Supplemental Methods

VarSifter is written in Java (Oracle). VarSifter includes the TableSorter.java class (Milne, et al., 2004), which is copyrighted by Oracle and released with an open license (see software for license).

Requirements

VarSifter is written in Java, and therefore can be run on any operating system with Java installed. VarSifter has been successfully tested on Windows XP and XP 64-bit, Mac OS X, and the CentOS and Gentoo distributions of GNU/Linux. Basic VarSifter functionality requires Java JRE 1.5 or newer, but Java JDK 1.6 is recommended, as the custom query/filter functionality requires Java JDK 1.6 or better (JDK is the full Java Development Kit). VarSifter has been successfully tested on Oracle Java and the IcedTea build of the open source OpenJDK.

VarSifter runs best with a current 64-bit processor, and a 64-bit Java version (which requires a 64-bit Operating System). The amount of RAM required depends on the number of samples you wish to view. A machine with 1 GB RAM will allow all variants in ~30 exomes; 8GB RAM will allow all variants in ~180 exomes.

Additionally, the Java Universal Network/Graph Framework 2 (JUNG2) is also required (VarSifter has been tested with version 2.0.1.) This is included in the "package" download, but can also be downloaded from: http://jung.sourceforge.net/.

Performance

A dual 2.66 Ghz Intel Xeon Mac Pro with 7 GB RAM running Mac OS X 10.6.6 was used to evaluate performance. Oracle Java Development Kit version 1.6.0_24 64-bit was used for testing. VarSifter (version 0.12) memory usage was measured using the Java command *jmap –histo:live <process ID>*, as this accurately measures the memory in use, not simply the size of the heap Java has allocated for future use. jmap is therefore more

accurate and reproducible when measuring memory of Java programs. The average

difference between jmap and the Apple Activity monitor readings was ~100MB, which is

attributed to the Virtual Machine overhead (and is not included in Supp. Fig. 4.)

Microsoft Excel 2011 for Mac (version 14.0.1) memory usage was measured using Apple

Activity Monitor (version 10.6), "Real Mem". Excel generally required ~74MB before

opening the data file, and this value was subtracted from the total memory usage in Supp.

Fig. 4. VarSifter startup time was measured from pressing enter to run the java command

until VarSifter appeared fully loaded on the screen. The spreadsheet loading time was

measured from pressing "Finish" to begin loading of a tab-delimited text file until the

display appeared fully loaded (the spreadsheet program was restarted for each file, and

the initial Excel loading time was not counted.)

Performance measurements for Excel are only available for 60 samples or less, as

Excel for Mac is only available as a 32-bit version, which limits the amount of memory

that can be addressed (and will no longer be a limitation once a 64-bit version is

available.)

Data Files

Exome files with increasing numbers of samples (Sup. Fig. 4) were generated

from the indicated number of individuals (10-160) from the ClinSeq™ project (Biesecker,

et al., 2009). The files represent all variants identified from exome capture using the

SureSelect Human All Exon Kit (Agilent Technologies Inc., Santa Clara, CA) followed

by sequencing with a Genome Analyzer IIx (Illumina, Inc., San Diego, CA). These files

include 21 annotation columns. Sequence analysis and variant determination was

performed as previously described (Wei, et al., 2011).

License

As VarSifter was developed as part of the authors' official duties for the United States Government, the software cannot be copyrighted. The software is therefore released with the following licensing statement:

"PUBLIC DOMAIN NOTICE

National Human Genome Research Institute, National Institutes of Health

This software/database is "United States Government Work" under the terms of the United States Copyright Act. It was written as part of the authors' official duties for the United States Government and thus cannot be copyrighted. This software/database is freely available to the public for use without a copyright notice. Restrictions cannot be placed on its present or future use.

Although all reasonable efforts have been taken to ensure the accuracy and reliability of the software and data, the National Human Genome Research Institute (NHGRI) and the U.S. Government does not and cannot warrant the performance or results that may be obtained by using this software or data. NHGRI and the U.S. Government disclaim all warranties as to performance, merchantability or fitness for any particular purpose.

In any work or product derived from this material, proper attribution of the authors as the source of the software or data should be made."

VarSifter includes the TableSorter.java class, which is copyrighted by Oracle, and released with an open license (see software for license.)

VarSifter source code, binaries, and example data files are available at

http://research.nhgri.nih.gov/software/VarSifter/.

References

Biesecker, L.G., *et al.* (2009) The ClinSeq Project: piloting large-scale genome
   sequencing for research in genomic medicine, *Genome Res*, **19**, 1665-1674.
Milne, P., *et al.* (2004) TableSorter.java. Oracle.
Wei, X., *et al.* (2011) Exome sequencing identifies GRIN2A as frequently mutated in
   melanoma, *Nat Genet*, **43**, 442-446.