GensearchNGS: Interactive variant analysis

B. Wolf, P. Kuonen, *University of Applied Sciences Western Switzerland*
T. Dandekar, *University of Würzburg, Germany*

Introduction

NGS data analysis is increasingly popular in the diagnostics field. This is thanks to advances in sequencing technologies which improved the speed, quantity and quality of the produced data. Due to those improvements, the analysis of the data requires an increasing amount of technical knowledge and processing power. Several software tools exist to handle these technical challenges involved in NGS data analysis. This poster shows recent advancements in one of them, GensearchNGS, specifically in regards to variant analysis. We look at recent improvements made in variant calling, variant annotation and variant filtering made in GensearchNGS. The improvements range from improved functionality to vastly improved performance to lower the infrastructure requirements to perform NGS data analysis.

Variant calling

The variant calling algorithm in GensearchNGS has been completely rewritten, using a more efficient architecture.
The variant calling model has been based on the one used in Varscan 2. Thanks to a modular architecture which makes full use of multithreading, GensearchNGS is able to call variants over 15 times faster than Varscan 2, while providing in the same analysis results. Not only was the speed increased thanks to the new architecture, but there was also a reduction in memory consumption during the analysis. This reduces the infrastructure requirements to perform the analysis and adds more flexibility for the researcher to analyse the same sample multiple times with the same settings.
The speed increases where achieved by using a modular multithreaded architecture shown in figure 1, separating the computationally intensive processing steps from those doing input and output operations.
More detailed results about the new variant calling are currently being published. The variant calling algorithm will be released as part of a free tools collection called GNATY.

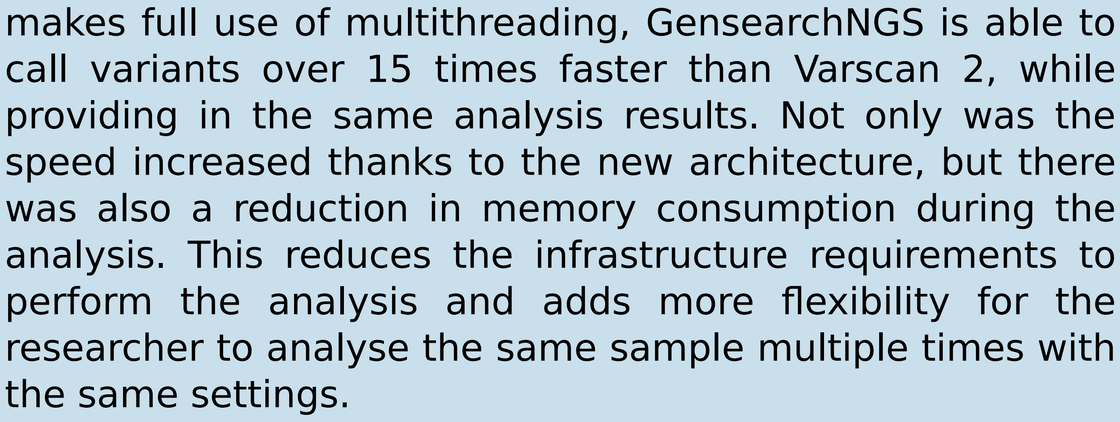

Figure 1: UML diagram of the new architecture of variant caller

Variant annotation

For the subsequent annotation of the called variants, various new data-sources have been integrated, such as Human Phenotype Ontology and the clinical predictions from Ensembl, which give the user more information about the clinical relevance of the called variants. An initial prototype of the integration of interactome data from different sources, such as CCSB or BioGRID, is also presented, further increasing the available information for variant effect prediction. The addition of annotation data has been accompanied by various optimizations, keeping memory requirements and analysis times stable.

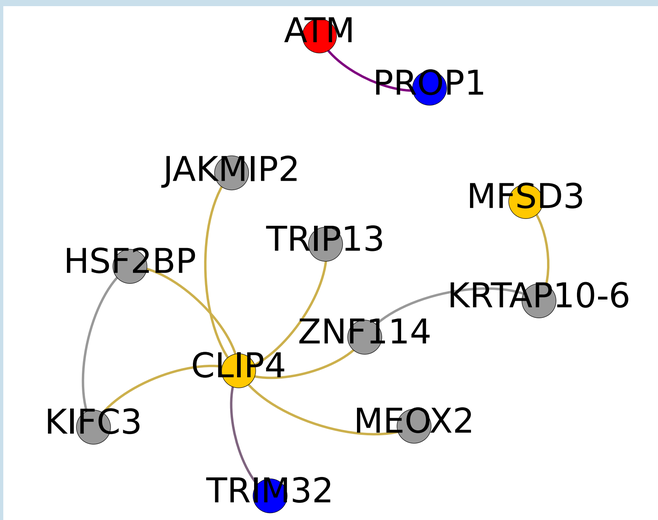

Figure 2: Interactome graph annotated with the variants of a patient

Interactive variant filtering

The interactive variant filtering makes it possible to filter variants, updating the displayed variant list on the fly depending on the chosen filters. This feature has been improved to run on machines with limited hardware.
New filters, related to the newly integrated annotations have been added, making it for example possible to quickly filter variants connected to a certain phenotype.

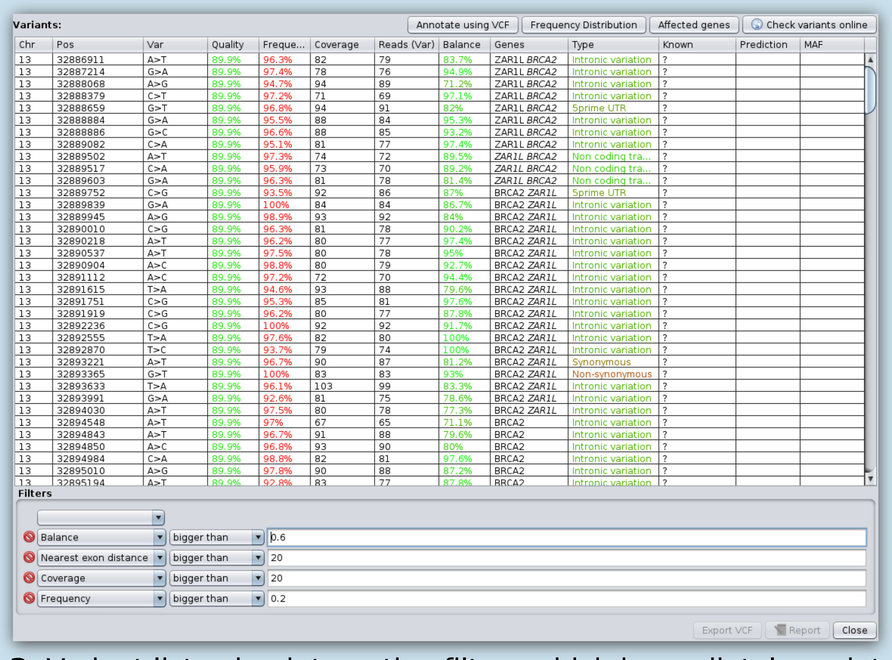

Figure 3: Variant list using interactive filters which immediately update the list

Similar improvements have also been made to the visualizer, allowing for a faster visualization requiring fewer resources, while integrating more data, such as the previously mentioned databases.

Contact: Beat wolf, beat.wolf@hefr.ch

Website GensearchNGS:
http://www.phenosystems.com/

PhenoSystems
GensearchNGS
MOLECULAR GENETICS MADE EASY

Hes·so FRIBOURG FREIBURG
Haute Ecole Spécialisée de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences Western Switzerland

BIOZENTRUM UNIVERSITÄT WÜRZBURG