Saphetor SA varsome.com



Unlocking the power of variants through globalized discovery and prioritization

Precision medicine company Saphetor has built the world's largest community and knowledge base for human genome variant identification, annotation, and interpretation. Through its open access platform and its clinical and application programming interface (API)-powered bespoke implementations, Saphetor helps researchers and clinicians accelerate understanding, diagnosis, and treatment of genetic conditions.

Saphetor, a precision medicine company with a global footprint and headquartered in Lausanne, Switzerland, is building the world's most comprehensive community and knowledge base for fast and accurate disease-associated variant identification, annotation, and interpretation. At the core of its mission is the company's platform, VarSome, a proprietary, decentralized, and high-performance genomically-indexed data management system that uses automated classification algorithms and artificial intelligence (AI) to unlock the information contained in next-generation sequencing data generated from whole genomes, exomes, and gene panels (Fig. 1). VarSome is rapidly emerging as the global standard platform for human genome data, with a network of more than 400.000 professional users on six continents and more than 130 standardized and continuously updated major datasets freely searchable online (Fig. 2).

VarSome is highly versatile, with open access and bespoke implementations designed to address the needs of clinicians and researchers from both academia and industry, all working toward the common goal of using genetic knowledge to diagnose and predict disease to improve outcomes.

"At Saphetor, backed by our world-class and multi-disciplinary team of life scientists and software engineers, we believe in the value of building a community that helps us reach our goal through continuous, network-driven feedback loops," said Andreas Massouras, Founder and CEO of Saphetor. "More specifically, we have designed a platform to bring together researchers in academic labs, clinical labs, and biopharma who are in search of optimizing the discovery and prioritization of variants contained in the vast amount of genomic data continuously being generated on a global scale."

A trove of information waiting to be unpacked

The publication of the first draft of the human genome in 2001 ushered in a new era of discovery on the role of genetic variation in disease. To date, hundreds of variants have been linked to dozens of conditions, including a number of rare diseases, but the genetic underpinnings of thousands of other diseases remain to be elucidated. The early hope for a straight and rapid path to solving the genetic component of most diseases based on the ability to directly interrogate the genome turned out to be more elusive than anticipated.



Fig. 1 | VarSome—streamlining genetic variant identification, annotation, and interpretation. VarSome maintains, harmonizes and integrates more than 130 genomic databases and over 32 million publications, along with unique user-generated annotations and links, to create one of the most comprehensive genomic-variant databases in the world. The platform's industry-leading variant annotation and classification tools are helping clinicians and researchers around the world work much more efficiently with next-generation sequencing data.

On the one hand, the biology of the genome has revealed itself to be more intricate than anticipated due to the high degree of inherent structural variability of the genome and the complex interactions among different genetic components. And on the other hand, the sheer explosion of information has resulted in computational challenges associated with increased data siloing and with handling large and diverse datasets.

Solving both the biological and the computational challenge requires solutions that can simultaneously organize the exponentially growing amounts of new biological information and make it actionable. Efforts to address the biological problem through increased next-generation sequencing (NGS) approaches, including whole genome sequencing (WGS) and whole exome sequencing (WES) has resulted in the emergence of hundreds of data repositories, each with its own reporting and query formats, creating a formidable barrier to discovery.

Researchers or clinicians investigating a specific variant are faced with the onerous task of harmonizing different sources of information about a genetic event that in itself is complex to begin with. A variant's coding effect on different transcripts, its genomic location, its neighboring variants, genes it may affect, its population frequency, its function, its interactions with other omic levels, its relevant phenotypes, related literature and clinical studies, are all important pieces of information that together can provide a better understanding of disease relevance of a particular variant, but the fragmented nature of the current data landscape presents a significant challenge to apply this information.

"We were frustrated with the amount of time and effort it takes to find comprehensive information on gene variants and the inconsistency of it in the fragmented landscape of human genomics, and so we decided to act on it," said Massouras. "We set it as our goal to develop a platform that could be both a powerful annotation tool and a search engine for human genomic variants, enabling the sharing of knowledge on specific variants."

VarSome—a new gold standard in genomic data management and interpretation

Saphetor's goal was to develop an interrogable and user-augmentable but centralized platform for sharing and analyzing pathogenic and benign variants detected in genomic studies worldwide. The result is VarSome, a search and sequence annotation engine for human genomic variation that enables users to analyze variants in their genomic context, retrieve relevant related information from multiple databases, and predict potential disease associations, all on one integrated platform. VarSome is a platform capable of harmonizing disparate sources of structured and unstructured information, including scientific publications, gene, protein and other databases, to mention a few. Genomic indexing of the billions of data points results in a streamlined molecular database that can be easily expanded, shared, and interrogated.

Saphetor has differentiated itself from others in the space by focusing on optimizing information extraction from unstructured sources. In particular, the company's machine learning (ML)-based AI approach to publication tagging, a complex task critical to generating the backbone of distilled evidence necessary to analyze new variants and evaluate their potential significance, is unparalleled in the industry. With over 32 million publications in its molecular database, VarSome provides the largest, high confidence biological contextualization tool for variant evaluation in the world, and is rapidly being adopted as the gold standard for variant annotation by central repositories such as ClinVar, the variant portal of the United States National Center for Biological Information (NCBI).

A second component of the VarSome platform that sets it apart from others is its variant classifier tool for evaluating the potential pathogenicity of known and novel variants. Variants can be associated with one or more phenotypes, and conversely, phenotypes can also be associated with one or more variants, making the evaluation of the clinical significance of any given variant a very complex task. It also translates into variants falling along a gradient that ranges between variants that are almost certainly pathogenic and those that are almost certainly benign.

The strength of VarSome's predictions comes from running its algorithms through its deep knowledge base extracted from existing biological and clinical databases, and from applying the rigorous standards for the interpretation of sequence variants defined by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP). The resulting classification into 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign' is determined using types of variant evidence that include population data, computational data, functional data, and segregation data. This classifier tool is fully controllable by the user through adjusting the weights imposed on the different rules and specifications embedded in the algorithm.

"Paraphrasing AI pioneer Daniel Willis, 'our genome knows much more about our medical history than we do'," said Massouras. "At Saphetor we believe we can accelerate the pace at which this knowledge is extracted and ultimately applied to improving disease detection, treatment and prevention."

One platform—many solutions

Saphetor's core platform, VarSome provides highly efficient variant identification, annotation, and interpretation with industry-leading performance and accuracy—annotation of up to 2,000 variants/ second, classification on 1 CPU threads, scanning of >100 billion data points, and interpretation of whole genomes in under one hour. Saphetor has developed several implementations of VarSome tailored to address specific needs of different user communities.



Fig. 2 | VarSome—powering a global network of genetic variant knowledge. An emerging global standard platform for human genome data, VarSome is the world's largest community and knowledge base for human genome variant data with more than 400,000 professional users. VarSome.com is responsible for helping users around the globe connect with each other to share vital information to support diagnostic efforts. This map shows the connections made at time of writing.

- VarSome.com—a universal, free to use implementation of VarSome designed to share information on novel variants and connect with others to collaborate on defining those variants further. VarSome.com enables anyone, anywhere to query—on an Internet search engine—the VarSome molecular database, instantly providing access to the most relevant and up-to-date information available about their variant of interest. True to the versatility of the platform, users can input search terms ranging from gene names to variant call format (VCF) files containing raw sequence data. VarSome.com is the main driver of knowledge creation around disease relevant variants through community building, the pillar of all other VarSome implementations.
- VarSome Clinical—an annotation platform for labs generating NGS, including WGS and WES data. VarSome Clinical enables fast and accurate variant discovery through annotation, and interpretation of whole genomes, exomes, and gene panels. Starting from an Illumina, MGI FASTQ or VCF file, VarSome Clinical offers a complete pipeline for annotation and classification, with an option to generate comprehensive clinical reports. VarSome Clinical is CE in vitro diagnostic-IVDcertified (to confirm that the device complies with European regulations) and US Health Insurance Portability and Accountability Act-HIPAAcompliant and is increasingly being embedded in clinical sequencing workflows to help clinicians reach faster and more accurate diagnoses and inform treatment decisions for genetic conditions. VarSome Clinical provides a tool for clinicians to dramatically improve outcomes for its patients while also more directly contributing to the larger body of knowledge on clinically relevant variants.
- VarSome API—a bespoke implementation of VarSome that allows pharma and biotech companies to gain full access to Saphetor's molecular database through its high-performance variant

annotation application programming interface (API). Through a simple REST API, data requests are returned as JSON, a setup that allows seamless integration with Python, R, Perl, Java, or any other programming language. VarSome API removes the need for companies to establish their own data mining and curating teams—through VarSome API they can gain instant access to the largest, most comprehensive, and most up-to-date database currently available. VarSome API is the tool of choice for partners seeking to accelerate the development of their therapeutics and diagnostics programs in genetic diseases.

VarSome provides a comprehensive and robust variant interpretation engine, uniquely enhanced by the network effect of the platform. Building on its genomically indexed molecular database, Saphetor has the ability to power research and development of therapeutics and diagnostics as well as form the backbone of a clinical NGS analysis pipeline.

"As the shift from reactive to proactive medicine continues to accelerate, and the generation and use of genomic data becomes ubiquitous in healthcare, research, and consumer applications, the need for faster and more precise tools to harness this explosion of information becomes more pressing," said Massouras. "At Saphetor we are poised to continue powering this revolution in the life sciences and help improve outcomes for patients worldwide."

Andy Cosgrove
Chief Commercial Officer
Saphetor SA
Lausanne, Switzerland

- Tel: +1-508-244 -182
- Email: andy.cosgrove@saphetor.com