

Safe variant annotation sharing across multiple laboratories

B. Wolf, P. Kuonen, D. Mazzoleni, J. Stoppani, *University of Applied Sciences Western Switzerland*
T. Dandekar, *University of Würzburg, Germany*
D. Atlan, *Phenosystems, Belgium*

Introduction

Annotating genetic variants heavily benefits from the usage of databases that already document the concerned variants. Due to the massive amount of possible variants and their complex effects, data sharing is essential to improve understanding of their consequences. Unfortunately many variants are only documented in private databases, either because of legal restrictions or because of the lack of time to publish the findings on public databases.

To solve this problem, we want to extract the information from where it already exists, which are the NGS data tools used in various laboratories. They often already integrate an internal variant database. We propose to augment those tools in a way to make it possible to share the variant annotations anonymously and securely among other users of the software, while giving the users full control on how to share their data.

Methods

For our first prototype we choose to extend the graphical NGS data analysis software GensearchNGS, developed by Phenosystems SA, with the ability to share variant annotations across multiple laboratories. Our approach is based on an original concept we call Trusted-Friend-Computing (TFC). The basic idea of TFC is the ability to share resources through a network of trusted friends. To implement this concept in GensearchNGS, we extend the open source POP-Java programming language with the required functionalities. POP-Java is a Java language extension that dramatically eases development of distributed Java applications.

The TFC concept also allows us to explore the possibility to share computing resources across multiple laboratories, but, in the current project, we will focus on variant annotation sharing.

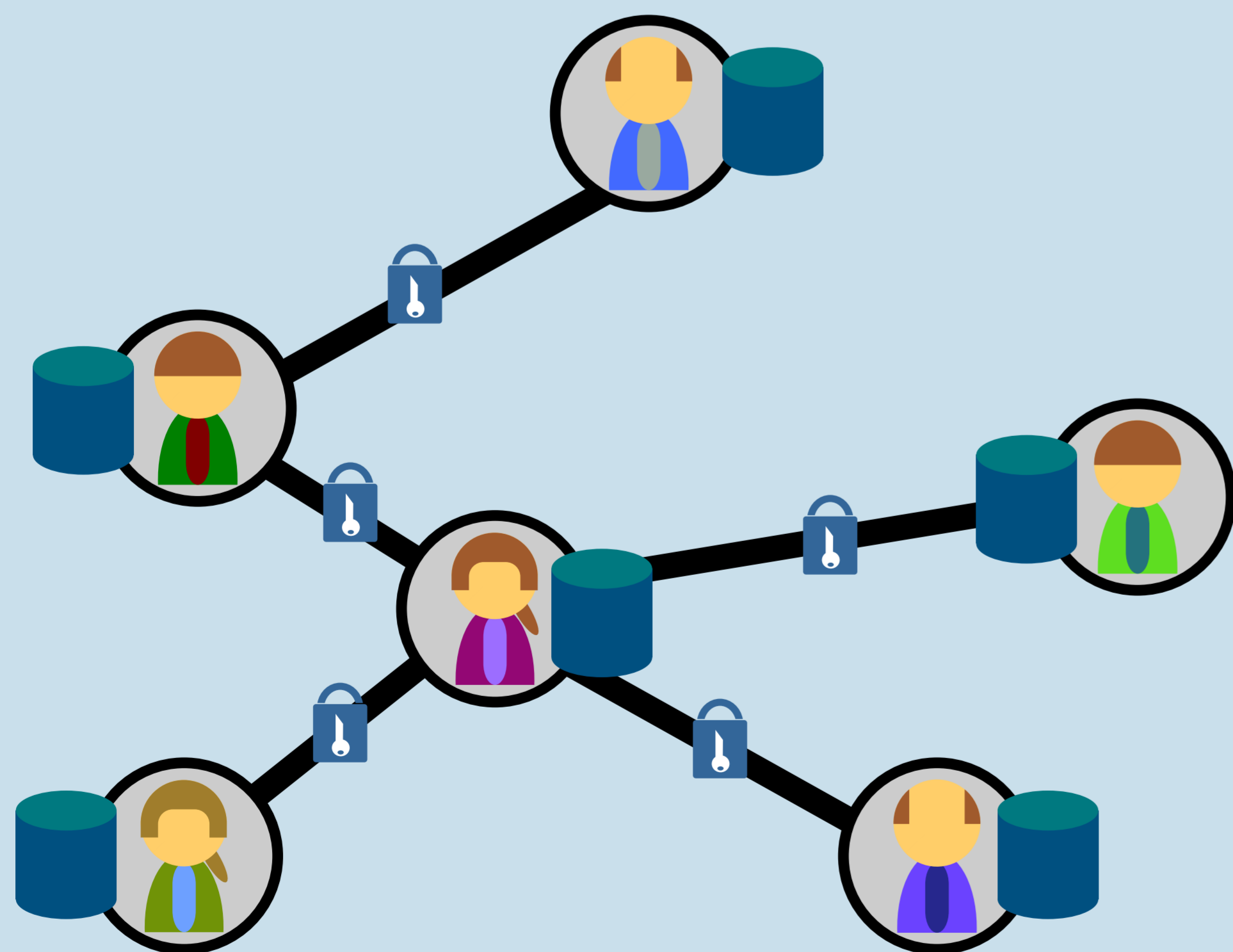


Figure 1: Example of an TFC network used to share variants

Results

Our initial prototype uses the development version of TFC, which is integrated into POP-Java, and adds a distributed variant database to GensearchNGS.

Using the already existing internal database in GensearchNGS, we added the ability to search for a specific variant among our friends. The core idea is that we do not transfer the content of the internal database, but only the information if a variant is known or not and certain (configurable) information like the classification.

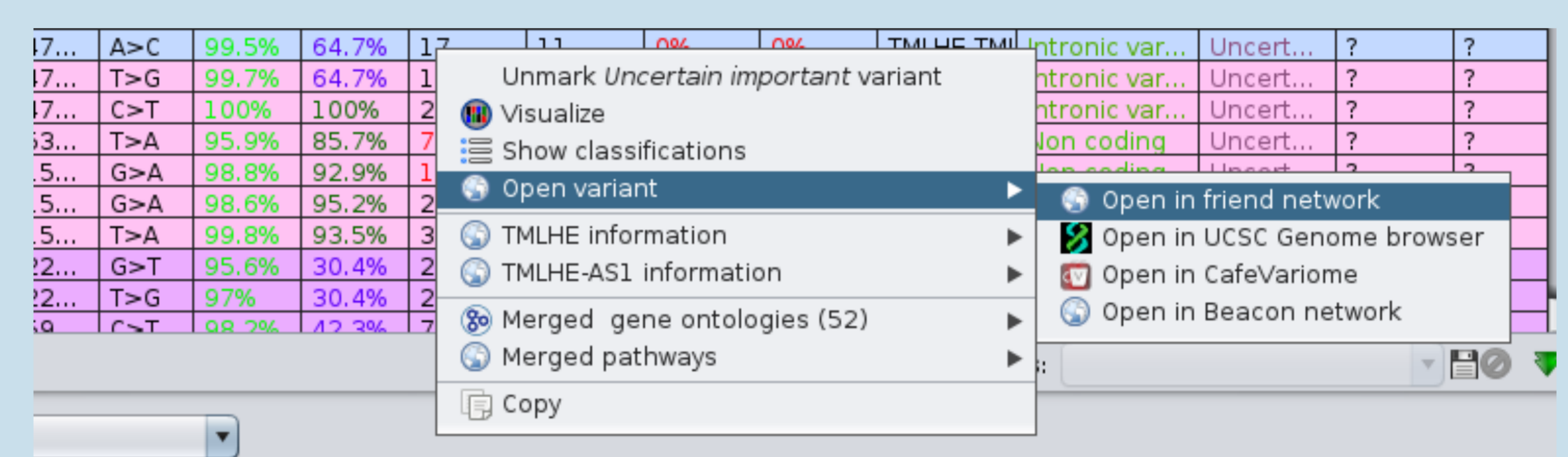


Figure 2: Transparent integration of the prototype into the variant list

Institute	Contact	Matchin...	Type
My institute	asdf@asdf.com	0	
Institute	beat@destop	1	Uncertain import...
Laboratory 1	info@lab1.com	1	Uncertain
Laboratory 2	info@lab2.com	1	Benign
Laboratory 3	info@lab3.com	0	
Laboratory 4	info@lab4.com	2	Uncertain
Laboratory 5	info@lab5.com	2	Uncertain

Figure 3: List of friendly laboratories and their classification of a specific variant

At the moment, the access right functionality is not yet implemented, as are many convenience features, but we can already show the potential of the concept. With a minimal amount of effort, laboratories can share overview information of their internal databases with selected other laboratories.

This concept can be expanded to many more use cases. Such as more in-depth analyses (GWAS, MAF etc.) or distributed computing use-cases like sequence alignment.

Conclusion

Sharing variant annotations is essential to improve the quality NGS data analysis. We acknowledge the technical and legal restrictions which hinder wide spread sharing, and propose thus a more controlled way to share variant annotations among a group of trusted laboratories. We hope that our work can serve as a stepping stone towards a more open variant data sharing culture.

We extended an existing open-source programming language to allow for an easier development of this type of applications, not only to share data, but also computing resources in general.

Grants : CTI no. 18781.1 PFES-ES