

Integrated variants from 13,000 complete genomes available to public in Kaviar database

September 23 2015

The Institute for Systems Biology (ISB) and the Inova Translational Medicine Institute (ITMI) announced today a new release of Kaviar, the most comprehensive collection of human genomic variants currently available to the public. This release expands on the January 2015 release most notably by the addition of 3842 whole genome sequences provided by ITMI. Inova, a not-for-profit healthcare system based in Northern Virginia, founded ITMI to transform healthcare from a reactive to a predictive model.

First described in Glusman et al., *Bioinformatics*. 2011, Kaviar answers the question: "Has this variant been seen before, and if so, how often?" Kaviar currently lists 169 million SNV (single nucleotide variant) sites and 48 million indels and substitutions.

Kaviar combines 31 public data sources plus 4622 private whole genome sequences. Kaviar integrates genome variation data from 77,238 unrelated individuals, including the 1000 Genomes Project's data, UK10K COHORT allele frequencies representing 3781 individuals, the Exome Aggregation Consortium (ExAC) 63,000 exomes, and 808 whole genomes from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Diversity is enhanced by the inclusion of data from the Simons Foundation Diversity Project and several population-specific data sources. Very rare variants in private data (those observed in fewer than 3 individuals) are omitted from Kaviar to protect the privacy of the

individuals.

Kaviar is accessible at <http://db.systemsbiology.net/kaviar/>, where users may query the database via a web interface. Kaviar accepts queries for genomic locations, then reports which variants have been observed at those locations, and at what frequency. Kaviar can also be queried programmatically via a web service. Users may also download the complete Kaviar database in variant call format (VCF) and use common software tools to query it.

ISB produces regular updates to Kaviar, incorporating updates to dbSNP, newly obtained genome sequences, and improvements in the reference assembly. An upcoming release will contain genotype frequencies in addition to the allele frequencies.

Provided by Institute for Systems Biology

Citation: Integrated variants from 13,000 complete genomes available to public in Kaviar database (2015, September 23) retrieved 28 April 2024 from <https://medicalxpress.com/news/2015-09-variants-genomes-kaviar-database.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--