



User Manual

- Introduction3
 - What is GeneTalk?3
 - Who develops GeneTalk?3
 - Current features.....3
 - Who may use GeneTalk?3
 - How to cite3
- Account and Data management3
 - Registering an account and login3
 - Updating your profile.....4
 - Uploading variant files4
 - Data Sharing.....5
 - Messaging5
- Data Analysis.....5
 - Automatic processing of VCF files.....5
 - Filtering6
 - Setting up Gene Panels6
 - Annotating and Rating7
 - Search GeneTalk.....8
- Tutorial8
- Filtering Recommendations for sample collections9
 - Identify *de novo* mutations in Trios9

Introduction

What is GeneTalk?

GeneTalk is a web-based platform that serves as an expert exchange network for the assessment of personal and potentially disease relevant sequence variants. GeneTalk assists a clinical geneticist who is searching for information about specific sequence variants and connects this user to other users with expertise for the same sequence variant. If you have a patient with a rare unknown disease GeneTalk is the platform to find the second of its kind.

Who develops GeneTalk?

GeneTalk is an implementation of bioinformatics algorithms for filtering Next-Generation sequencing data by IT professionals. GeneTalk was started by Peter Krawitz and Tom Kamphans in 2011. In 2012 GeneTalk received funding from the Federal Ministry of Economics and Technology of Germany. Since then we started hiring software engineers and staff for sales, marketing and customer relations.

GeneTalk is a framework that warmly welcomes any kind of collaboration. If you are interested in contributing your expertise or if you want to join the team, please contact:

peter.krawitz@gene-talk.de and tom.kamphans@gene-talk.de.

Current features

- Variant specific knowledge base
- messaging system for variant annotations and gene panels for filtering
- Preprocessing engine that automatically annotates variants in a VCF file
- Genotype frequency filters comprising more than 1000 high quality genomes, and more than 5000 high quality exomes
- Interactive and expert curated gene panels

Who may use GeneTalk?

GeneTalk may be used by anyone who is interested in analyzing Next-Generation sequencing data sets. The only requirement is that you have an up to date internet browser and that you have the consent to analyze the genetic data.

How to cite

T. Kamphans, P.M. Krawitz, *GeneTalk: an expert exchange platform for assessing rare sequence variants in personal genomes*, Bioinformatics 2012

Account and Data management

Registering an account and login

To register a new account click on the sign up link in the login screen. Registered users may either use the login screen or the sign in field at the upper right corner to log into their account.

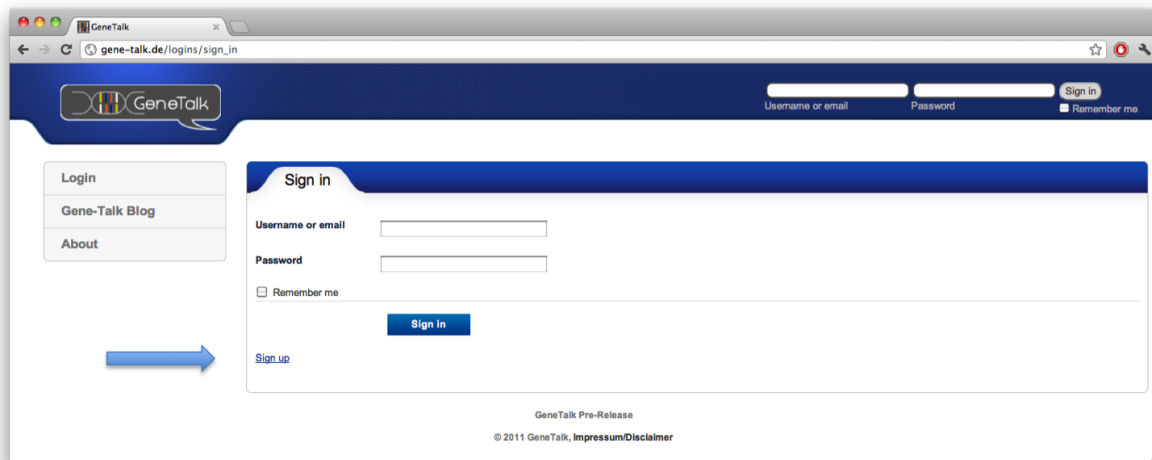


Figure 1: click on sign up to register for a new GeneTalk account. Registered users may either use the login screen or the sign in field at the upper right corner to log into their account.

Updating your profile

You can update your profile information by clicking the edit button on the My Profile screen. In the abstract field you should provide information about your field of expertise. The text of this field will be visible to other users, when they read your comments in GeneTalk.

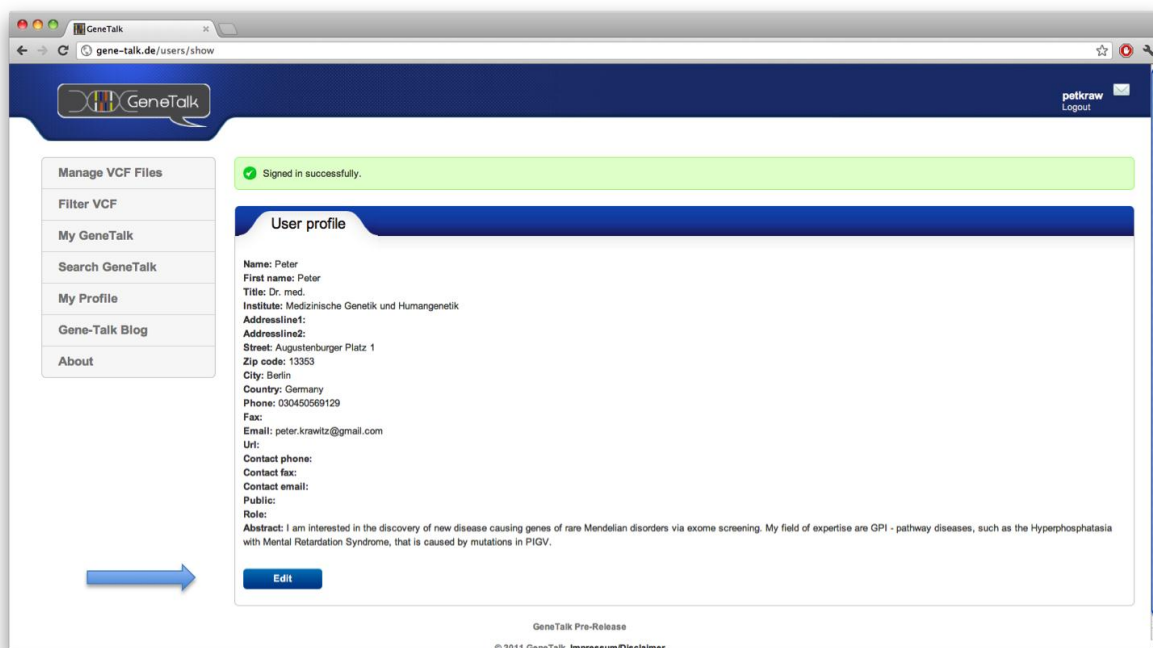



Figure 2: You can update your profile by clicking the edit button on the My Profile screen. In the abstract field you should provide information about your field of expertise. The text of this field will be visible to other users, when they read your comments in GeneTalk.

Uploading variant files


GeneTalk supports [VCF](#) format, which is the standard format for reporting variant data from Next-Generation sequencing experiments. GeneTalk requires that you use [GRCh37/hg19](#) as reference sequence. Please ask your sequencing facility to provide you with such a variant file!

You can simply upload your variant data on the My VCF Files screen. All uploaded VCF files will appear in the List Files section.

Data Sharing

In GeneTalk you can basically share everything, entire data sets, gene sets for filtering or linkage intervals. Whenever you see the sharing icon, , this indicates that you can share this kind of data with other users. When you click on this icon the sharing menu opens and you may specify with which user you intend to collaborate on that data simply by typing his user name.

Messaging

Your mailbox allows you to keep track of all your correspondence in GeneTalk. You will be notified about all data sharing requests. When you comment on an annotation in GeneTalk, other users might want to get into contact with you. They may either do so by directly replying to your comment or they might send you a mail through their messaging center. You can enter your messaging center through My Profile or simply by clicking on the mail icon .

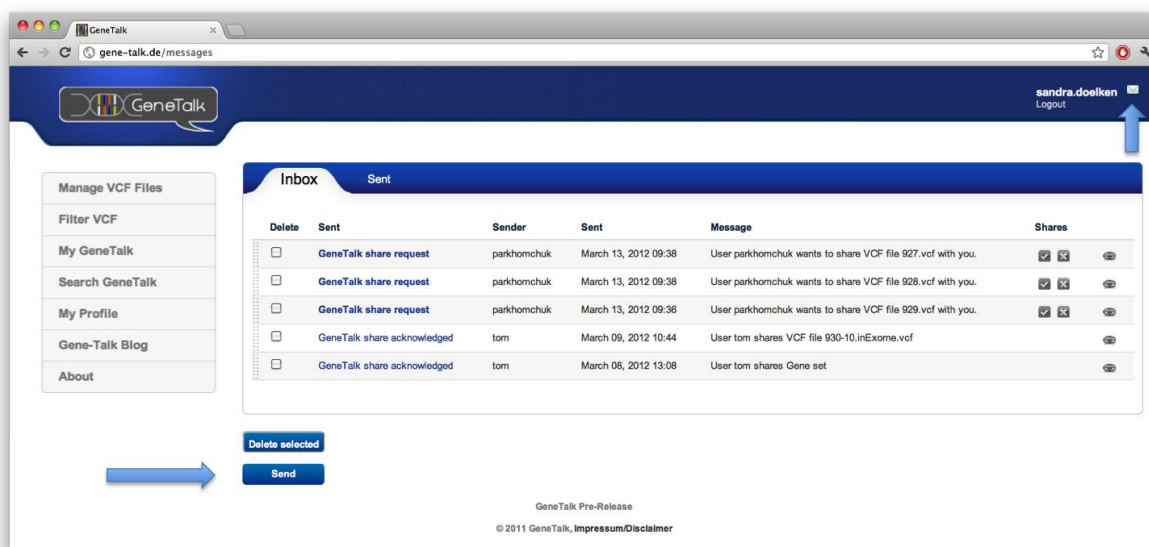


Figure 3: In your messaging center you can manage all your correspondence in GeneTalk. You will be notified about data sharing requests as well as comments from other users regarding you annotations.

Data Analysis

Automatic processing of VCF files

There are a few helpful processing steps GeneTalk performs automatically for you in the background after file upload. We check, whether your uploaded file is proper variant call format (VCF) and will inform you about any format inconsistencies in your uploaded file.

All genotypes in your variant list are aligned to the current release of dbSNP and the effect of a variant on the protein level is computed, based on transcript databases from the NCBI. This information is added to your VCF file and may then be used for further filtering.


If you upload exome or whole genome data GeneTalk estimates the gender of your sample by looking at the genotype of certain polymorphic SNP positions on the X chromosome. The according gender symbol is then displayed next to your sample:



Filtering

The Filter VCF menu provides you with several filtering options that you may adjust to reduce the number of relevant mutations in your variant file. The Functional Filter may be used to filter for e.g. non-synonymous variants, the Linkage Filter and Gene Set Filter may be used to restrict your variant list to a certain gene set of interest or to a linkage interval. The inheritance filter restricts your analysis to genotypes that are in accordance with a certain mode of inheritance. Note, currently the filter settings are only applied to a single VCF file. Combined filtering of more than one VCF file is not yet possible. This means: In order to filter for mutations that are in accordance with a dominant mode of inheritance you should check the box “heterozygous variants only”. In order to filter for mutations that are in accordance with a recessive mode of inheritance you should check the boxes “homozygous variants only” and “two or more heterozygous hits per gene, that also fulfill the other filtering settings”. Note the last criteria may be used to screen for candidates fulfilling a compound heterozygous mode of inheritance. Note, as this filter does not yet take any parental genotypes into account, it merely filters for two or more heterozygous variants in the same gene, after applying the functional filter, linkage filter, gene set filter, and frequency filter, but not the annotation filter.

The Frequency Filter sorts out all genotypes that occur more often than an adjustable cutoff frequency in the population. Currently we use the [1000 genomes data](#) as well as the [5000 exomes data](#) as background population. The Annotation Filter allows you to look for variants with expert annotations.

Your selected filtering settings are displayed at the bottom and you can start the filtering process by clicking on Filter. After a few seconds your results will be shown and you may display additional information such as the HGVS code by clicking on the  icon.

The basic filtering features are also explained in the introduction to GeneTalk video tutorial on the GeneTalk web site.

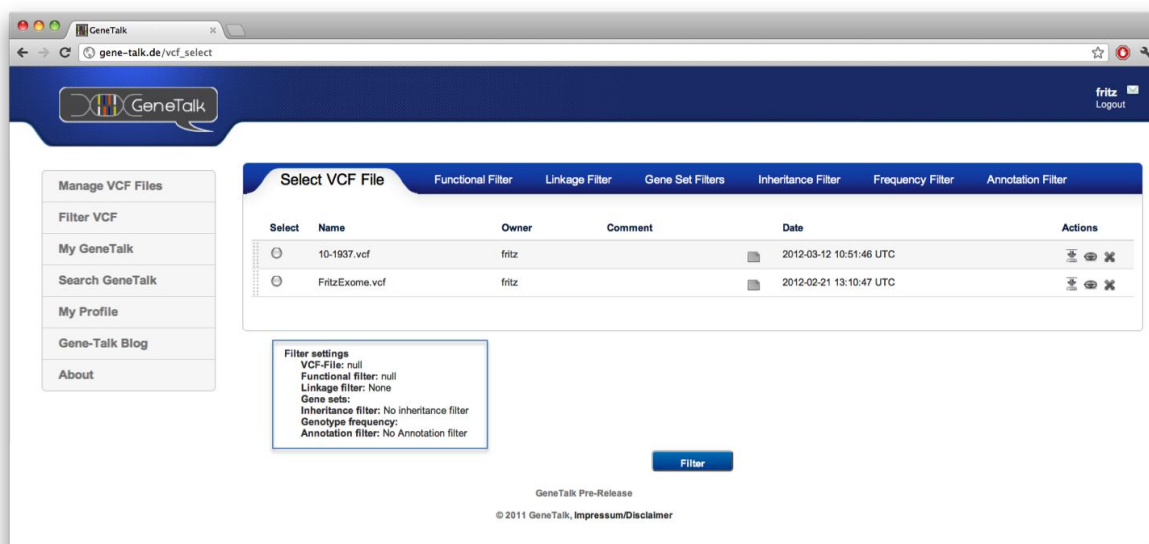


Figure 4: The Filter VCF menu helps you to reduce the number of relevant mutations in your variant list. The Functional Filter restricts your analysis to e.g. nonsynonymous variants only. With the Linkage Filter and the Gene Set filter you may focus on a specific linkage interval or a certain list of genes. The Inheritance Filter may be used to adjust for a certain mode of transmission. The Frequency Filter sorts out all variants that occur more often in the population than an adjustable cutoff frequency.

Setting up Gene Panels

In GeneTalk Gene Panels serve two purposes: You can use Gene Panels for filtering your variant data and also to define your field of expertise. If you declare a Gene Panel as your field of expertise other users

will see your user name in the annotation track for all their variants that lie in these genes or genomic intervals.

You can create a gene panel filter in My GeneTalk > Gene Panels, either by entering the official gene symbol or Entrez ID. In this case the exon boundaries of all reviewed transcripts of the Consensus cDNA database are used. We include the coordinates of the canonic splice acceptor and donor site. This means that we add 2 base pairs of the intronic sequence. If there is no CCDS transcript available, you have to edit the intervals of your interest by hand in the following way: (chr,start,end) by coordinates with respect to GRCh37/hg19.

You can also make your Gene Panel available to the GeneTalk community by publishing it. Other users may then subscribe your Gene Panel for filtering. They will also be notified about any change you make to your Gene Panel. By this means everyone stays up to date. There is a detailed tutorial about Gene Panels on the GeneTalk webpage and on youtube:

http://www.youtube.com/watch?feature=player_embedded&v=VlKj6vT5d4w

Annotating and Rating

In GeneTalk a genotype annotation is simply added by clicking the **Annotate** button and contains human expert knowledge about a specific genomic variant. An annotation can be a link to a genetic disorder or a link to a currently published scientific article in which the effect of this specific mutation is described. However, an annotation may also discuss technical issues of detecting artefacts at a specific locus or may be an assumption about a genotype-phenotype association. GeneTalk users may comment on annotations, and rate its trustworthiness and its potential to be disease causing. Genotype annotations that are rated to be disease causing by most GeneTalk users are included in the GeneTalk pathogenic annotation track. In Filter VCF all variants may be filtered for expert annotations.

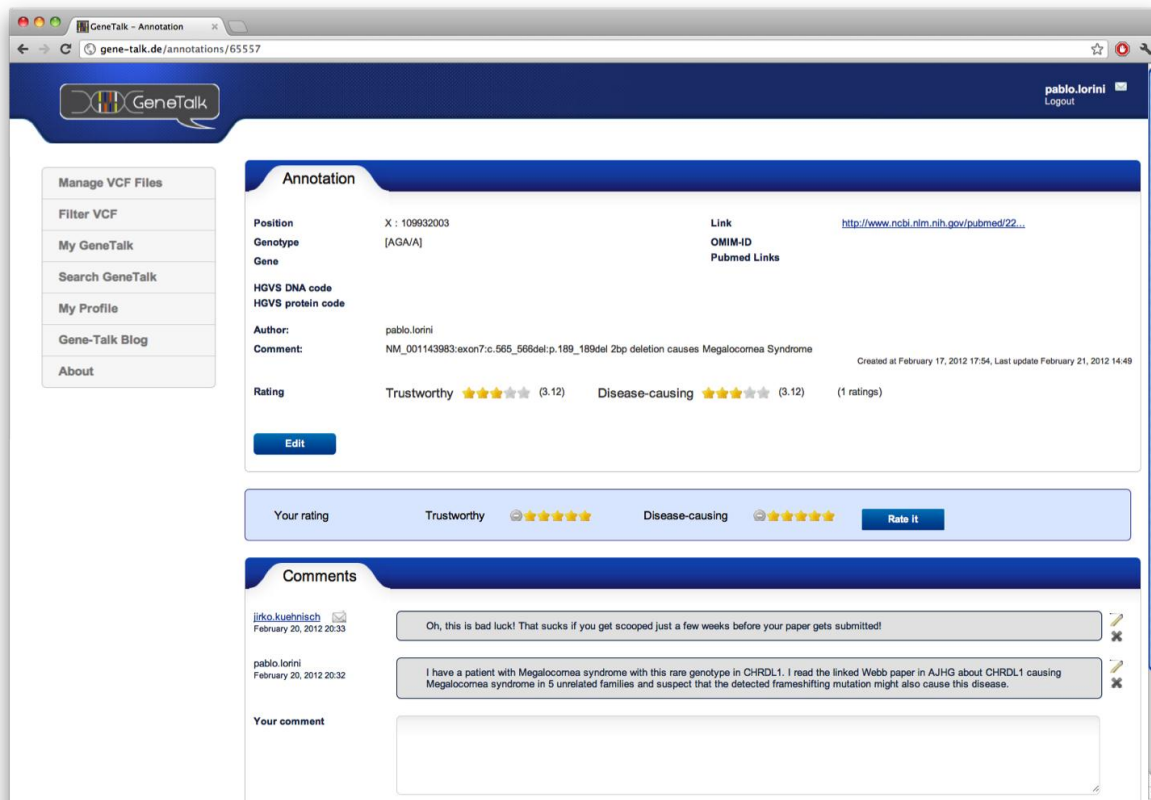



Figure 5: GeneTalk users may add locus-specific annotations that link e.g. human expert knowledge about a genetic disorder to a genotype. Users may comment on annotations as well as rank their trustworthiness and potential to be disease-causing.

Search GeneTalk

In Search GeneTalk you can look for gene or locus specific expert annotations. Detailed information about an annotation will be displayed when clicking the eye icon, .

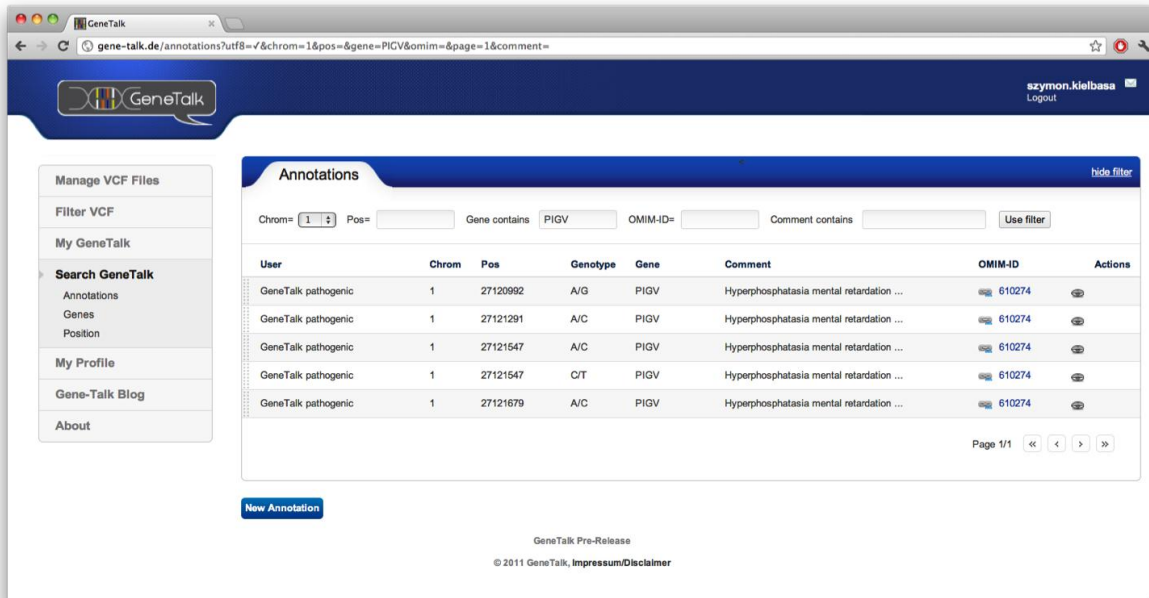



Figure 6: In Search GeneTalk you can look for gene or locus specific expert annotations. The annotation menu opens by clicking the eye icon, .

Tutorial

You can find online tutorials about “how to filter” and “how to use Gene Panels” on the GeneTalk webpage.

You can log into a demo account to reproduce everything shown in the tutorials and test all GeneTalk features. The dataset provided in this account is assembled from the 1000 genomes project with an additional pathogenic mutation.

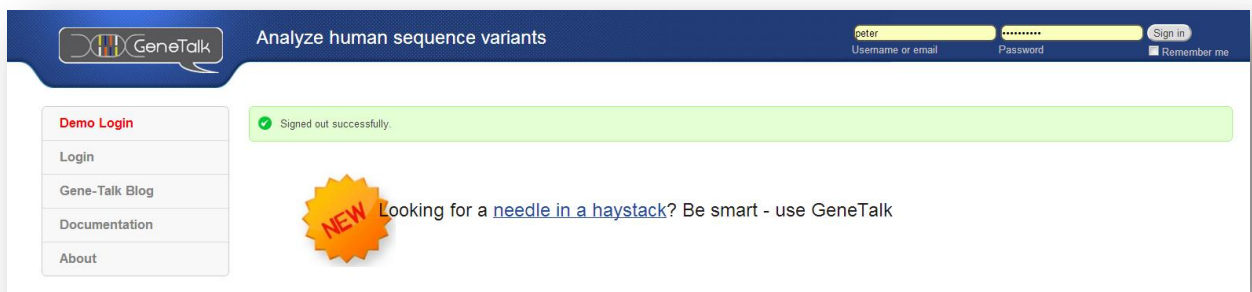


Figure 7: Log into the Demo account to find all the data presented in the tutorials.

Filtering Recommendations for sample collections

Identify *de novo* mutations in Trios

If you have variant data of an affected patient and its healthy parents, you should first define their relationship with the create collections tool in the manage files menu and set the status of being affected or not appropriately. Then the inheritance filtering menu allows you to reduce your sample collection to variants that follow a certain segregation pattern. For *de novo* mutations you should filter for dominant segregation. This will remove all variants from the collection that are in disagreement with a dominant pattern of inheritance. That means only variants that are heterozygous in the affected samples but not heterozygous in the unaffected relatives will remain in your filtered set. Be aware that this filtering step will not remove variants that are heterozygous in the patients from the collection unless you also provided genotypes for the unaffected relatives, as these variants might still indicate *de novo* mutations in the affected samples. If you would like to restrict your variant set to highly likely *de novo* mutations you should additionally use a coverage or quality cutoff.