MedSavant: A high-performance search engine for genetic variants

Marc Fiume[1], Orion Buske[1], James Vlasblom[1], Eric Smith[1], Andrew Brook[1], Khushi Chachcha[2], Sergiu Dumitriu[1], Christian R. Marshall[3], Kym Boycott[4], Marta Girdea[1], Peter Ray[3], Gary Bader[5], Michael Brudno[1,3]

[1]*University of Toronto, Toronto, Ontario, M5S 1A1, Canada*

[2]*University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada*

[3]*The Hospital for Sick Children, Toronto, Ontario, M5G 1X8, Canada*

[4]*University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada*

[5]*The Donnelly Centre, Toronto, Ontario, M5S 3E1, Canada*

Recent advancements in DNA sequencing technologies have economized the sequencing of individual genomes, creating potential to improve diagnostics and treatment for patients affected by genetic diseases. A significant challenge in utilizing these technologies in the clinical setting has been in the implementation of data analysis strategies that can be used to easily and efficiently identify causative genetic mutations from the large number of variants discovered through sequencing. A common approach is to annotate and iteratively filter potentially causal genetic mutations using ad-hoc combinations of computational scripts and their parameters; however, this method necessitates informatics expertise, is slow, non-interactive, and does not scale to the extent needed for clinical and other large-scale sequencing applications.

To facilitate the discovery of causative genetic mutations, we have developed MedSavant, an integrated solution for the storage, annotation, filtration, prioritization, and visual inspection of variants. It is entirely graphical, interactive, and scalable to manage datasets generated by large-scale sequencing projects. To accomplish this, the system employs big data analytics technologies optimized for genomic datasets that are capable of delivering the results of complex dynamic queries nearly instantaneously while using significantly less storage resources compared to the standard flat file equivalent[1].

The MedSavant platform was designed for use in large-scale research and clinical genome sequencing projects. We have created two MedSavant installations to demonstrate its use in these respective applications. The first installation contains over one billion unique variant entries from ZZZ of the 1000 Genomes Project[2] individuals. This installation is publically accessible via an online portal. The second installation contains clinical data from 425 individuals from the FORGE Consortium, a set of projects whose aim is to discover the etiology of rare genetic disorders. Using MedSavant equipped with the internally-developed Mendel App, we independently discovered the causal gene in 16 tested FORGE Projects.

Motivation

Next-generation DNA sequence analysis holds the promise of improving diagnostics and treatment for individuals affected with genetic diseases. A significant challenge to fulfilling its promise is in implementing data analysis strategies that can be used to easily and efficiently identify causative genetic mutations from the large number of variants discovered through sequencing in the clinical

setting. A common approach is to annotate variants with informative metadata (e.g. genomic context, predicted harmfulness), filter for potentially causal genetic mutations based on complex criteria, and iteratively refine the previous steps after manual inspection of the results. This method utilizes a combination of ad-hoc and loosely integrated computational scripts that rely on flat files as the data-transfer medium.

This existing paradigm for genomic variant analysis, which involves serial processing of flat files with manual inspection as an endpoint, inherently requires a substantial amount of informatics expertise to run, is non-interactive and time-consuming, and thus does not scale to the extent needed for clinical and other large-scale sequencing applications. It is well appreciated in bioinformatics and in other scientific domains that visually-guided real-time exploration significantly aids the understanding of big datasets, as is evidenced by the success of tools like ABySS-Explorer[3], Savant Genome Browser[4], and Galaxy[5]. These tools deliver computationally-intensive analytics through accessible user interfaces.

A few graphical applications for variant searching have previously been developed. VarB[6] enables visual exploration and rudimentary filtration of genomic variants based on size, quality, depth, and codon effect; but these features alone are not enough to resolve causal genetic mutations. VarSifter[7] and the SNP and Variation Suite[8], the latter being a commercial tool, are other desktop solutions for management, filtration, and visualization of genetic variants; however, their architecture places significant limitations on performance and restricts access to the data to a single computer. VarSifter loads the complete volume of genotyped variants into memory, and for this

reason is practical only for exome analysis. As yet, there is an unmet need for accessible software to perform dynamic visual analyses of large volumes of genetic variant data detected through sequencing, with an emphasis on facilitating disease etiology discovery.

Introduction

We introduce MedSavant, an integrated solution for the storage, annotation, filtration, prioritization, and visual inspection of variants that is entirely graphical, interactive, and scalable to manage datasets generated by large sequencing projects. MedSavant is a standalone client-server application that uses a custom high-performance database engine to store and perform faceted search queries on data. This design results in significantly improved performance and user-friendliness over standard flat file or client-side alternatives.

While standardized text-based flat file formats have commonly been used as a data-transfer medium for genomic datasets (e.g. BED, GFF, and the Variant Call Format (VCF)) and enabled interoperability between computational tools, they do not facilitate faceted search - i.e. searches based on arbitrary metadata affiliated with the records. This makes the interactive refinement of searches for causal genetic mutations difficult, as these are typically based on a complex set of quality, contextual, and other criteria that necessitate reprocessing these flat files whenever the search parameters change. As a result, sets of candidate causal mutations often contain large numbers of poor quality or otherwise irrelevant variants that are frequently manually filtered, rather than resorting to further parameter refinement and reprocessing. This problem is exacerbated as

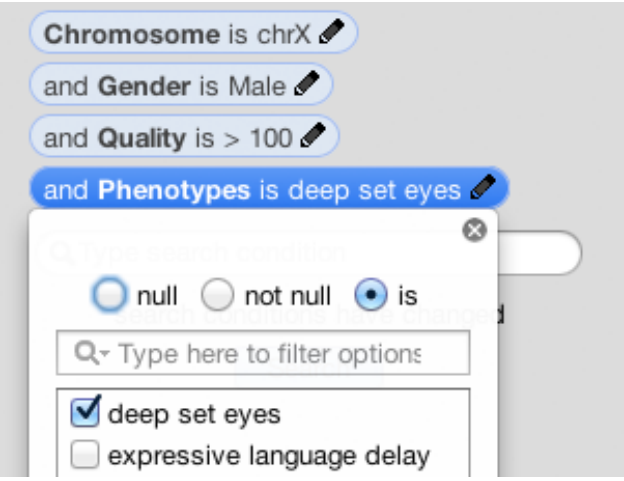
both the size and diversity of genomic datasets continue to increase.

MedSavant

Data and Annotation MedSavant employs the Infobright[9] high-performance database and a custom query engine to quickly deliver results of faceted searches on genomic variant datasets. The database stores and compresses tables column-wise, and so uses substantially fewer resources and executes faster than flat file-based processing pipelines: MedSavant databases require 10-20X less disk space compared to raw VCF files, while simultaneously making it possible for common queries to execute in seconds or less. The facets that are searchable by MedSavant are numerous and customizable. They include both low-level genotype information (e.g. base quality, chromosome position) and high-level information (e.g. ethnicity, family, gender, phenotype). MedSavant also contains an internal annotation engine similar to ANNOVAR[10] and SnpEff[11] that automatically appends to all uploaded VCF files user-specified searchable annotations such as harmfulness prediction scores (e.g. SIFT[12], PolyPhen-2[13]), allele frequencies from population sequencing studies (e.g. 1000 Genomes Project[2], NHLBI Exome Sequencing Project[14]), and ontology relationships (e.g. Gene Ontology[15], Human Phenotype Ontology[16], OMIM[17]). Table shows a list of data and annotations stored by MedSavant. Custom annotations are also supported.

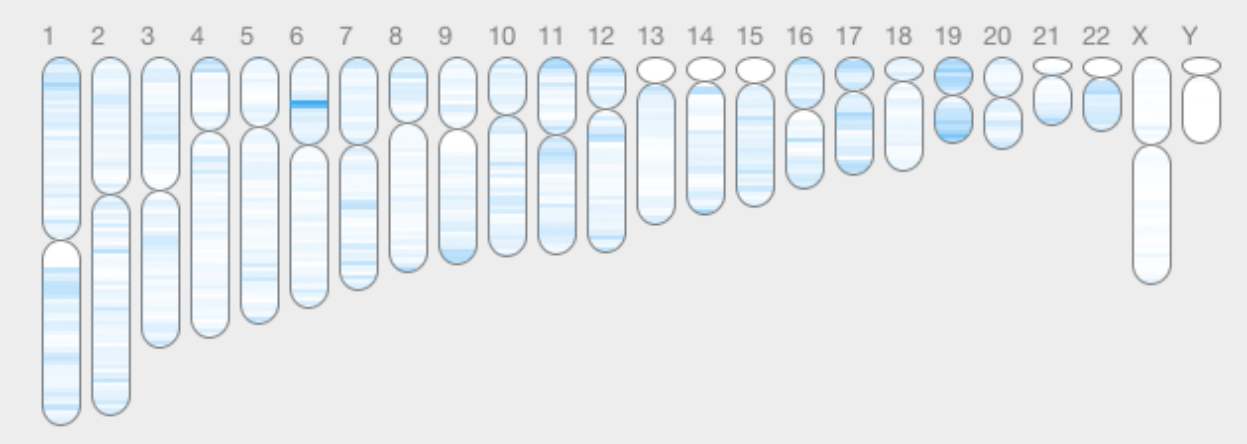
Search The MedSavant user interface manages remote authenticated access to the database, and offers a simple means for query construction and visualization of results. A list of search conditions and familiar widgets for tuning them are provided. Search conditions can be grouped,

unioned, and intersected, yielding a very expressive query language. Queries can combine conditions based on both genotype and phenotype information; the synergistic use of both types of data allows for segmentation of mutations based on disease subgroups, an important functionality when investigating complex diseases like autism spectrum disorders or cancer. The user interface for faceted variant search is shown in Figure and specific examples of possible queries are provided in later sections. Searches can be constructed incrementally, one facet at a time, with nearly instant feedback about parameter effectiveness at each step. The dynamic provision of guidance during filter construction allows for rapid exploration of the parameter space and is one of the most significant time-saving advantages of the MedSavant platform. Moreover, MedSavant saves search states internally, producing no intermediary files that need to be managed by the user. Once a search has been fine-tuned, it can be saved and reused for reproducing results on different cohorts or on samples whose genotypes will be processed in the future.

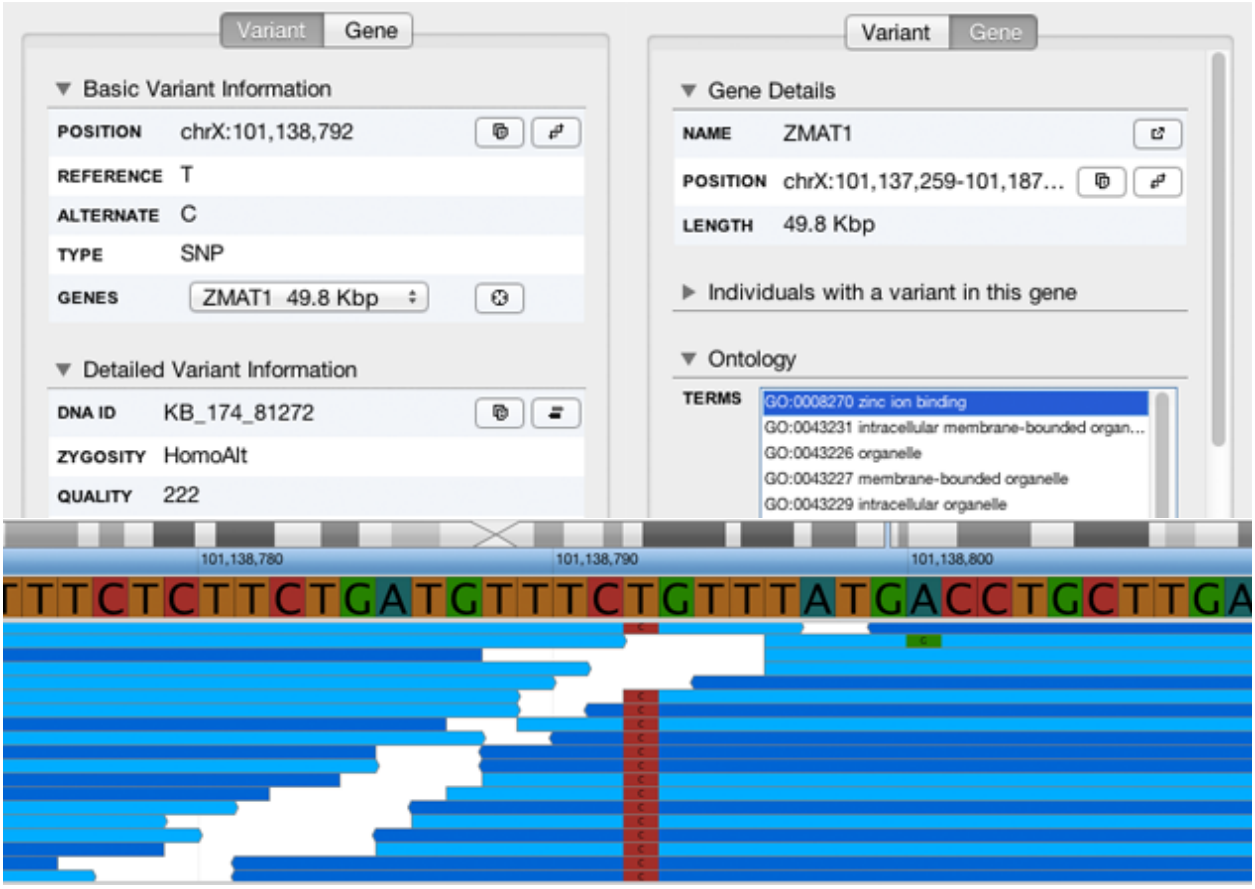


Inspection Candidate mutations can be manually inspected in various levels of detail as illustrated in Figure. The distribution of variants across the genome is depicted as a heatmap on a karyotypic

ideogram. The distribution of variants can also be charted per searchable facet (as a histogram or pie chart), and between searchable facets (as a scatter plot). The full list of candidate variants is also represented in spreadsheet format with an associated Inspector that displays detailed information about selected variants, including all information contained in the original VCF file and a list of nearby genes. Further, the Inspector presents information regarding the function and relevance of nearby genes, including associated terms in the Gene Ontology, Human Phenotype Ontology, and OMIM. It also utilizes GeneMania[27], a service that finds related genes based on protein interaction and other networks, to suggest other genes to consider.



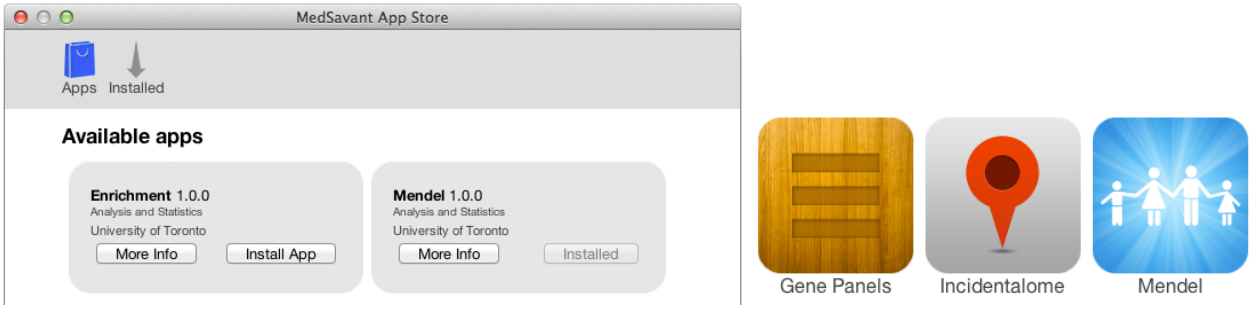
Chromo…	Position	Vari…	Quality	Zyg…	American F…	Harmfulnes…
chrX	99,941,705	SNP	222	HomoAlt	0.98	0.196
chrX	100,509,862	SNP	222	HomoAlt	0	0.979
chrX	101,138,792	SNP	222	HomoAlt	1	0.459
chrX	107,978,343	SNP	222	HomoAlt	0.07	0.405
chrX	107,979,475	SNP	222	HomoAlt	0.07	0.418



Visualization It is often informative to inspect variants in their genomic context to ascertain their predicted functional consequence. Links in the Inspector offer the ability to navigate to and browse the genomic context of a candidate variant (including the corresponding read alignments from which the genotype prediction was initially made) using the embedded Savant Genome Browser. Savant is a next-generation genome browser that contains unique visualization modes for identifying and validating genetic variants. It contains multiple representations for exploring variant datasets: a traditional track view that displays variants per sample along a linear genomic coordinate system, and a streamlined non-linear display showing a condensed map, allele frequencies,

and linkage disequilibrium derived from variable positions. Linking of the candidate variants list with the genome browser optimizes the inspection of results by removing the need to manually navigate to regions of interest as is required in traditional multi-tool workflows.

Apps More data processing may be required to resolve disorders from filtered variants, for example by leveraging statistical analysis or by cross-sectioning the results with external datasets. MedSavant offers an Application Programming Interface (API) to foster integration with additional data processing tools developed by the community. Third-party Apps can leverage the MedSavant platform to access data, perform searches, do custom analyses, and present results within the graphical framework. A section of MedSavant, called the Clinic, is reserved for apps that perform push-button analyses for clinically-oriented workflows. Apps can be published to the MedSavant App Store, making useful additions to the platform available to all users. Two internally developed Apps are discussed below.



Enrichment App A single phenotype may be caused by genetic mutations at different positions in the genome, and so it is informative to aggregate potentially deleterious mutations based on genes, gene functions, and pathways, to concentrate analyses on relevant and noticeably perturbed

biological functions. The Enrichment App aggregates variants by user-specified gene-lists, and terms in the Gene Ontology, Human Phenotype Ontology, or OMIM. Aggregation is a basic form of mutational burden analysis, and facilitates the process of identifying biological functions that are affected within the sequenced population, and ultimately understanding the genetic mechanisms of diseases.

Mendel App Cohort, pedigree, and inheritance model information help to resolve the segregation of variants with genetic conditions. For example, VAAST[28] is a popular probabilistic disease-gene finder that can use such information to discover the causal gene(s) for genetic disorders given a modest number of genotyped samples. The Mendel App is an internally-developed disease-gene finder that adds the ability to resolve disorders through case-control and pedigree analysis. Given complete pedigree information, Mendel can also perform segregation based on known inheritance models. Once installed Mendel is made available in the MedSavant Clinic; a clinical application of this utility is discussed in the section that follows.

Case Study

The MedSavant platform was designed for use in both large-scale research and clinical genome sequencing projects. We have created two MedSavant installations to demonstrate its use in these respective applications. The first installation contains over one billion unique variant entries from ZZZ of the 1000 Genomes Project[2] individuals. This installation is publically accessible via an online portal. The second installation contains clinical data from 425 individuals from the FORGE

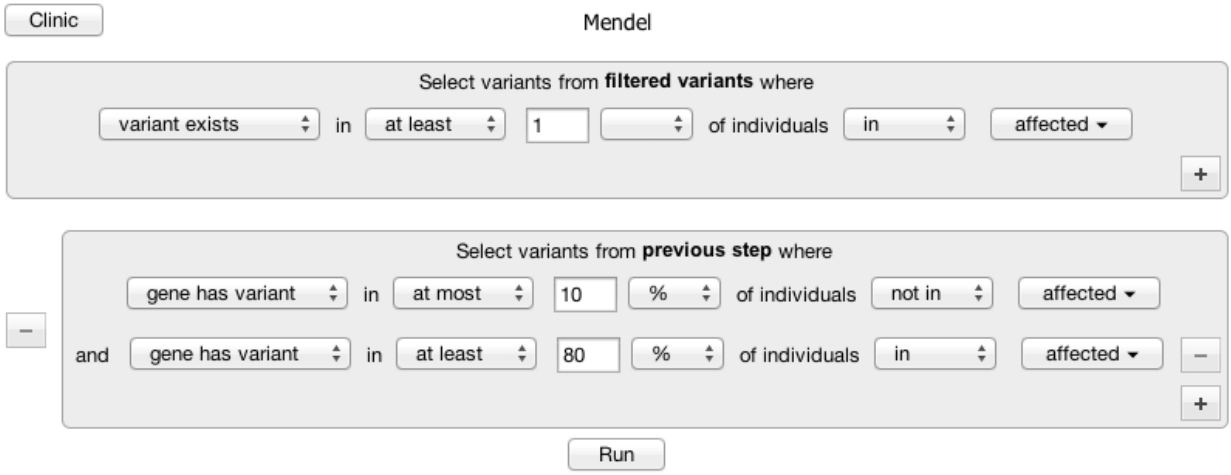
Consortium, a set of projects whose aim is to discover the etiology of rare genetic disorders. Using MedSavant equipped with the internally-developed Mendel App, we independently discover the causal gene in 16 tested FORGE Projects.

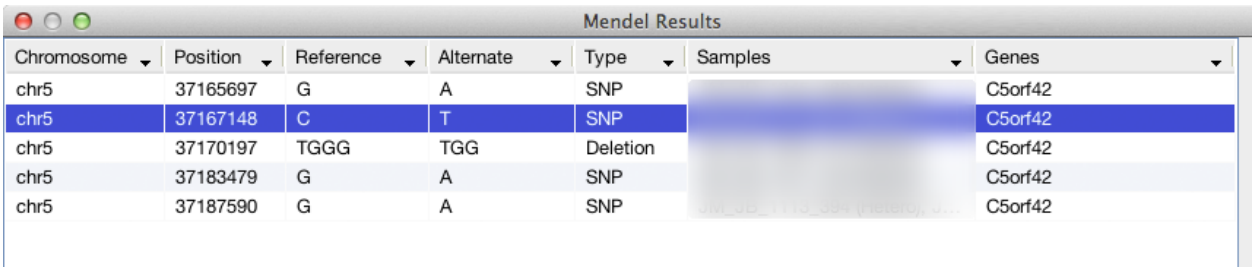
FORGE is a national consortium whose mission is to discover the etiology of a large number of rare genomic conditions. For example, Joubert syndrome is a rare brain malformation characterized by the absence or underdevelopment of the cerebellar vermis, an area of the brain that controls balance and coordination. Common symptoms in infants affected by Joubert syndrome include mental retardation, inability to coordinate voluntary muscle movements (ataxia), jerky eye movements (oculomotor apraxia), rapid breathing (hyperpnea), and decreased muscle tone (hypotonia). Joubert syndrome occurs in about 1 in 100,000 births, although there is a relatively high prevalence in the French-Canadian population, with several founder effects noted. SNP genotyping was previously performed on 9 affected individuals from seven families living in the Lower St. Lawrence region, and from these data it was previously discovered that mutations in C5orf42 cause Joubert syndrome in this French Canadian population[29].

20 FORGE disorders, including Joubert syndrome, were chosen for validation with Mendel based on the availability of genotype data, pedigree data, and consistency in their sequencing and genotyping pipelines. For each of the chosen disorders, genomic DNA from patients was captured using the Agilent SureSelect 50 Mb oligonucleotide library and sequenced with Illumina HiSeq2000, yielding 100bp paired-end reads. Putative PCR-generated duplicates were removed from the raw read data using Picard[30], alignment was performed using the Burrows-Wheeler

Aligner (BWA)[31], and custom scripts for SamTools[30], Pileup[30], and varFilter[30] were used to call variants.

We created a MedSavant database containing genotypes from all FORGE projects for which data was available, comprising 425 individuals and 138,640,418 variants identified from their samples. MedSavant was used to specify stringent quality filters (minimum coverage of 3; support from both strands; quality 50 for indels and 30 for SNVs; either exonic, splicing, or in UTR; allele frequency 0.05) yielding 427,765 variants. For each of the 20 chosen FORGE projects, Mendel was used to identify variants that segregate with the disorder using the remaining individuals, including family members of affected individuals and unaffected individuals from other projects, as controls. For Joubert syndrome, the Mendel expression shown in Figure produced missense and splicing mutations in C5orf42, with no other results. The positions of mutations identified by Mendel were manually inspected using the built-in genome browser. All 9 of the affected individuals were found to carry compound heterozygous mutations in this gene unlike any of the controls.





Using a similar workflow the causal gene was independently discovered for 16 of the 20 chosen disorders, listed in Table. Of the 4 disorders which were not rediscovered, the causal variant either did not pass the aforementioned quality filters or the causal variant was not included in the provided VCF files (e.g. a causal CNV was identified by other means). A summary of the Mendel expressions used for each confirmed disorder is provided in the Supplementary Information.

Conclusion and Future Work

The potential to gain new insights into genetic disease is currently encumbered by significant challenges in the data analysis methodologies that can be used to easily, efficiently, and intelligently identify causative genetic mutations from the large number of variants discovered through sequencing. Existing approaches either rely on complicated computational pipelines that include manual inspection only as an endpoint, or are desktop-based solutions that do not scale well for medium-to-large sequencing experiments. MedSavant is a genomic variant search engine built upon the client-server paradigm designed to meet the demands of even large population sequencing studies. It unifies the storage, annotation, filtration, prioritization, and visual inspection of variants

into a powerful yet simple-to-use graphical interface that is designed for users with all levels of computational expertise, with demonstrated utility in both clinical and research settings.

We are actively developing a version of the software that is capable of distributing queries across private computational clusters or across multiple instances running in the cloud, further increasing the performance and scalability of the platform. We also intend to work with third-party developers to create additional MedSavant Apps that leverage the platforms resources to deliver cutting-edge statistical and visual analytics of genomics data.

1. Danecek, P. *et al.* The Variant Call Format and VCFtools. *Bioinformatics* **27**, btr330–2158 (2011). URL http://dx.doi.org/10.1093/bioinformatics/btr330.

2. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010). URL http://dx.doi.org/10.1038/nature09534.

3. Nielsen, C. B., Jackman, S. D., Birol, I. & Jones, S. J. ABySS-Explorer: visualizing genome sequence assemblies. *IEEE transactions on visualization and computer graphics* **15**, 881–888 (2009). URL http://dx.doi.org/10.1109/tvcg.2009.116.

4. Fiume, M. *et al.* Savant Genome Browser 2: visualization and analysis for population-scale genomics. *Nucleic Acids Research* **40**, W615–W621 (2012). URL http://dx.doi.org/10.1093/nar/gks427.

5. Hillman-Jackson, J. *et al.* Using Galaxy to Perform Large-Scale Interactive Data Analyses. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 10** (2002). URL http://dx.doi.org/10.1002/0471250953.bi1005s38.

6. Preston, M. D. *et al.* VarB: a variation browsing and analysis tool for variants derived from next-generation sequencing data. *Bioinformatics* **28**, 2983–2985 (2012). URL http://dx.doi.org/10.1093/bioinformatics/bts557.

7. Teer, J. K., Green, E. D., Mullikin, J. C. & Biesecker, L. G. VarSifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics (Oxford, England)* **28**, 599–600 (2012). URL http://dx.doi.org/10.1093/bioinformatics/btr711.

8. Genetic association software for next-generation sequencing. http://www.goldenhelix.com/SNP$_V ariation/. Accessed : 2010 - 10 - 10.

9. Ślezak, D. & Eastwood, V. Data warehouse technology by infobright. In *SIGMOD '09: Proceedings of the 35th SIGMOD international conference on Management of data*, 841–846 (ACM, New York, NY, USA, 2009). URL http://dx.doi.org/10.1145/1559845.1559933.

10. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164 (2010). URL http://www.openbioinformatics.org/annovar/.

11. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).

12. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* **4**, 1073–1081 (2009). URL http://dx.doi.org/10.1038/nprot.2009.86.

13. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

14. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013). URL http://dx.doi.org/10.1038/nature11690.

15. Consortium, G. O. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* **32**, 258D–261 (2004). URL http://dx.doi.org/10.1093/nar/gkh036.

16. Robinson, P. N. *et al.* The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *American journal of human genetics* **83**, 610–615 (2008). URL http://dx.doi.org/10.1016/j.ajhg.2008.09.017.

17. Hamosh, A., Scott, A. F., Amberger, J., Valle, D. & McKusick, V. A. Online Mendelian Inheritance In Man (OMIM). *Hum. Mutat.* **15**, 57–61 (2000). URL http://dx.doi.org/10.1002/(sici)1098-1004(200001)15:1%3C57::aid-humu12%3E

18. Pico, A. R. *et al.* WikiPathways: pathway editing for the people. *PLoS biology* **6**, e184+ (2008). URL http://dx.doi.org/10.1371/journal.pbio.0060184.

19. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29**, 308–311 (2001). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC29783/.

20. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).

21. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research* **20**, 110–121 (2010). URL http://dx.doi.org/10.1101/gr.097857.109.

22. Davydov, E. V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* **6**, e1001025 (2010).

23. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).

24. Pruitt, K. D., Tatusova, T., Klimke, W. & Maglott, D. R. NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* **37**, D32–36 (2009).

25. Karolchik, D. *et al.* The UCSC Genome Browser Database. *Nucleic acids research* **31**, 51–54 (2003). URL http://view.ncbi.nlm.nih.gov/pubmed/12519945.

26. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Meth* **7**, 575–576 (2010). URL http://dx.doi.org/10.1038/nmeth0810-575.

27. Mostafavi, S., Ray, D., Farley, D. W., Grouios, C. & Morris, Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biology* **9**, S4+ (2008). URL http://dx.doi.org/10.1186/gb-2008-9-s1-s4.

28. Yandell, M. *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Res.* **21**, 1529–1542 (2011).

29. Srour, M. *et al.* Mutations in C5ORF42 cause Joubert syndrome in the French Canadian population. *Am. J. Hum. Genet.* **90**, 693–700 (2012).

30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). URL http://dx.doi.org/10.1093/bioinformatics/btp352.

31. Li, H. & Durbin, R. Fast and accurate short read alignment with BurrowsWheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). URL http://dx.doi.org/10.1093/bioinformatics/btp324.

Acknowledgements Put acknowledgements here.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to the Savant Development Team. (email: support@genomesavant.com).

Figure 1 Interface for specifying searches. Queries can be constructed based on both genotype and phenotype criteria.

Figure 2 Visualizing variant data at various resolutions. From top to bottom, a karyotypic heatmap shows the distribution of variants across chromosomes; a tabular view lists variants that pass the search criteria; the Inspector displays highly detailed information pertaining to the selected variant (left) and nearby genes (right); genome browser view of read alignments surrounding selected variant.

Figure 3 The MedSavant App Store provides access to third-party extensions to the platform. Apps can add complex search criteria, extend inspectors, or add novel visualizations or analyses. Clinically-oriented Apps are made available in the MedSavant Clinic.

Figure 4 Mendel expression used to identify harmful variants in C5orf42 (top), an independent discovery of the genes involvement in causing Joubert syndrome in 9 French-Canadians. The expression yielded variants in this gene only (bottom).

Table 1: Data stored by MedSavant. Information stored for patients and variants can be extended. External datasets can be connected through Apps available on the MedSavant App Store.

Genome	Patient	Annotations	Apps
Alternate Allele	Patient ID	Allele Frequencies (TGP[2], NHLBI[14])	WikiPathways[18]
Chromosome	Affected Status	Catalogues (dbSNP[19], COSMIC[20])	
Position	Gender	Conservation (PhyloP[21], GERP[22])	
Genotype	Family	Function (FATHMM[23])	
Quality	Father	Genes (RefSeq[24], UCSC[25])	
Variant Type	Mother	Harmfulness (Polyphen-2[13], SIFT[12], MutationTaster[26])	
Zygosity	Phenotype(s)	Ontology (GO[15], HPO[16], OMIM[17])	
more...	more...	more...	

Table 2: Sample information and result summary for the 16 FORGE disorders independently discovered using Mendel.

Disorder	Samples	Gene	Mendel Results
AD retinitis pigmentosa	2	RPE65	casual gene is 1 of 4 results
AR pontocerebellar hypoplasia with cortical atrophy	4	RARS2	causal gene is 1 of 4, but only nonsynonymous compound het
Floating Harbour syndrome	8	SRCAP	causal gene is only result
French Canadian Joubert	9	C5ORF42	causal gene is only result
French Canadian Joubert	3	TMEM231	causal gene is 1 of 4 results, only stopgain
Hadju-Cheney	5	NOTCH2	causal gene is only result
Hawk-Junction Microcephaly	1	WDR62	causal gene is 1 of 2
Hereditary leg dominant quadriparesis	3	DDHD2	causal gene is 1 of 45 results, but only compound het/hom
Hutterite CASS	1	SLC39A8	causal gene is only result
Hutterite Syndromic ID	1	THOC6	causal gene is 1 of 2
Mandibulofacial Dysostosis with Microcephaly	4	EFTUD2	causal gene is 1 of 2, but only frameshift/stopgain
MIC-CAP	5	STAMBP	causal gene is 1 of 3 results
MPPH-CM	5	CCND2	causal gene is only result
Multiple Intestinal Atresia	3	TTC7A	causal gene is only result
Sensory neuropathy	2	GPR172A /SLC52A2	causal gene is 1 of 2, but only homozygous