetic diseases are recessive and are caused by a variety of **different** variants - currently these would show up as two heterozygous variants. If one is on each copy of the gene then you may have a problem, but if they are both on the same copy of the gene then the other copy should be fine.

An example from within the PGP: PGP1 (hu43860C) is heterozygous for SERPINA1-E366K and SERPINA1-E288V. As I understand it, these are considered somewhat pathogenic ("PiSZ" genotype) when one is on each gene. Perhaps they are always heterozygous, but we would like to see that they are indeed on different copies of the gene and not both in the same gene.

History

#1 - 02/18/2011 02:48 PM - Evan Maxwell

- Assigned To set to Evan Maxwell

#2 - 02/23/2011 11:55 AM - Alexander Wait Zaranek

- File compound_het_report.png added

#3 - 07/07/2011 02:20 PM - Madeleine Ball

Evan Maxwell (emaxwell) has added a python module for phasing of trio data commit: <https://github.com/madprime/get-evidence/commit/fed5761e>

I've added this to the python pipeline [62a45dd5](#) and also created an interpretation of phased data [b03d0a30](#)

All we need is to add some boxes in the php side where genomes are uploaded to finish it all -- the processing step checks if there's a metadata file in the source directory and, if so, reads it and checks for shasums that identify two parent genomes to send to the trio phasing module.

Files

compound_het_report.png	199 KB	02/23/2011	Alexander Wait Zaranek
-------------------------	--------	------------	------------------------