Utilizing External Databases in Geneticist Assistant® NGS Workbench

April 2020

Megan McCluskey, Edward Bouton, Zhaorong Ma, James Todd

Introduction

Interpreting the significance of detected variants is a crucial aspect of research and clinical sequencing projects. Next generation sequencing applications such as whole exome sequencing can produce a large number of variant calls such that interpreting variants to generate a curated variant list is necessary. Numerous databases exist with information on variants including population frequencies, disease associations, calculated prediction scores, and more. Reviewing this information from various sources can be a time-consuming process.

Geneticist Assistant NGS Workbench assists with variant interpretation by providing access to multiple external databases in one location. Any imported variant can be reviewed in detail with the information from all databases displayed in a single view. Custom selected database fields can also be added to the variant table and can be used for filtering variants, facilitating the creation of prioritized variant lists.

Method

Download References

External databases are easily downloaded using Geneticist Assistant's built in Reference Downloader tool. This tool can be opened by clicking the "Ref" button at the top right of Geneticist Assistant.

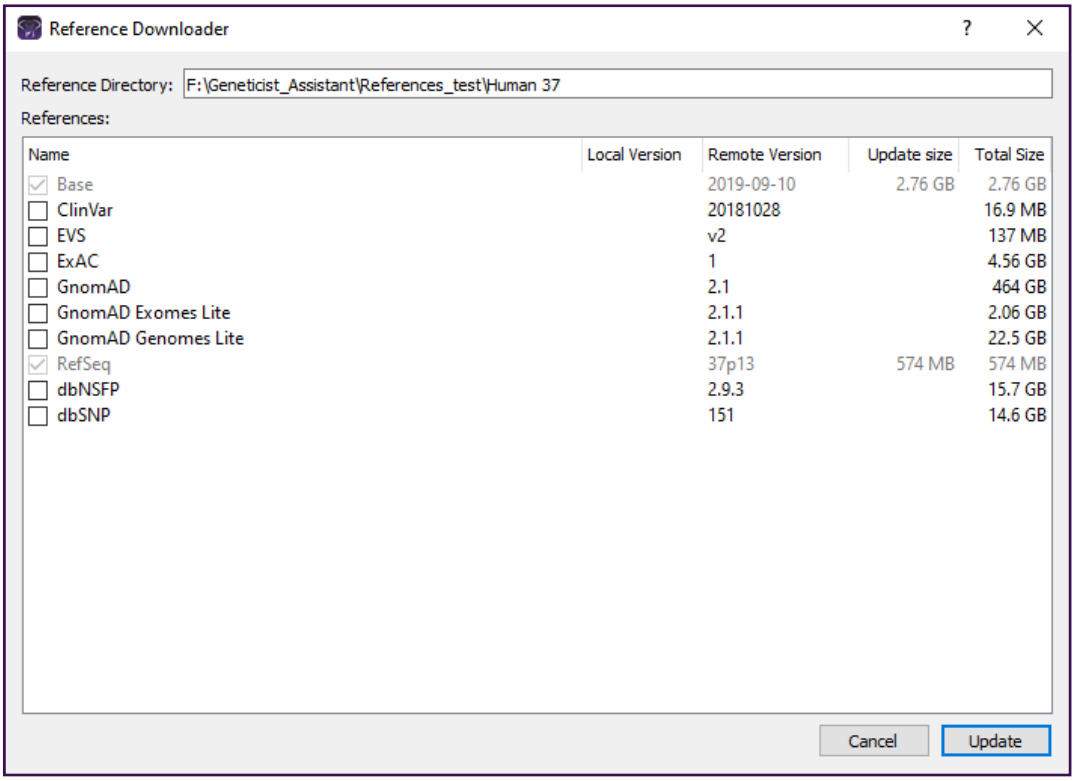
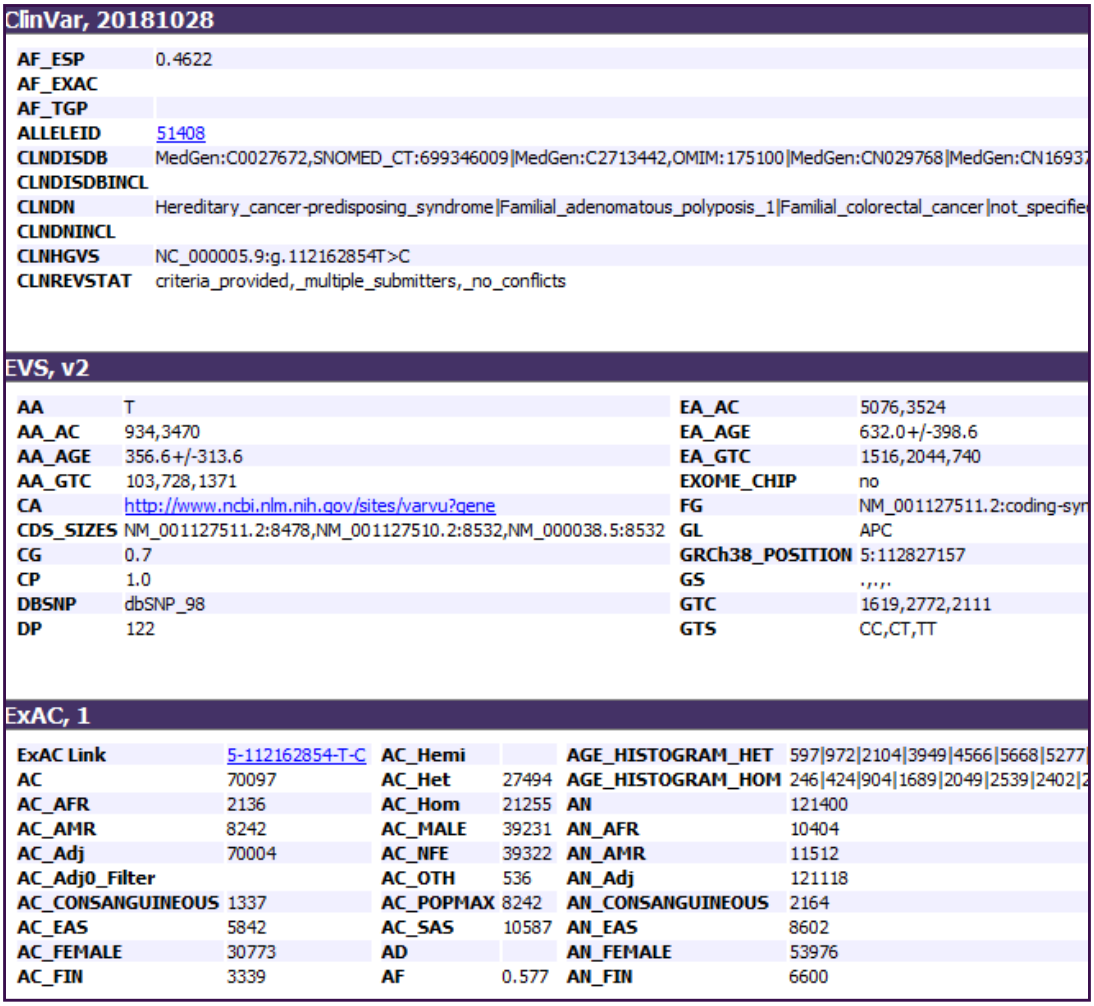


Figure 1: Reference Downloader tool – Available databases are displayed and can be selected for download to be included for reviewing variants.

With the Reference Downloader tool databases can be selected for download. The size of each database is listed so that users can ensure adequate space is available for the references before beginning download.

Review Database Information for a Variant

When reviewing a submitted sample in Geneticist Assistant, a Variant tab where variants can be reviewed in more detail can be opened by double-clicking on any variant. The Variant tab displays panes for each of the downloaded databases.



ClinVar, 20181028

AF_ESP	0.4622
AF_EXAC	
AF_TGP	
ALLELEID	51408
CLNDISDB	MedGen:C0027672,SNOMED_CT:699346009|MedGen:C2713442,OMIM:175100|MedGen:CN029768|MedGen:CN1693
CLNDISDBINCL	
CLNDN	Hereditary_cancer-predisposing_syndrome|Familial_adenomatous_polyposis_1|Familial_colorectal_cancer|not_specifie
CLNDNINCL	
CLNHGVS	NC_000005.9:g.112162854T>C
CLNREVSTAT	criteria_provided,_multiple_submitters,_no_conflicts

EVS, v2

AA	T	EA_AC	5076,3524
AA_AC	934,3470	EA_AGE	632.0+/-398.6
AA_AGE	356.6+/-313.6	EA_GTC	1516,2044,740
AA_GTC	103,728,1371	EXOME_CHIP	no
CA	http://www.ncbi.nlm.nih.gov/sites/varvu?gene	FG	NM_001127511.2:coding-syn
CDS_SIZES	NM_001127511.2:8478,NM_001127510.2:8532,NM_000038.5:8532	GL	APC
CG	0.7	GRCh38_POSITION	5:112827157
CP	1.0	GS	.,.,.
DBSNP	dbSNP_98	GTC	1619,2772,2111
DP	122	GTS	CC,CT,TT

ExAC, 1

ExAC Link	5-112162854-T-C	AC_Hemi		AGE_HISTOGRAM_HET	597|972|2104|3949|4566|5668|5277
AC	70097	AC_Het	27494	AGE_HISTOGRAM_HOM	246|424|904|1689|2049|2539|2402|2
AC_AFR	2136	AC_Hom	21255	AN	121400
AC_AMR	8242	AC_MALE	39231	AN_AFR	10404
AC_Adj	70004	AC_NFE	39322	AN_AMR	11512
AC_Adj0_Filter		AC_OTH	536	AN_Adj	121118
AC_CONSANGUINEOUS	1337	AC_POPMAX	8242	AN_CONSANGUINEOUS	2164
AC_EAS	5842	AC_SAS	10587	AN_EAS	8602
AC_FEMALE	30773	AD		AN_FEMALE	53976
AC_FIN	3339	AF	0.577	AN_FIN	6600

Figure 2: Database panes on the Variant tab report information for the variant from each downloaded database. Additional databases that may be shown include dbNSFP, GnomAD, and dbSNP.

Add Database Fields to the Variant Table

Fields from any database can also be added to the variant table that is shown on the Sample tab. This is done by right-clicking anywhere in the pane for the database and then selecting "Add Values to Variant Table."

The dialog that opens shows all the fields available for the database. To select the field to be added, double-click the row for the field. Changes can be made to the field configuration settings as needed. For instance, for fields that have multiple allele values it often works best to select to display only the matching allele. Similarly, for fields that include values for multiple populations, selecting only the value for the applicable population may be most useful. For fields where multiple values are included, selecting a single value is generally needed in order to use the field for filtering.

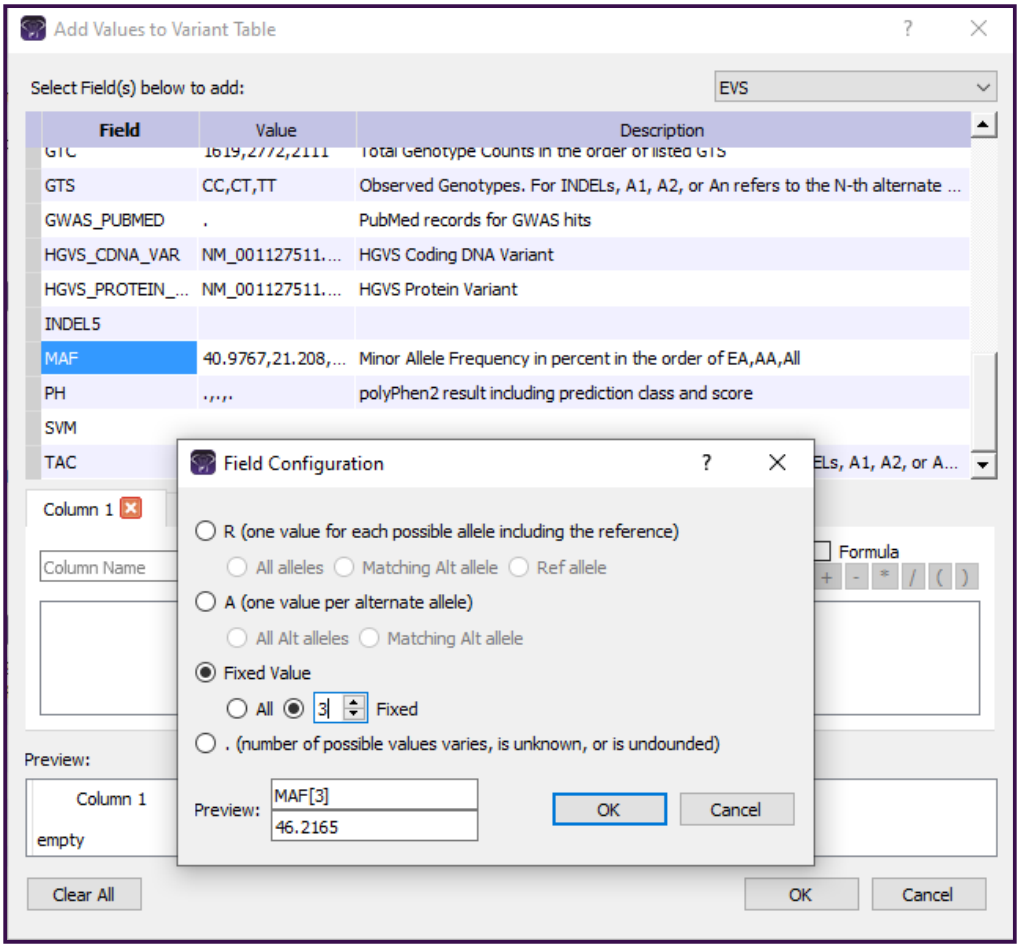


Figure 3: After double-clicking on the database field to be added you can specify the field configuration settings. Shown here is an example of a field that contains multiple values based on different populations. In this case the third value, the value for "All" populations, is selected.

Multiple fields from the same database can be added as needed by clicking the "+" button. Multiple fields can also be combined as a single column by using text or creating a formula.

Filter Variants

Once database columns have been added to the variant table variant filters can now be created based on these fields. While viewing the variant table on the Sample tab, open the "Filters" menu and select "Sample's Variants". This will open the Table Filters pane. In this pane all displayed columns of the table are listed and can be selected for filtering.

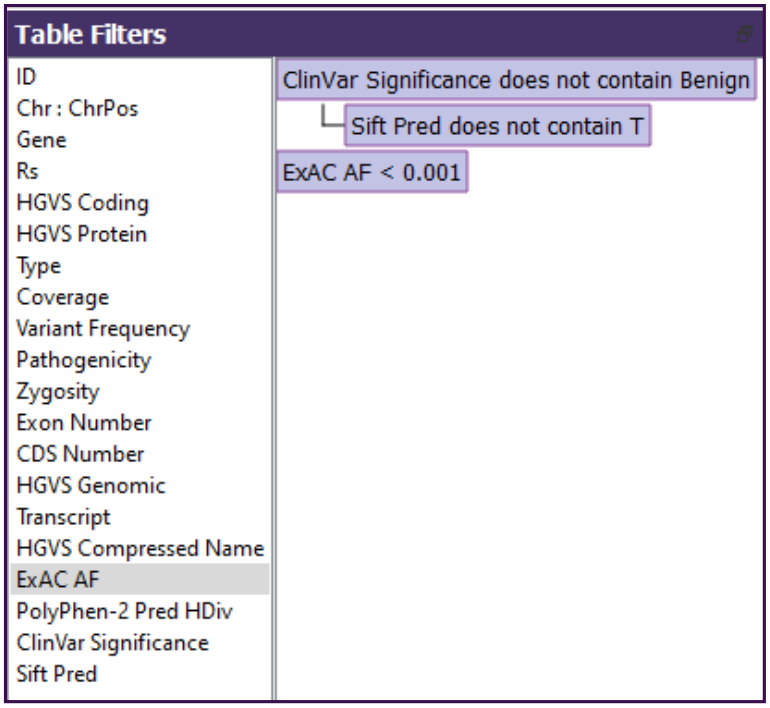


Figure 4: Variants can be filtered by database fields that have been added to the variant table. Multiple fields can be added as an "AND" or an "OR" combination. The above example shows a filter that will retain variants that have a ClinVar significance that does not contain Benign AND a SIFT prediction that does not contain T (tolerated) OR that have an ExAC allele frequency less than 0.001 (0.1%).

To select a field for filtering, click and drag the column name to the right side. A dialog opens where the parameters for filtering can be input. Click "OK" to add.

Additional fields for filtering can be added and the additional fields can be added as an "OR" or as an "AND" filter. To add as an "AND" filter, drag and drop the next field so that the cursor lies directly over the existing filter. Enter the parameters for the filter and click "OK" to add. When adding an additional field in this way, both filters must be met for a variant to be included.

To add an additional field as an "OR" filter, drag and drop the next field anywhere in the right side. Enter the parameters for the filter and click "OK" to add. When adding an additional field in this way, a variant will be included if any one of the filters are met.

Created filters can also be saved to be quickly applied in the future.

Conclusion

By displaying detailed information from multiple external databases in one location Geneticist Assistant makes variant review quicker and easier. Users can select which databases they prefer to download and can also add custom databases as needed. Database information can be reviewed in detail for each imported variant and custom selected fields can also be added as columns in the variant table to allow filtering by the database fields. This facilitates the process of narrowing down a large list of detected variants down to a prioritized list of potentially causal variants.