Identification of somatic and germline variants from tumor and normal sample pairs

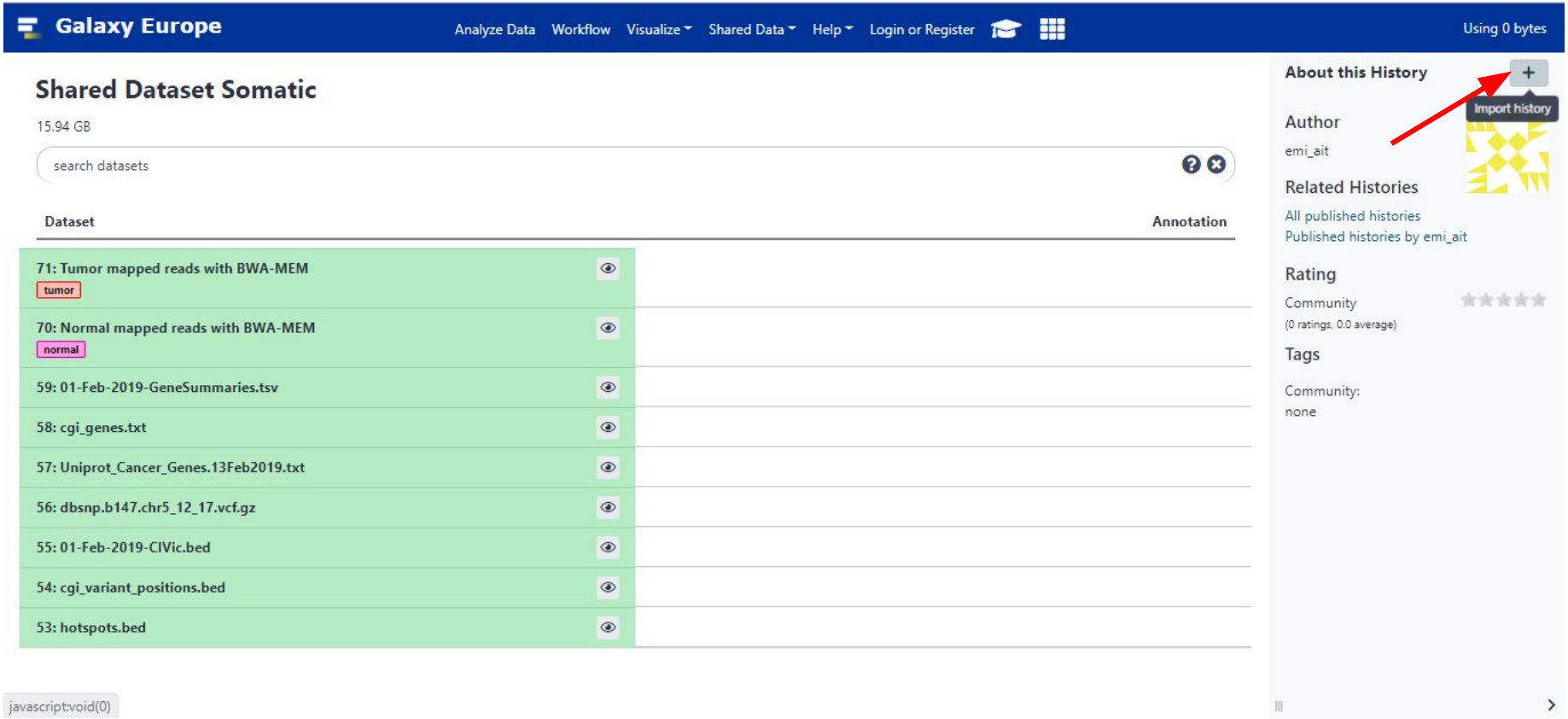
Somatic variants tutorial

Workflow

1. Mapped reads postprocessing
    a. Filtering on mapped reads properties
    b. Removing duplicate reads
    c. Left-align reads around indels
    d. Recalibrate read mapping qualities
    e. Refilter reads based on mapping quality
2. Variant calling and classification
3. Variant annotation and reporting
    a. Adding annotations to the called variants
    b. Reporting selected subsets of variants
    c. Generating reports of genes affected by variants
    d. Adding additional annotations to the gene-centered report

Starting from BAMs : Import Shared History

https://usegalaxy.eu/u/emi_ait/h/shared-dataset-somatic

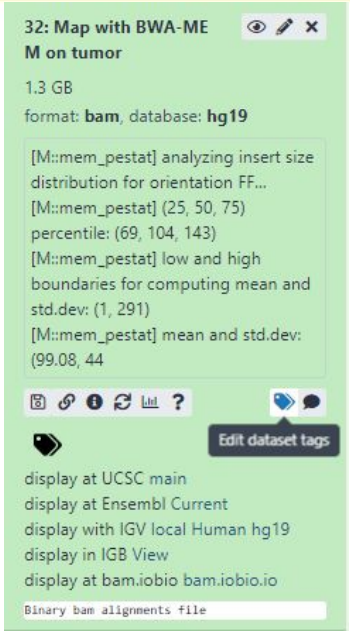


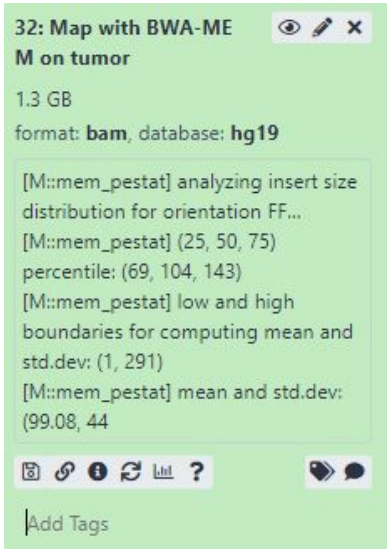
Prepare Data

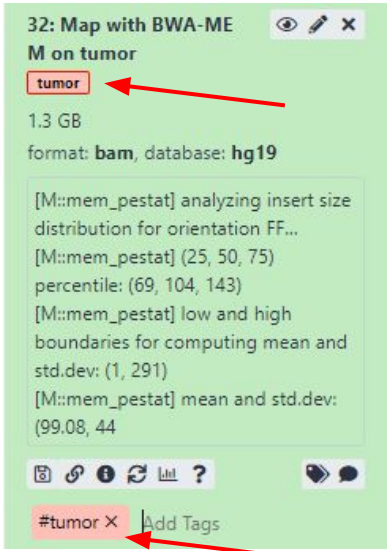
Tip: Adding a tag

- Click on the dataset
- Click on **Edit dataset tags**
- Add a tag starting with `#`

  Tags starting with `#` will be automatically propagated to the outputs of tools using this dataset.
- Check that the tag is appearing below the dataset name





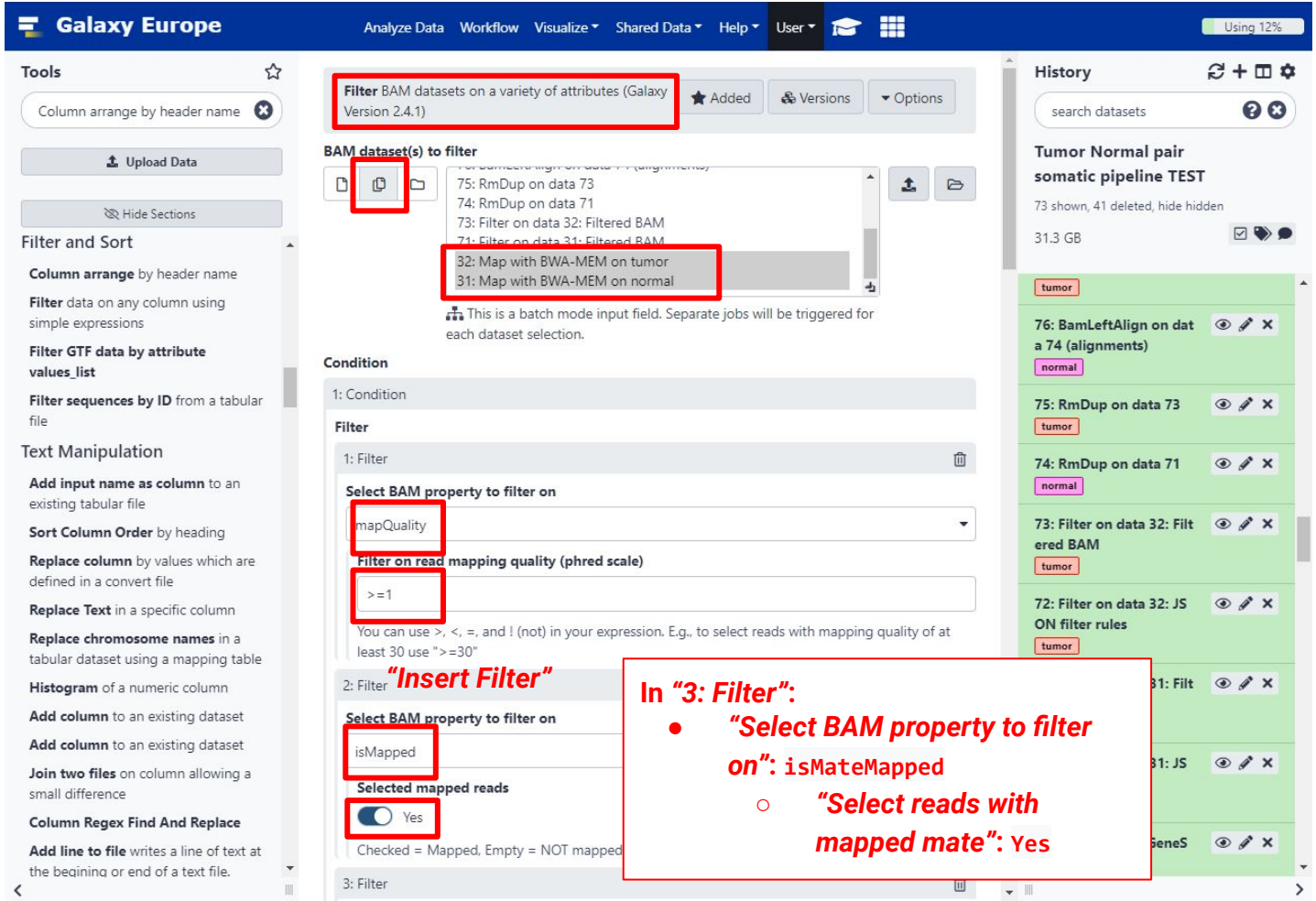


1. Mapped reads postprocessing

1. Mapped reads postprocessing

a. Filtering on mapped reads properties

Filtering for mapping status and quality



Filtering for mapping status and quality

There is not only one tool that can filter reads.

To Do : find another tool in Galaxy to perform the same operation

Filtering for mapping status and quality

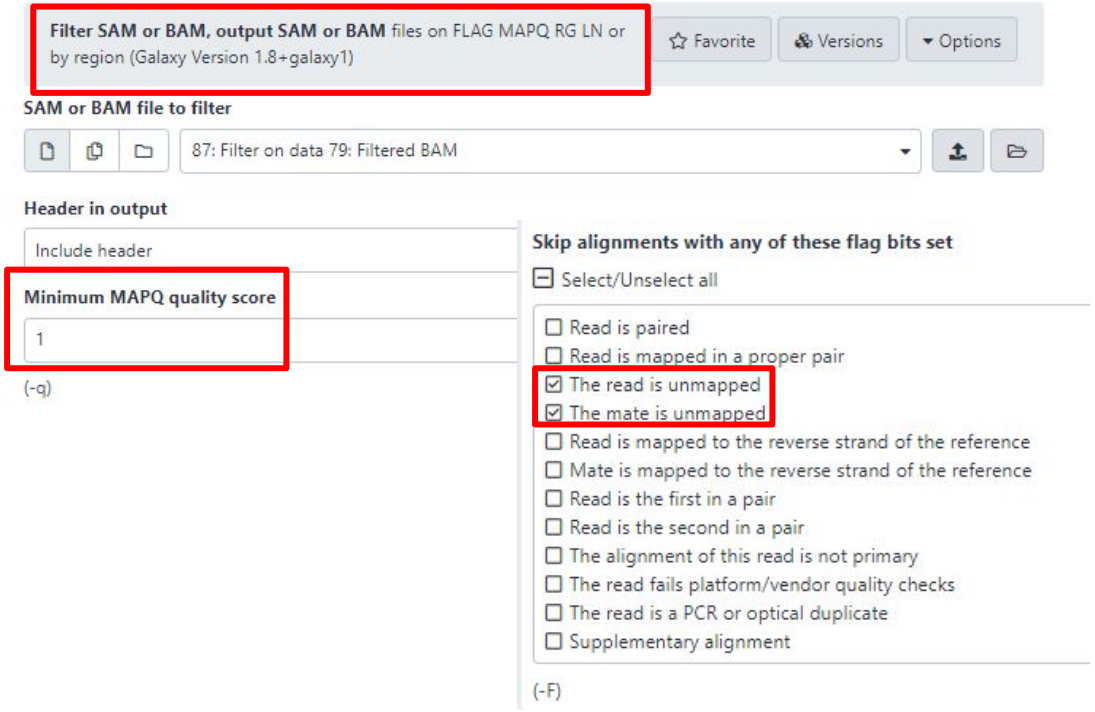
There is not only one tool that can filter reads.

To Do : find another tool in Galaxy to perform the same operation

Filter SAM or BAM, output SAM or BAM
based on samtools view

equivalent to

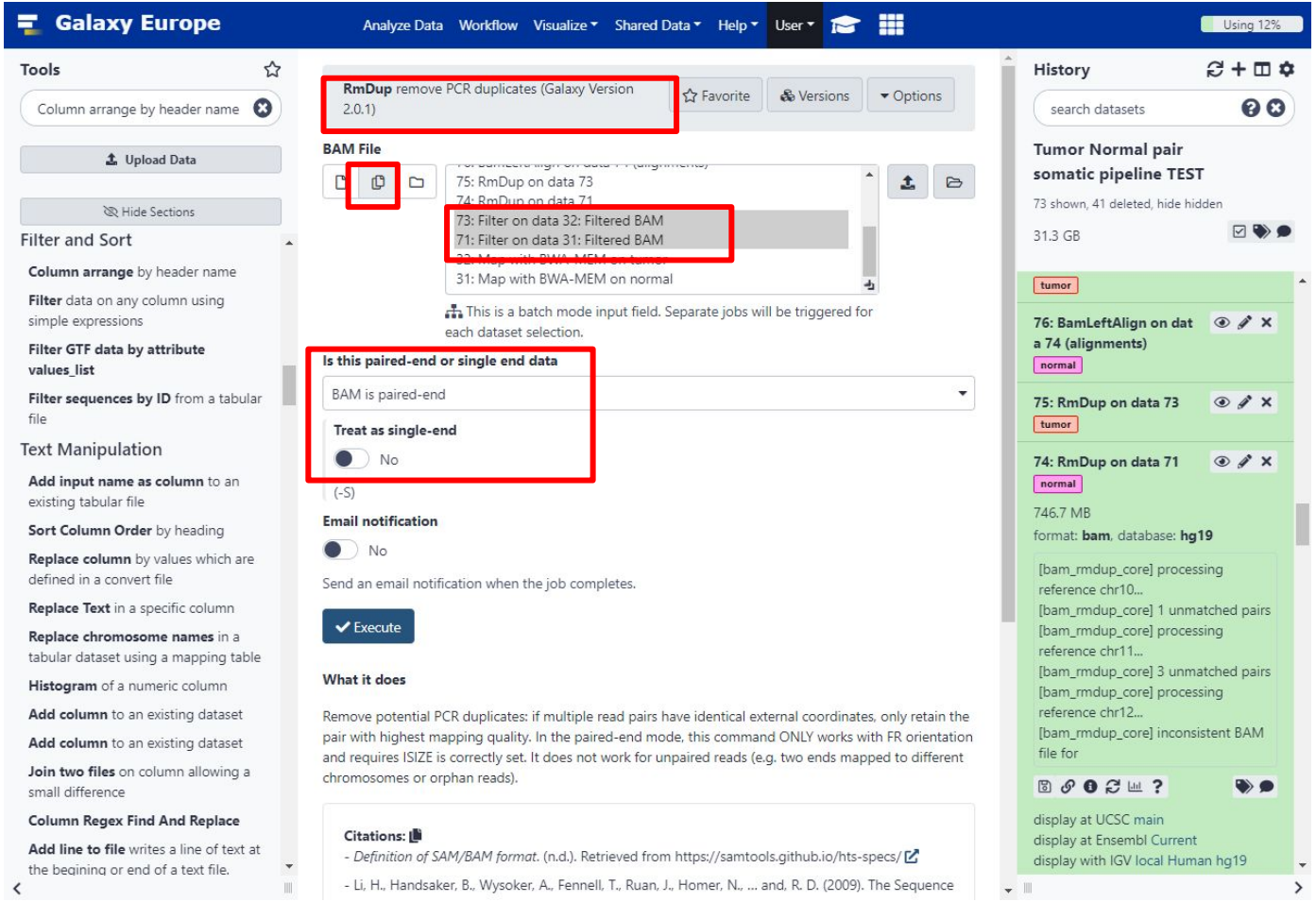
Filter BAM datasets on a variety of attributes
Based on bamtools filter



Mapped reads postprocessing

b. Removing duplicate reads

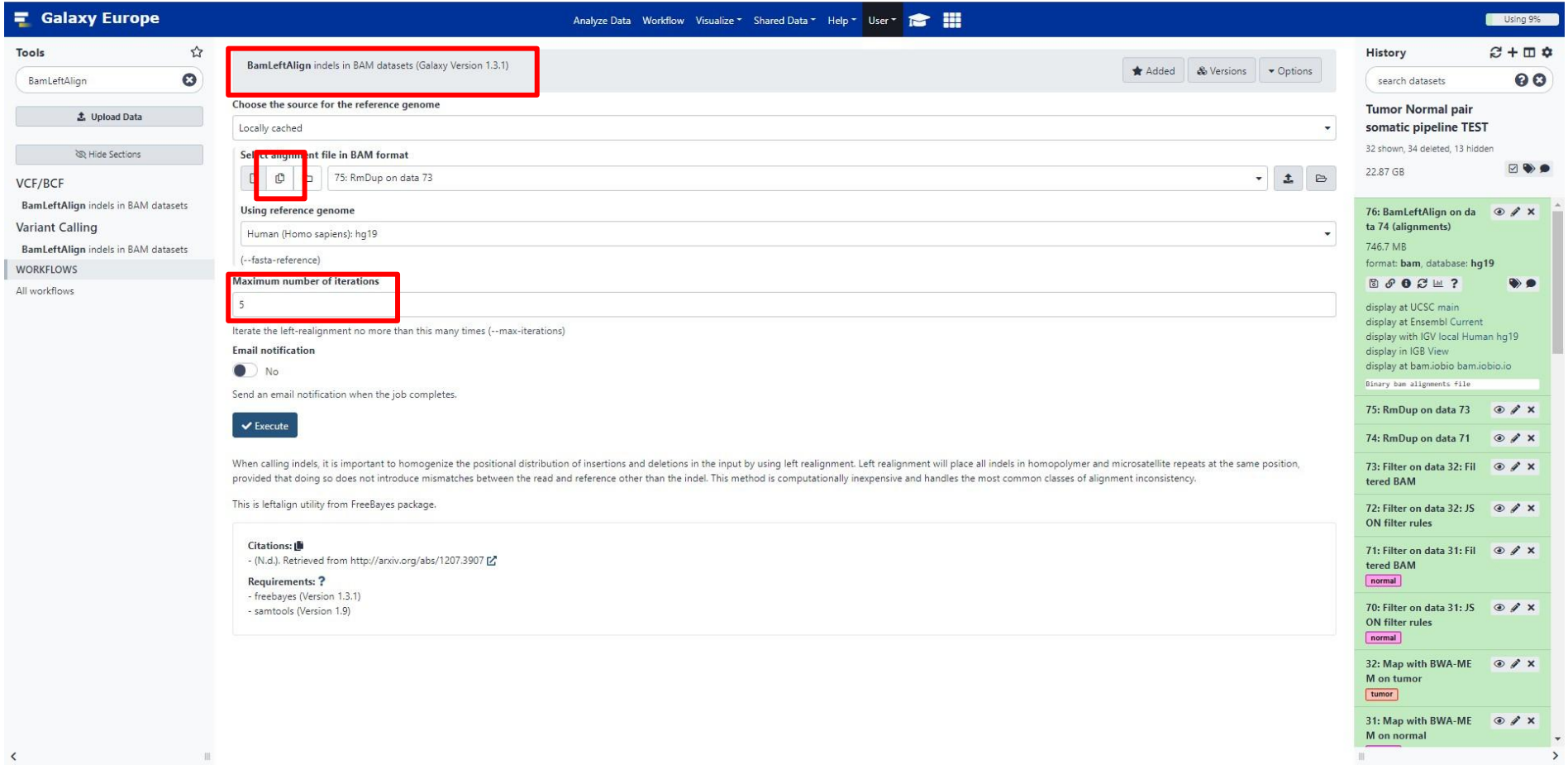
Remove duplicates with RmDup



Mapped reads postprocessing

c. Left-align reads around indels

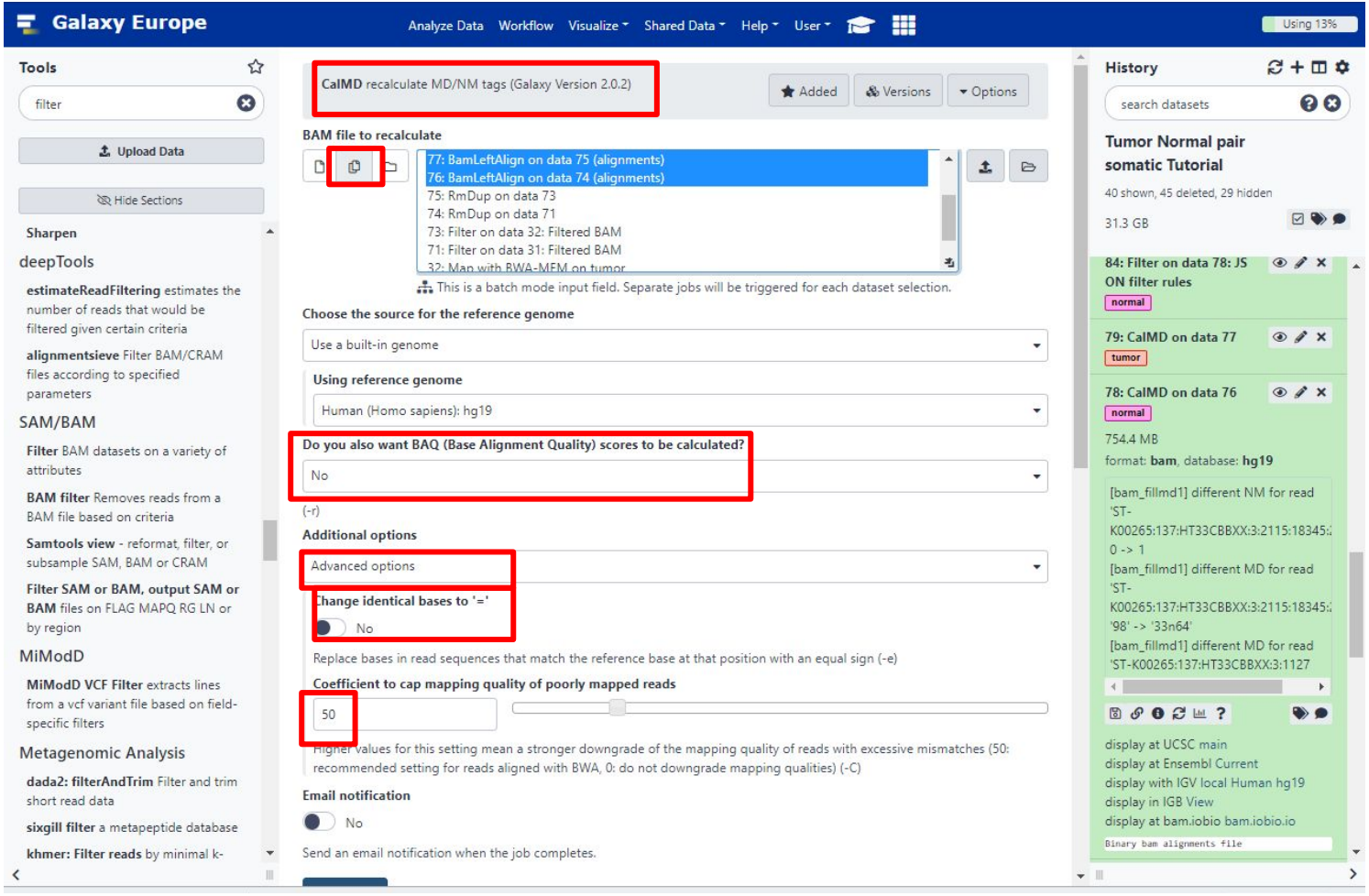
Left-align with BamLeftAlign



Mapped reads postprocessing

d. Recalibrate read mapping qualities

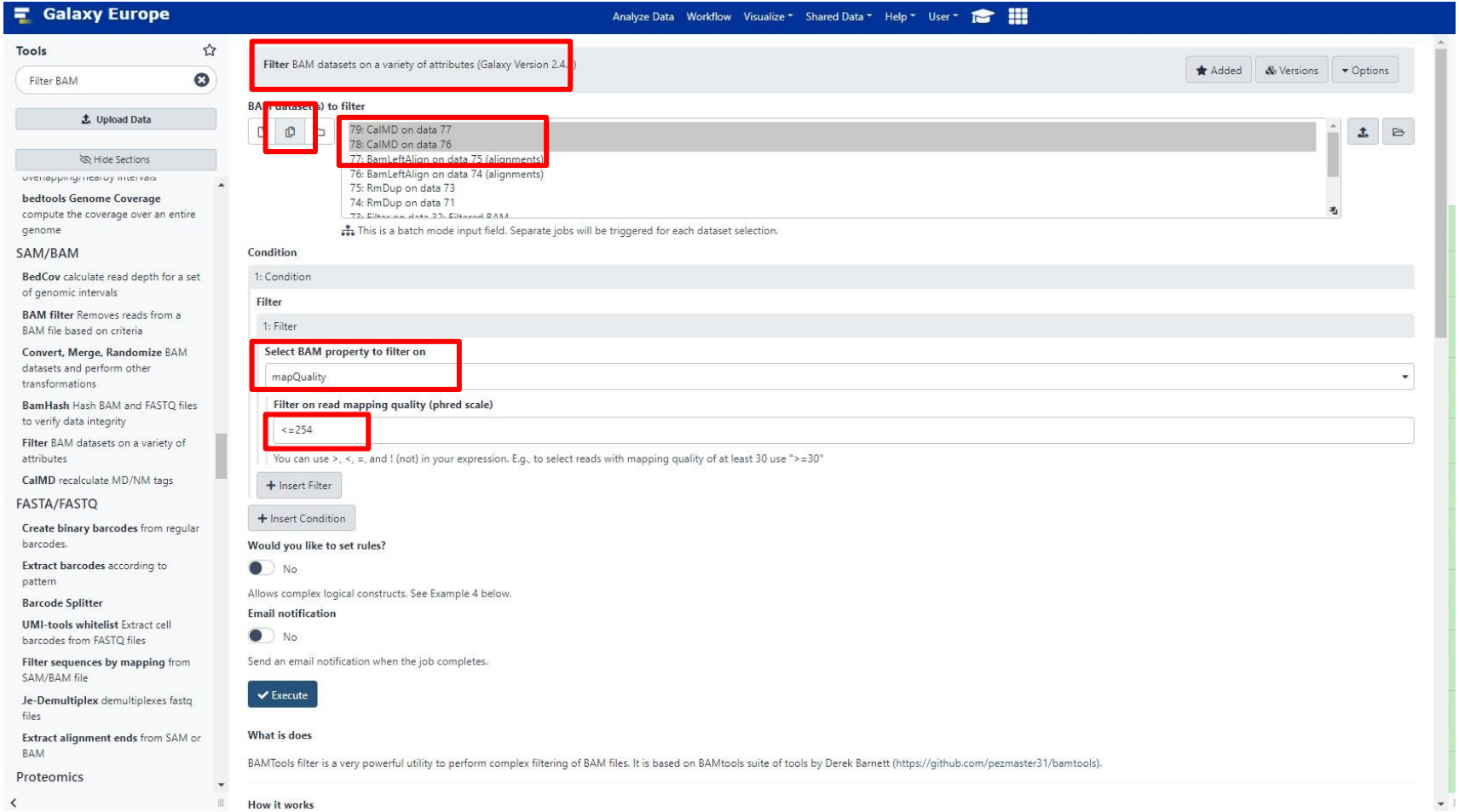
Recalibrate read quality scores with CalMD



Mapped reads postprocessing

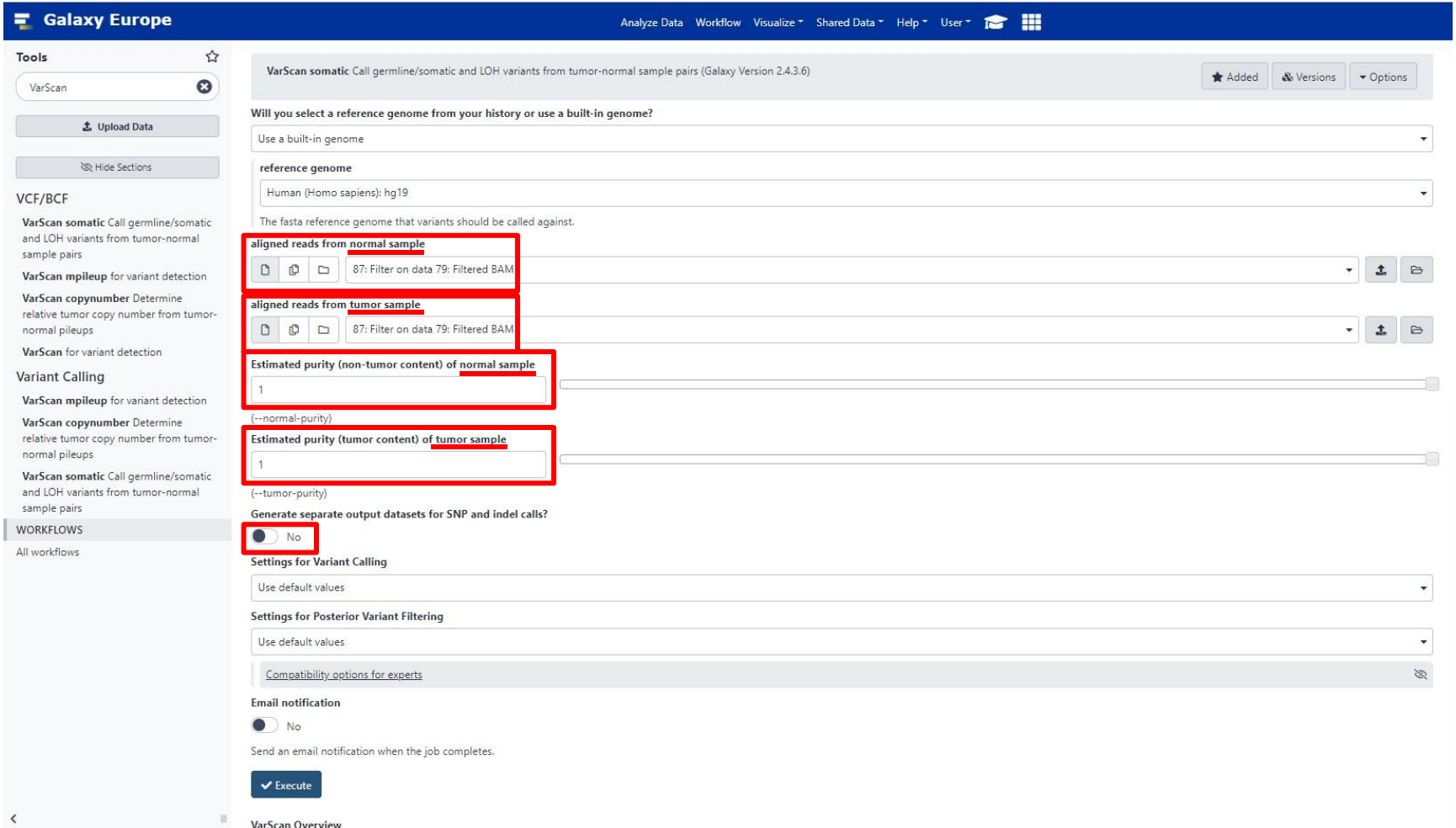
e. Refilter reads based on mapping quality

Eliminating reads with undefined mapping quality

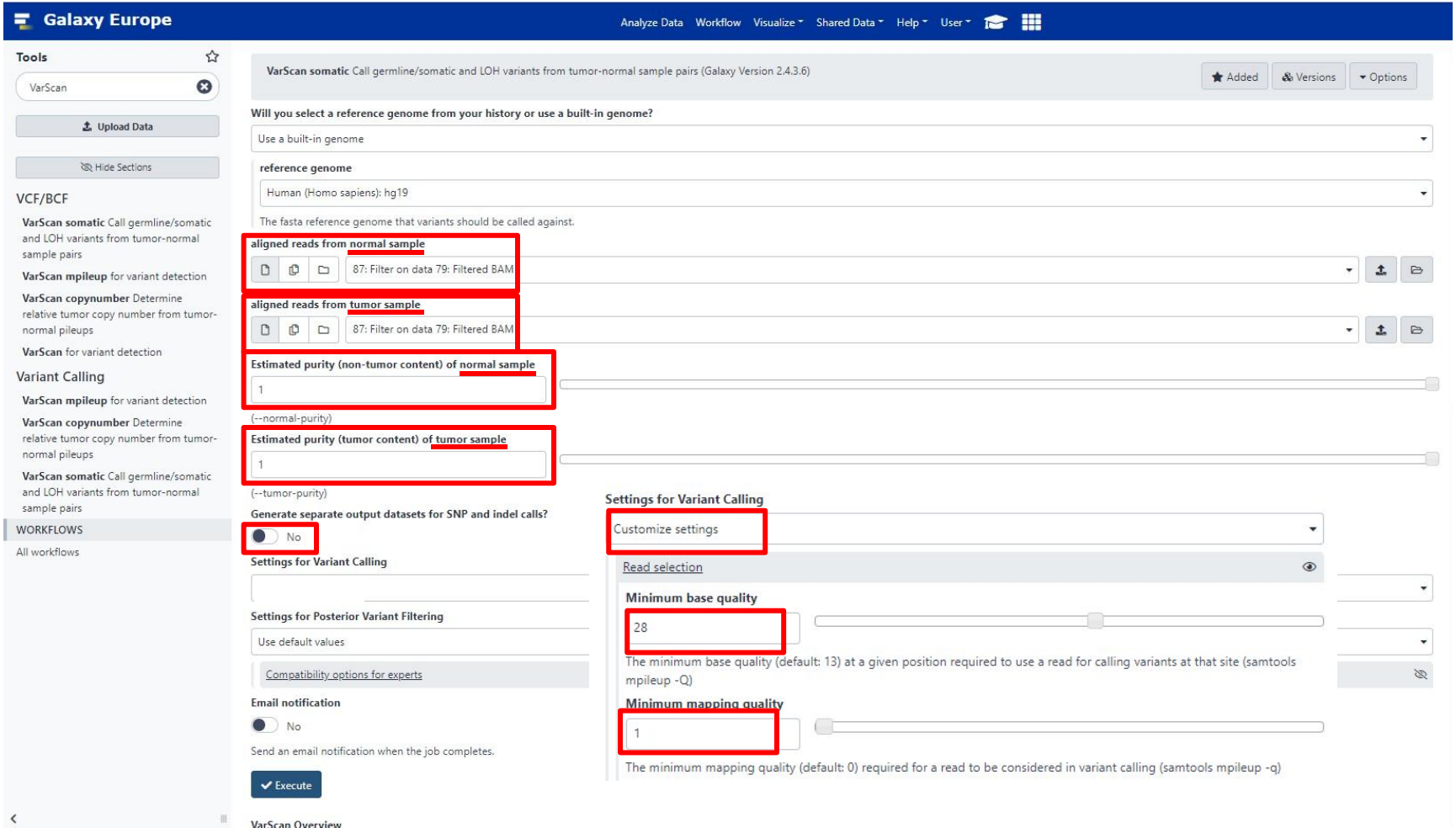


2. Variant calling and classification

Variant calling with VarScan somatic



Variant calling with VarScan somatic



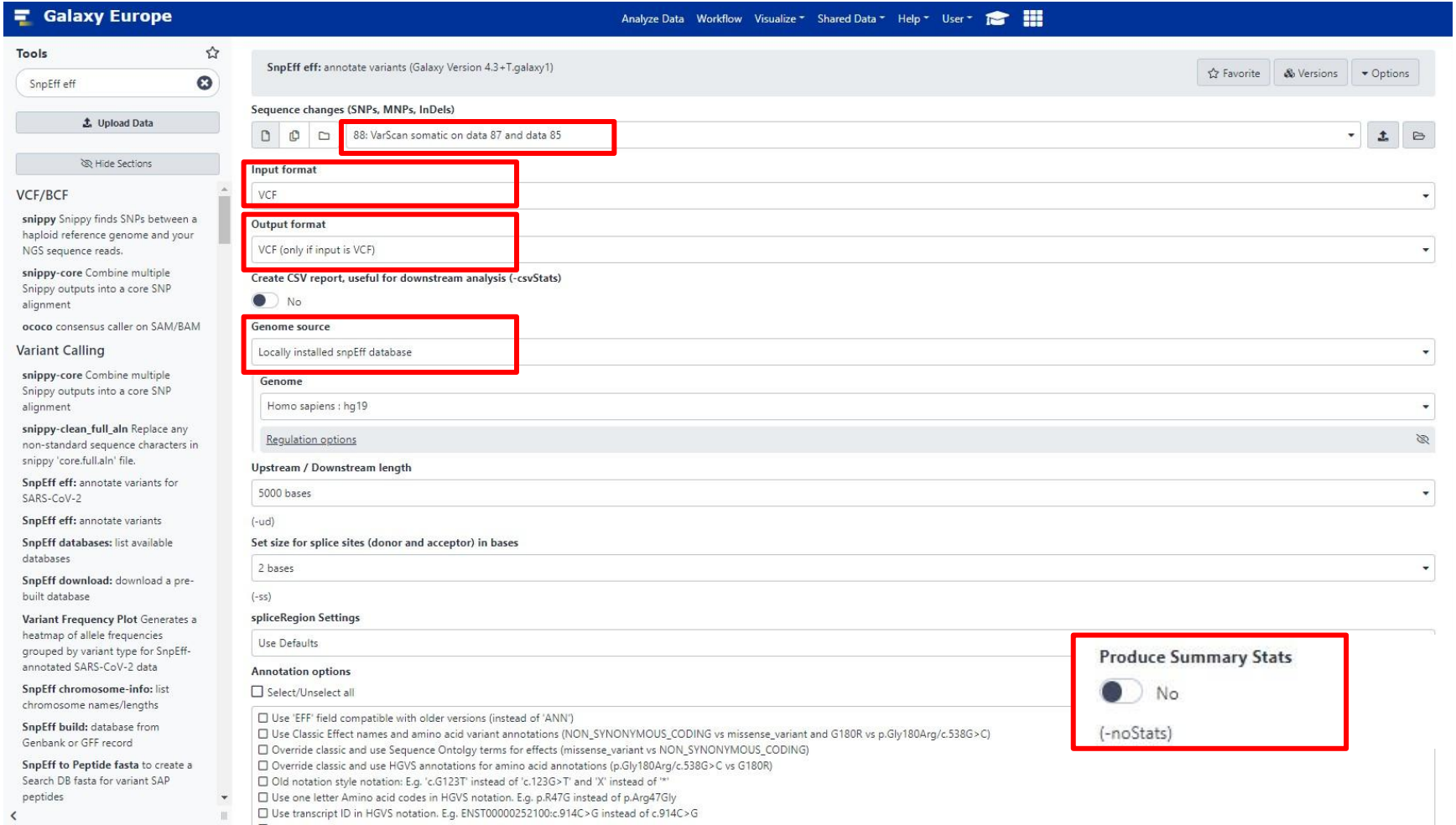
3. Variant annotation and reporting

Adding annotations to the called variants

a. Adding annotations to the called variants

a.1. Adding functional genomic annotations

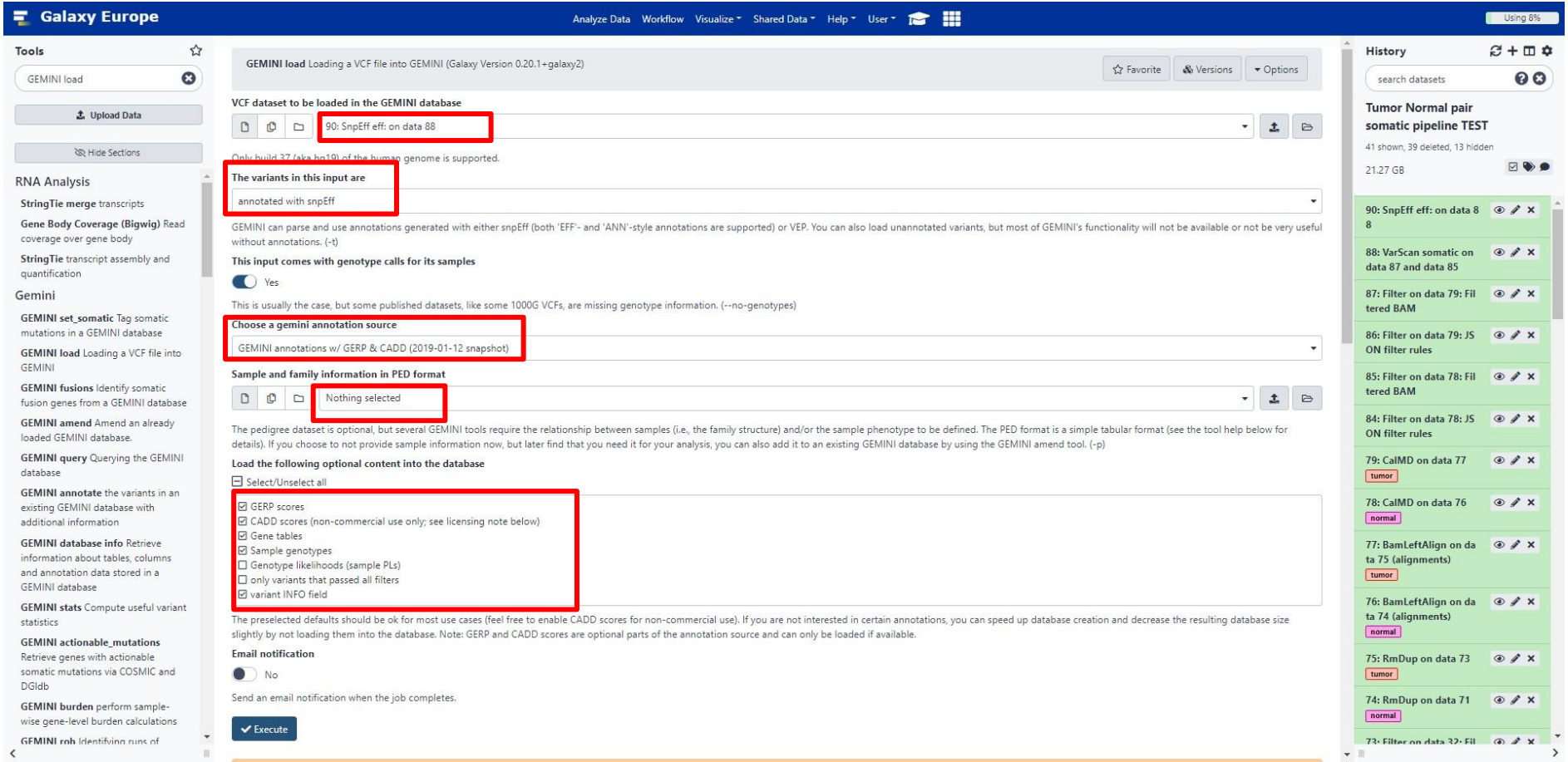
Adding annotations with SnpEff



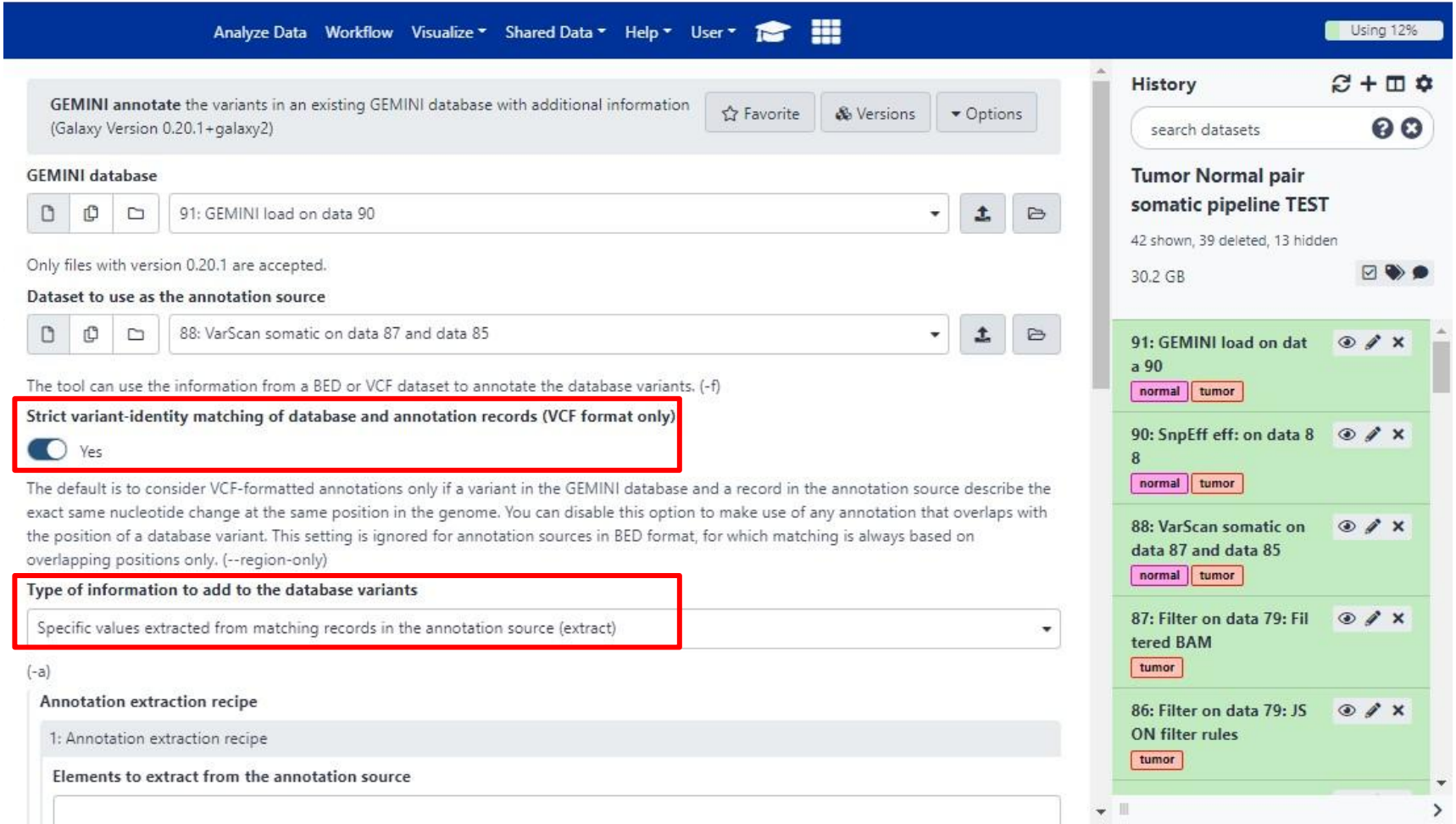
a. Adding annotations to the called variants

a.2. Adding genetic and clinical evidence-based annotations

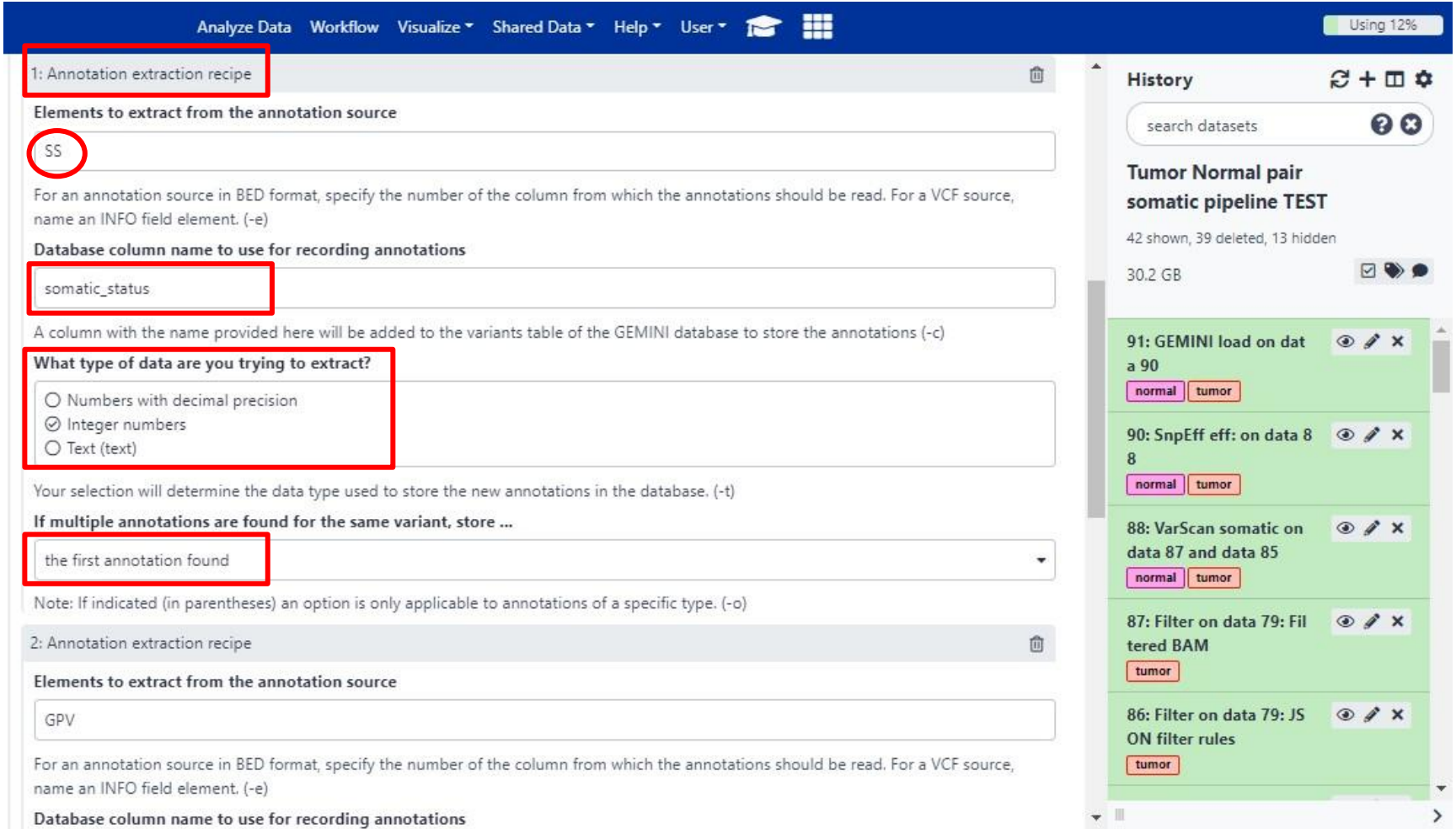
Creating a GEMINI database from a variants dataset



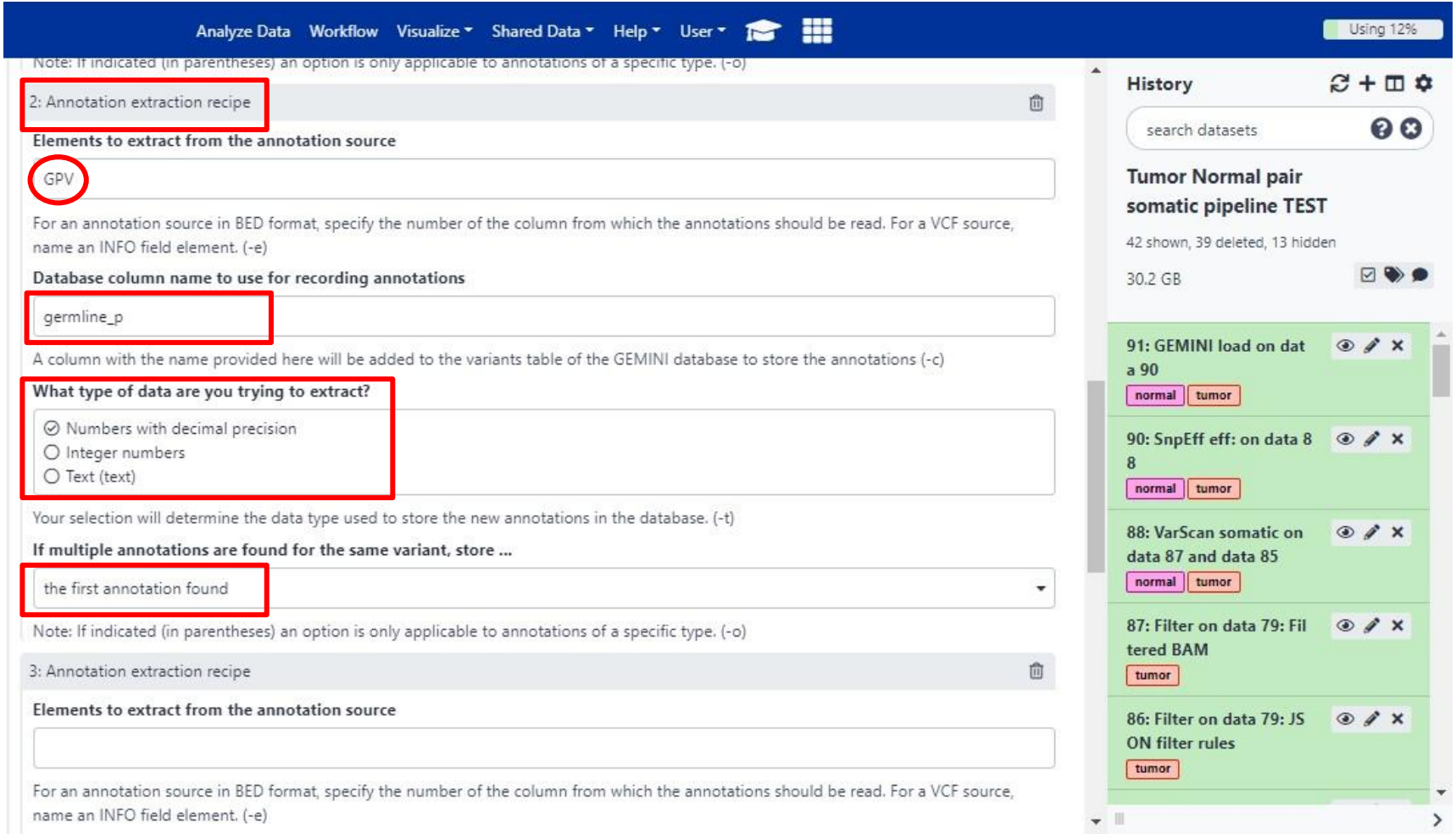
Making variant call statistics accessible



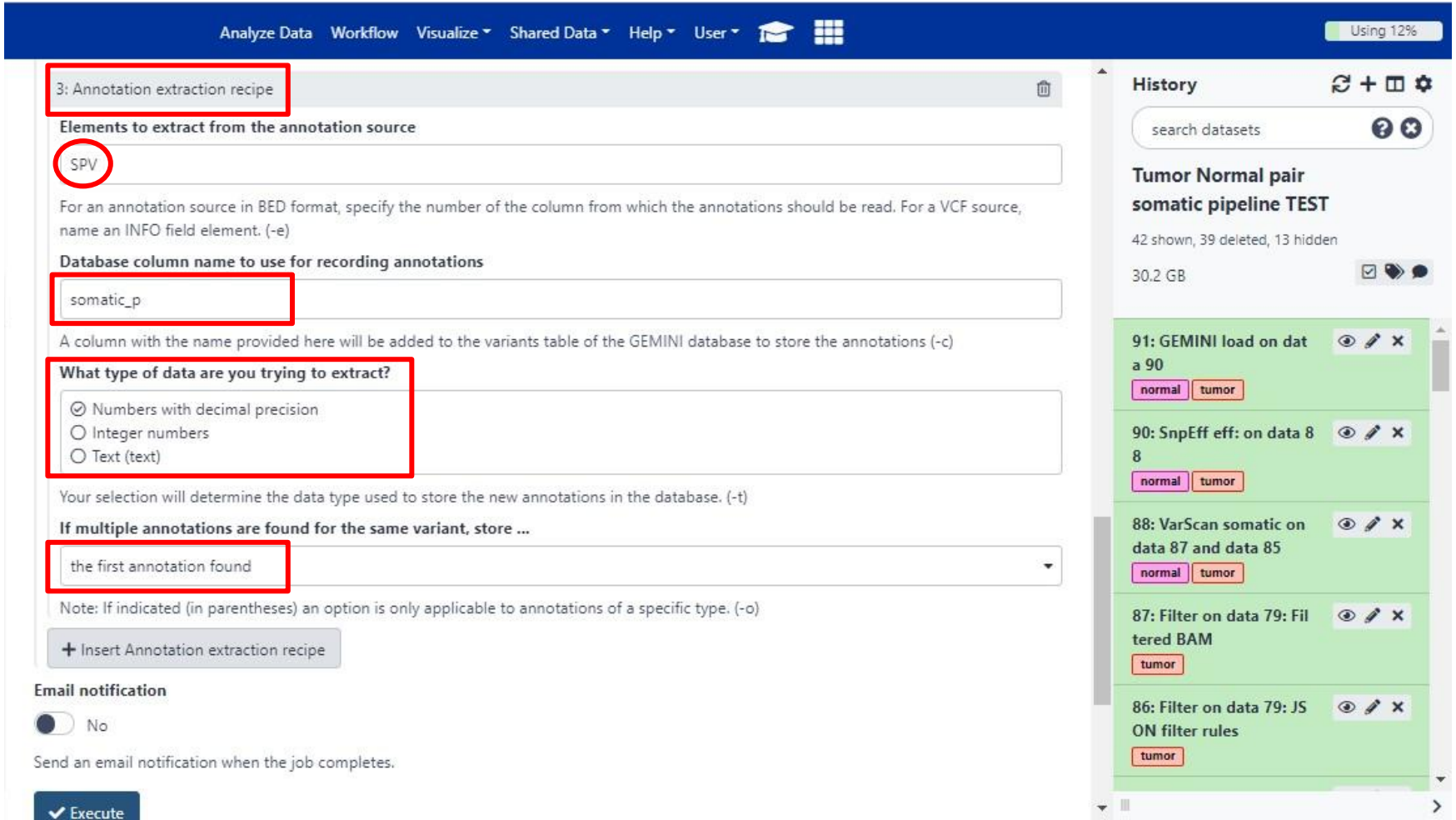
Making variant call statistics accessible



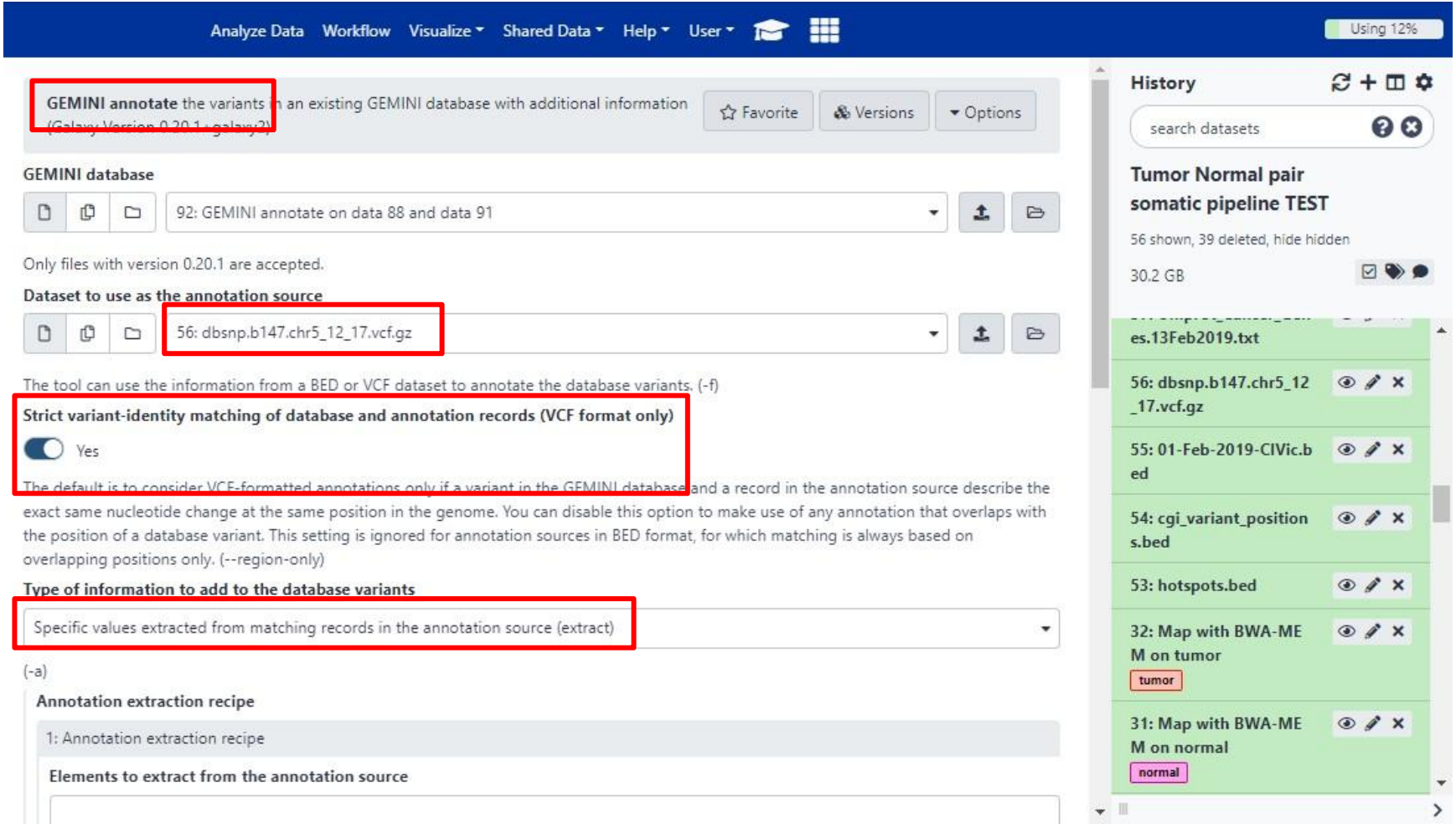
Making variant call statistics accessible



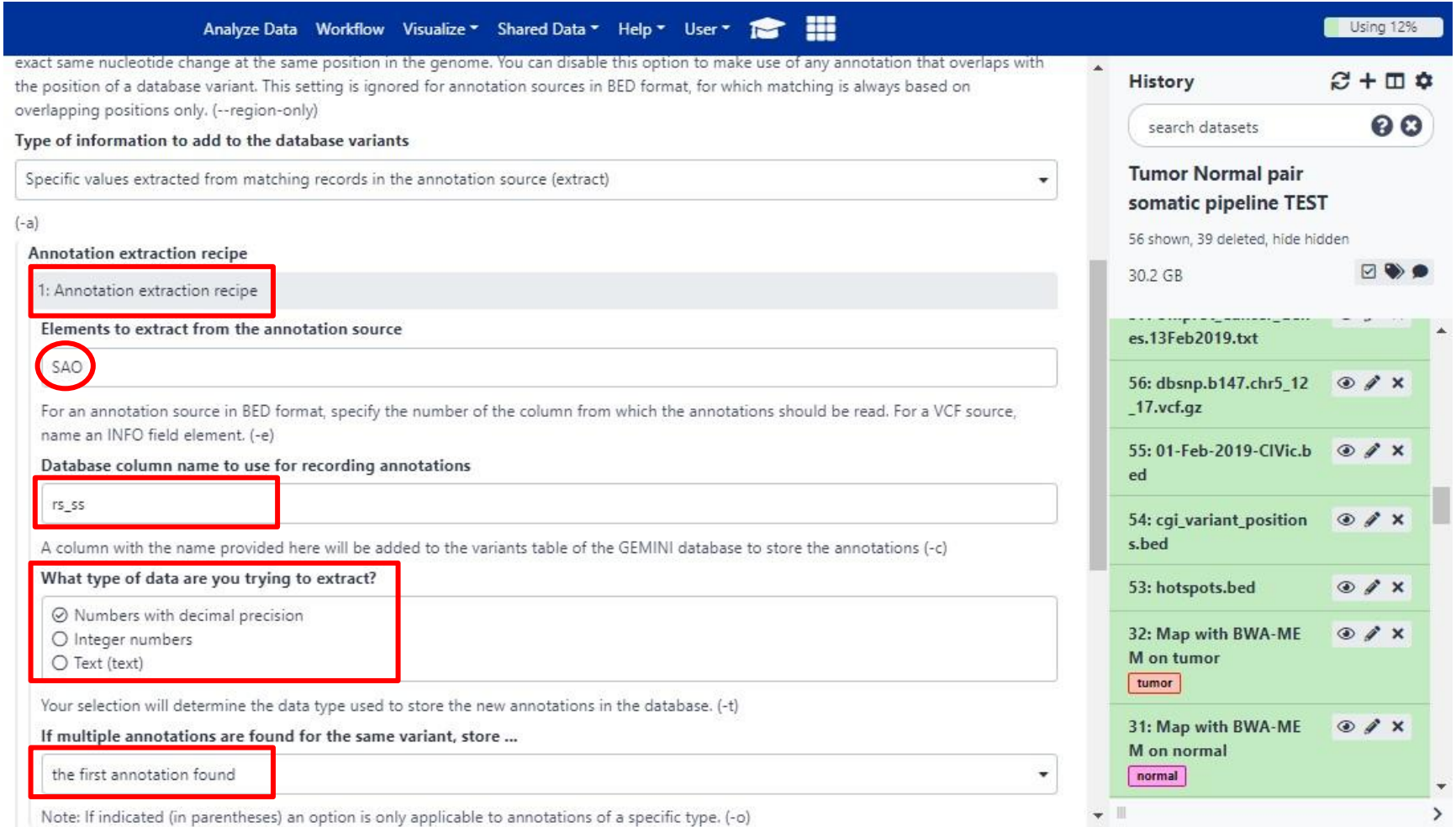
Making variant call statistics accessible



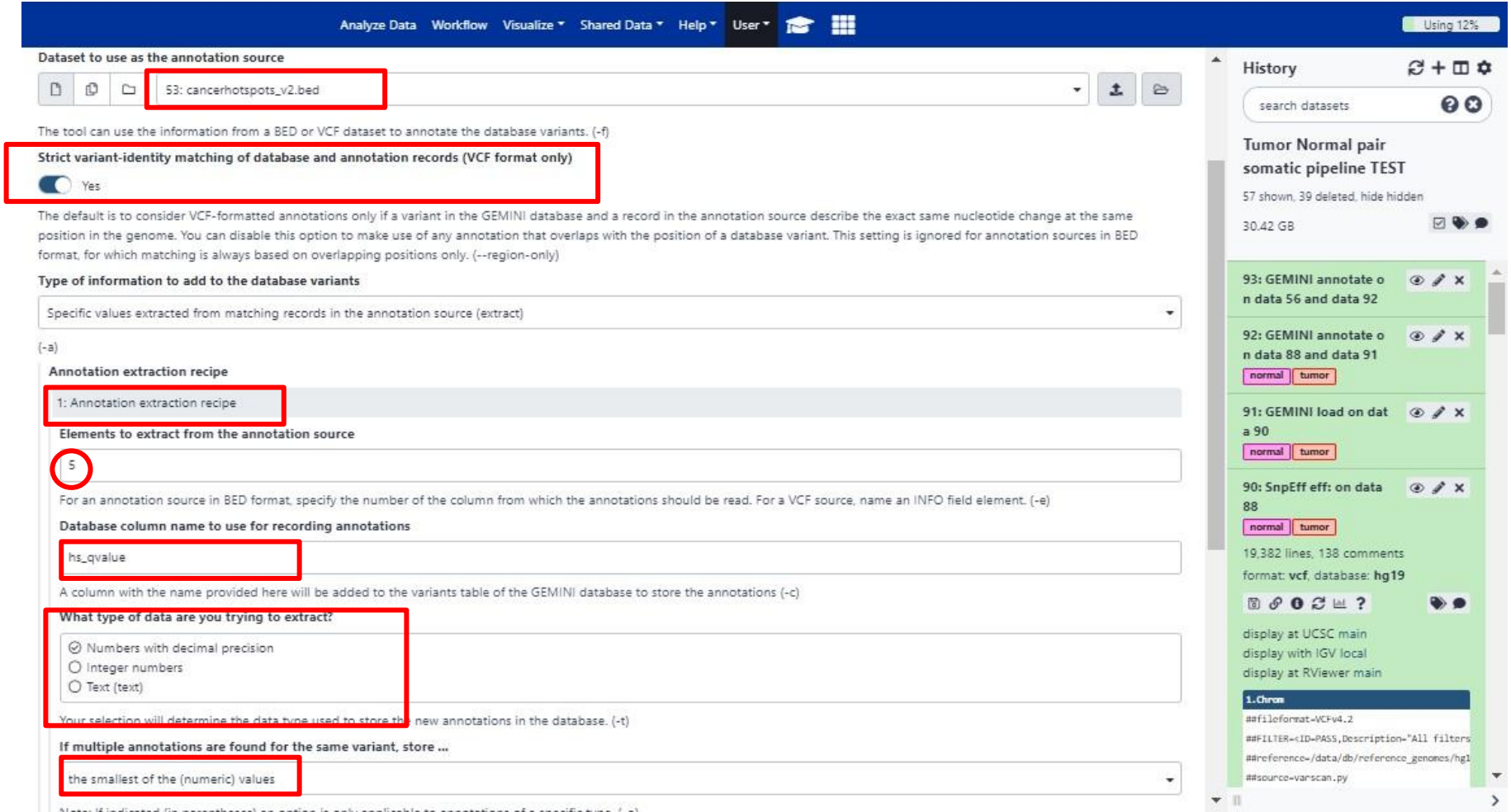
Adding further annotations from dbSNP



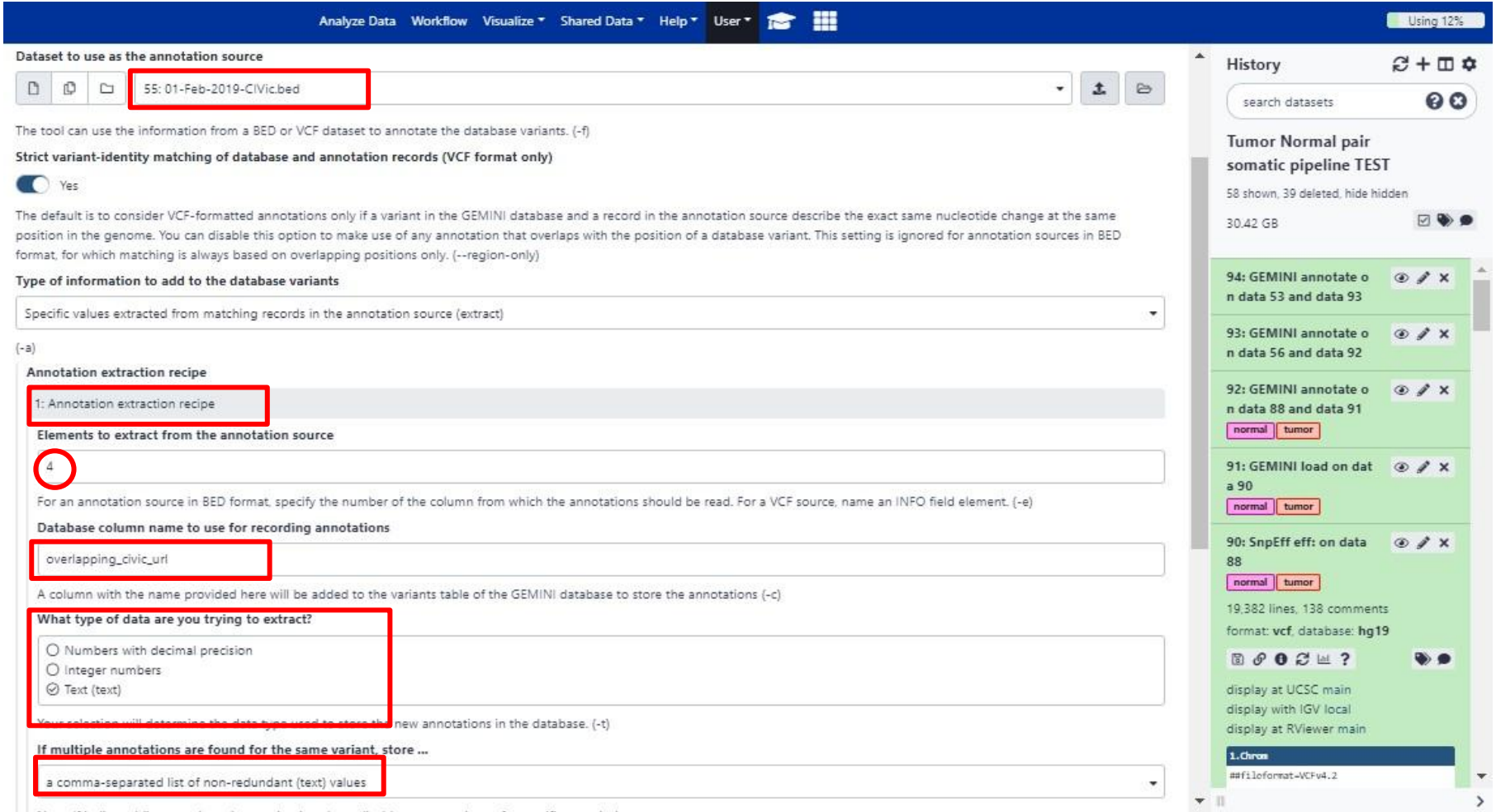
Adding further annotations from dbSNP



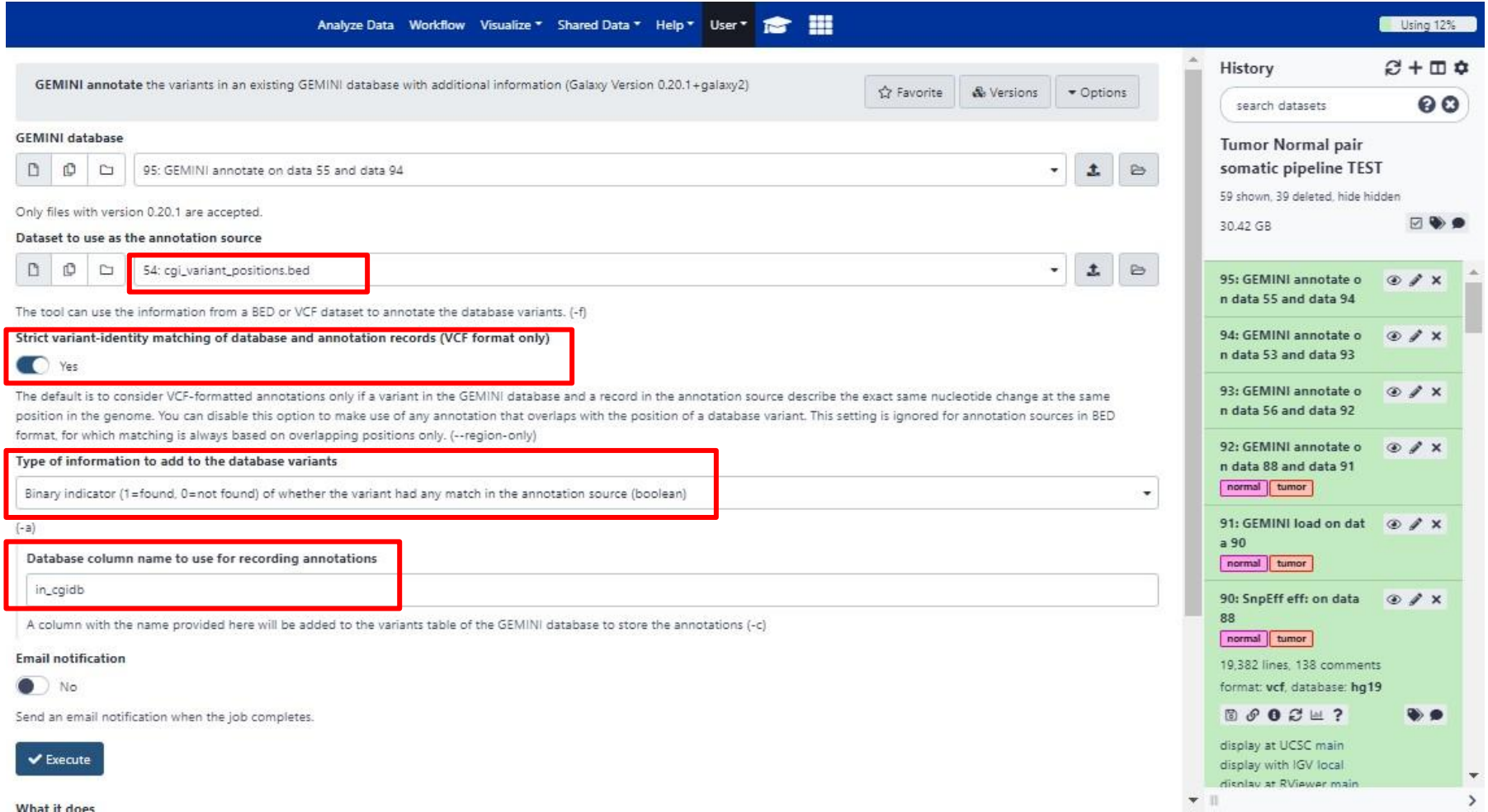
Adding further annotations from Cancer Hotspots v2



Adding links to CIViC

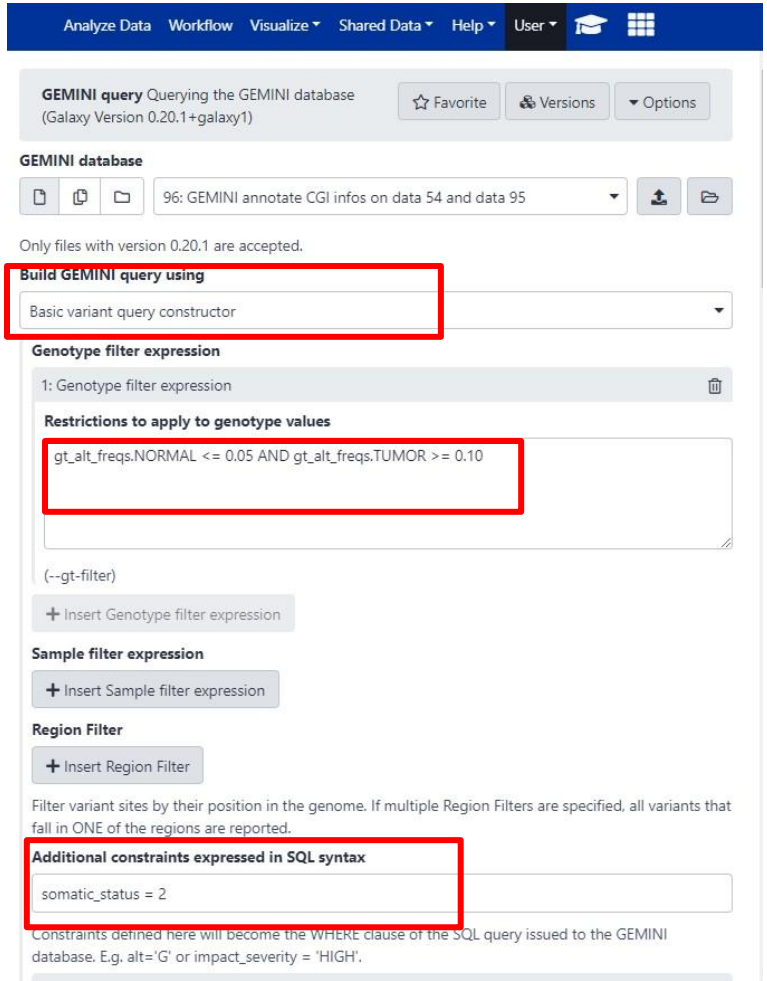


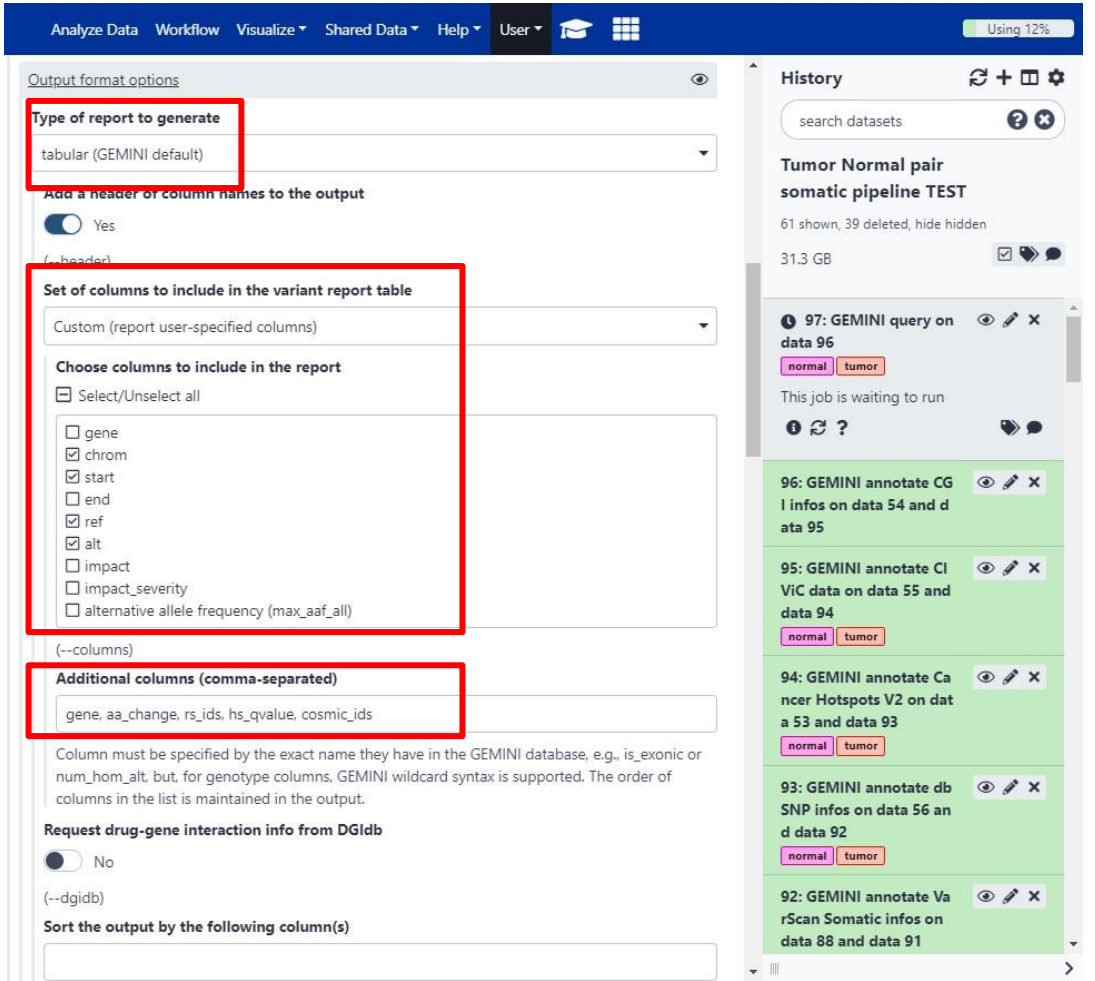
Adding further annotations from Cancer Genome Interpreter (CGI)



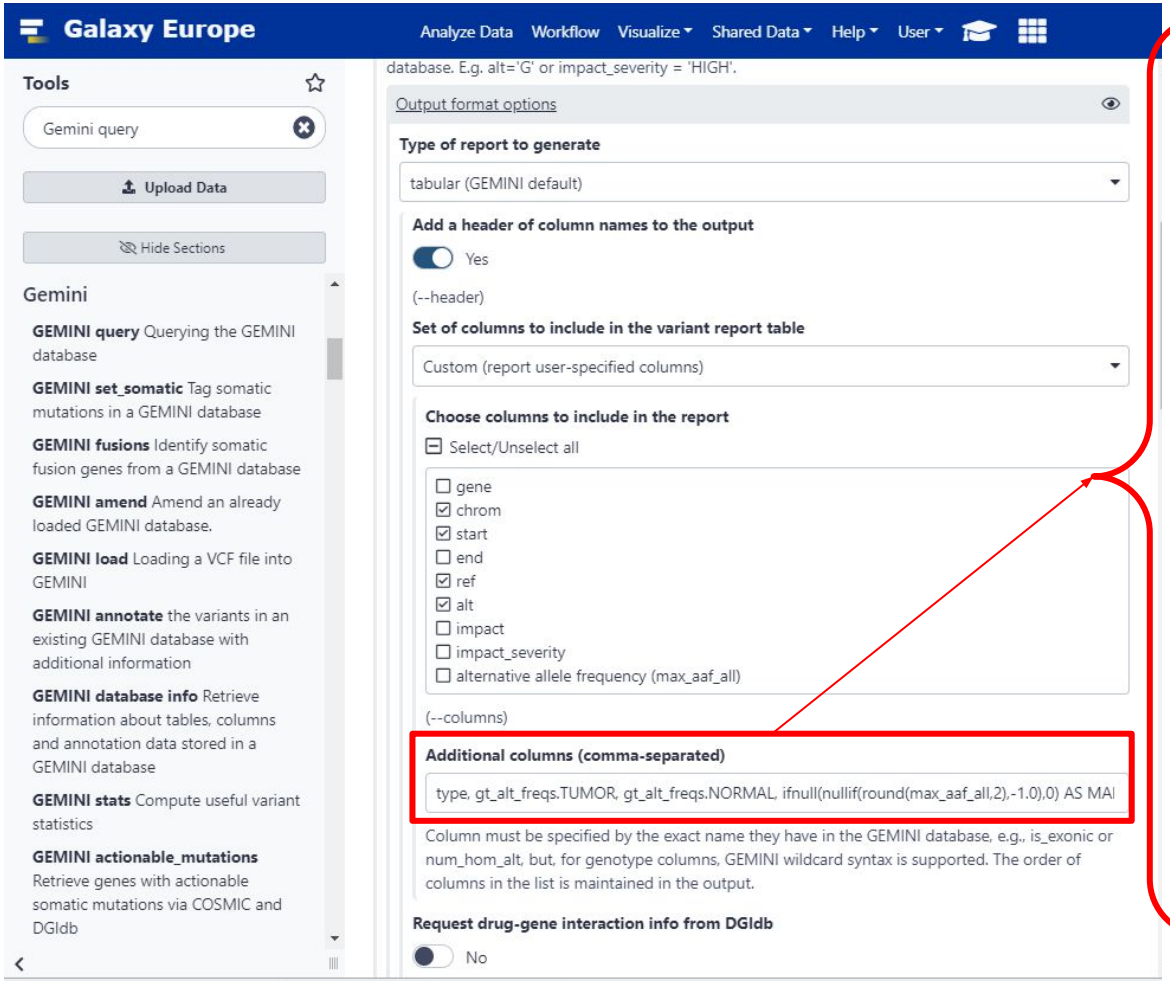
b. Reporting selected subsets of variants

Querying the GEMINI database for somatic variants





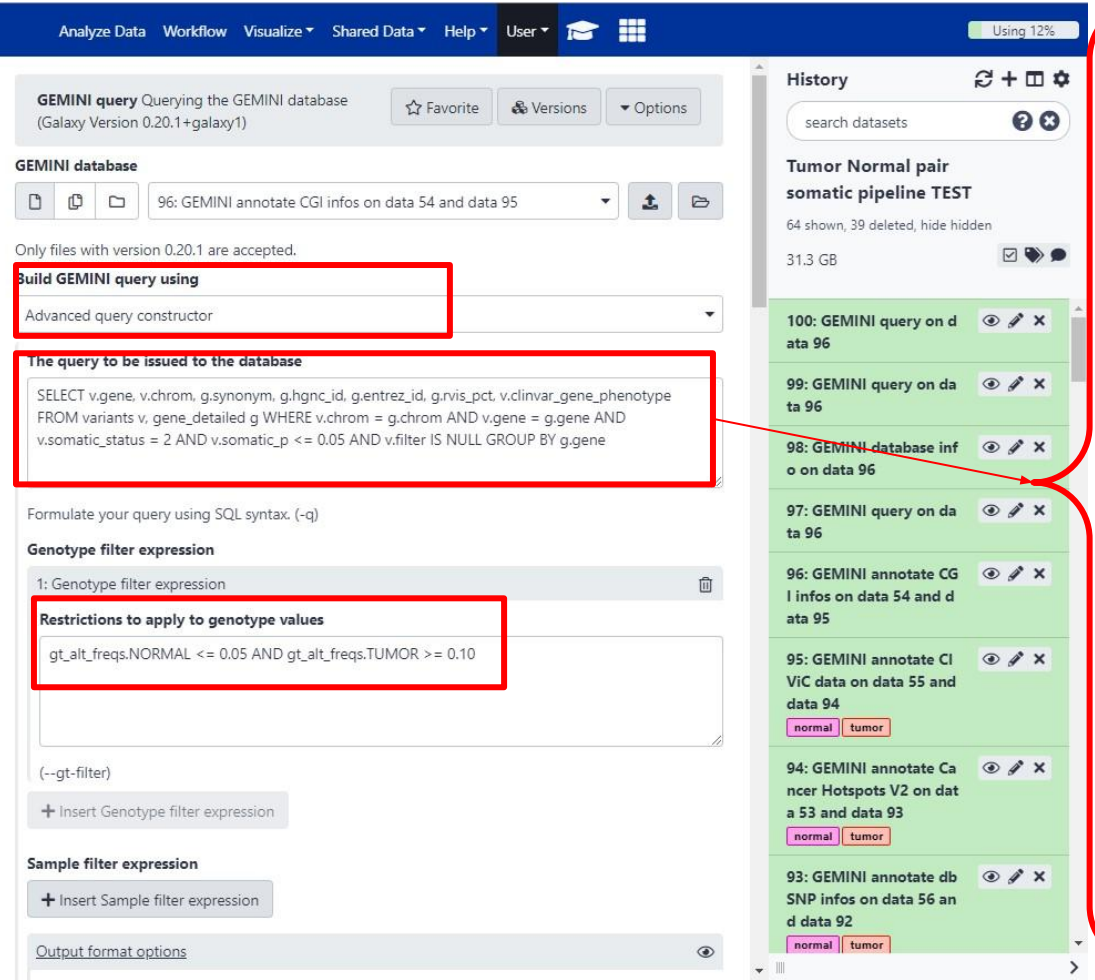
GEMINI SQL-based output formatting




```
type,
gt_alt_freqs.TUMOR,
gt_alt_freqs.NORMAL,
ifnull(nullif(round(max_aaf_all,2),-1.0),0)
AS MAF,
gene,
impact_so,
aa_change,
ifnull(round(cadd_scaled,2),'.') AS
cadd_scaled,
round(gerp_bp_score,2) AS gerp_bp,
ifnull(round(gerp_element_pval,2),'.') AS
gerp_element_pval,
ifnull(round(hs_qvalue,2), '.') AS
hs_qvalue,
in_omim,
ifnull(clinvar_sig,'.') AS clinvar_sig,
ifnull(clinvar_disease_name,'.') AS
clinvar_disease_name,
ifnull(rs_ids,'.') AS dbsnp_ids,
rs_ss,
ifnull(cosmic_ids,'.') AS cosmic_ids,
ifnull(overlapping_civic_url,'.') AS
overlapping_civic_url,
in_cgidb
```

c. Generating reports of genes affected by variants

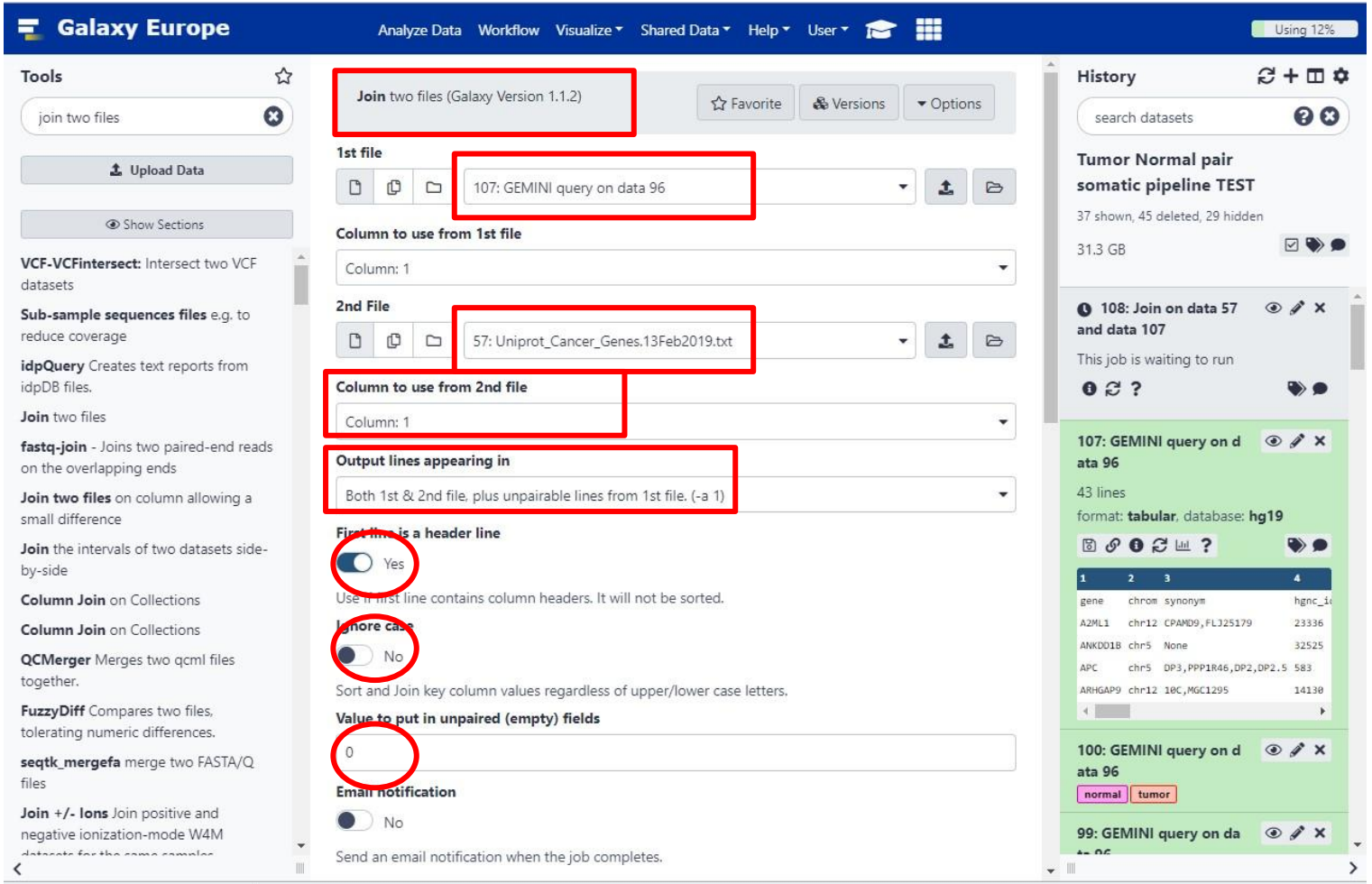
Turning query results into gene-centered reports



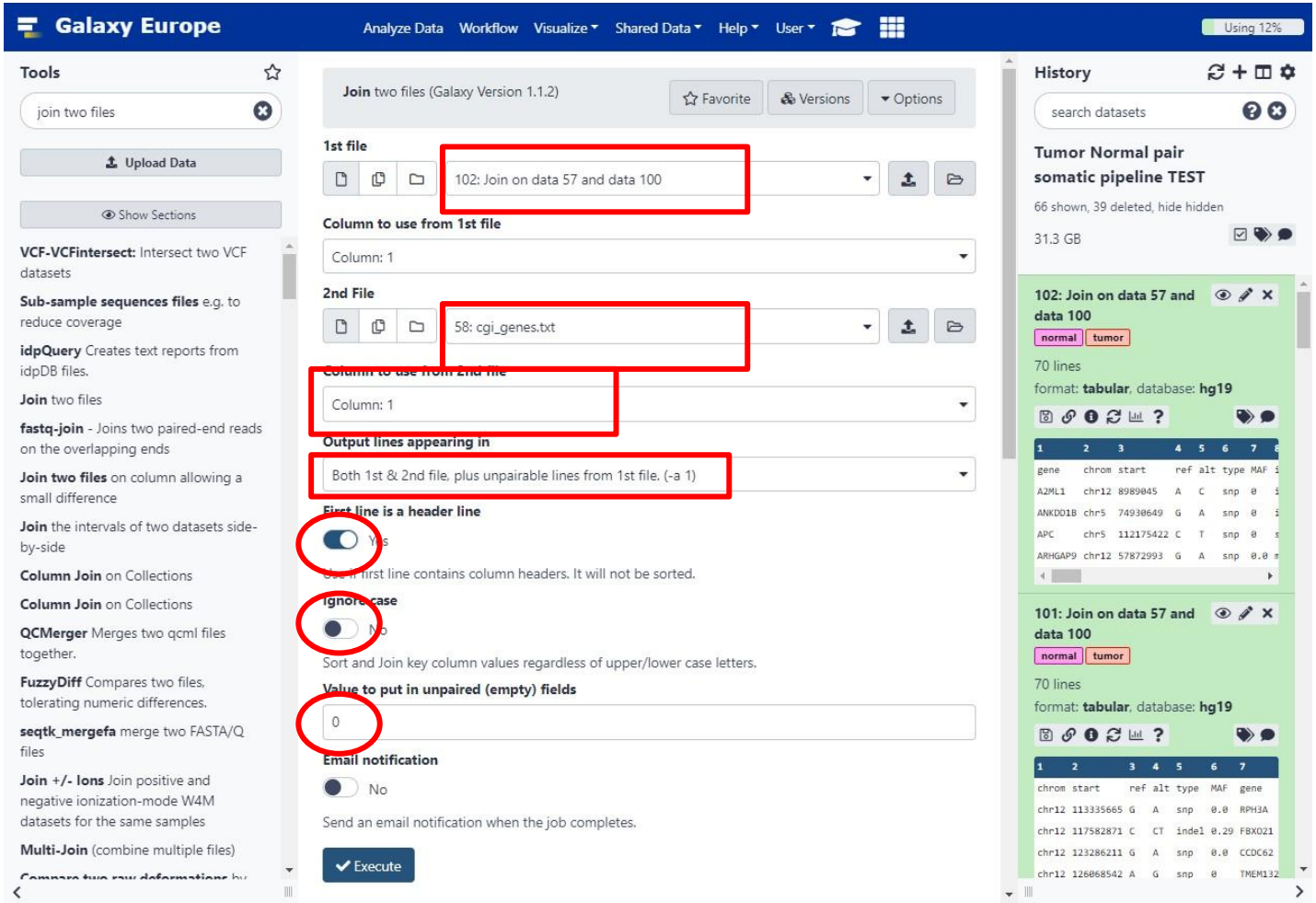

```
SELECT v.gene, v.chrom,
g.synonym, g.hgnc_id,
g.entrez_id, g.rvis_pct,
v.clinvar_gene_phenotype

FROM variants v,
gene_detailed g

WHERE v.chrom = g.chrom AND
v.gene = g.gene AND
v.somatic_status = 2 AND
v.somatic_p <= 0.05 AND
v.filter IS NULL


GROUP BY g.gene
```

d. Adding additional annotations to the gene-centered report

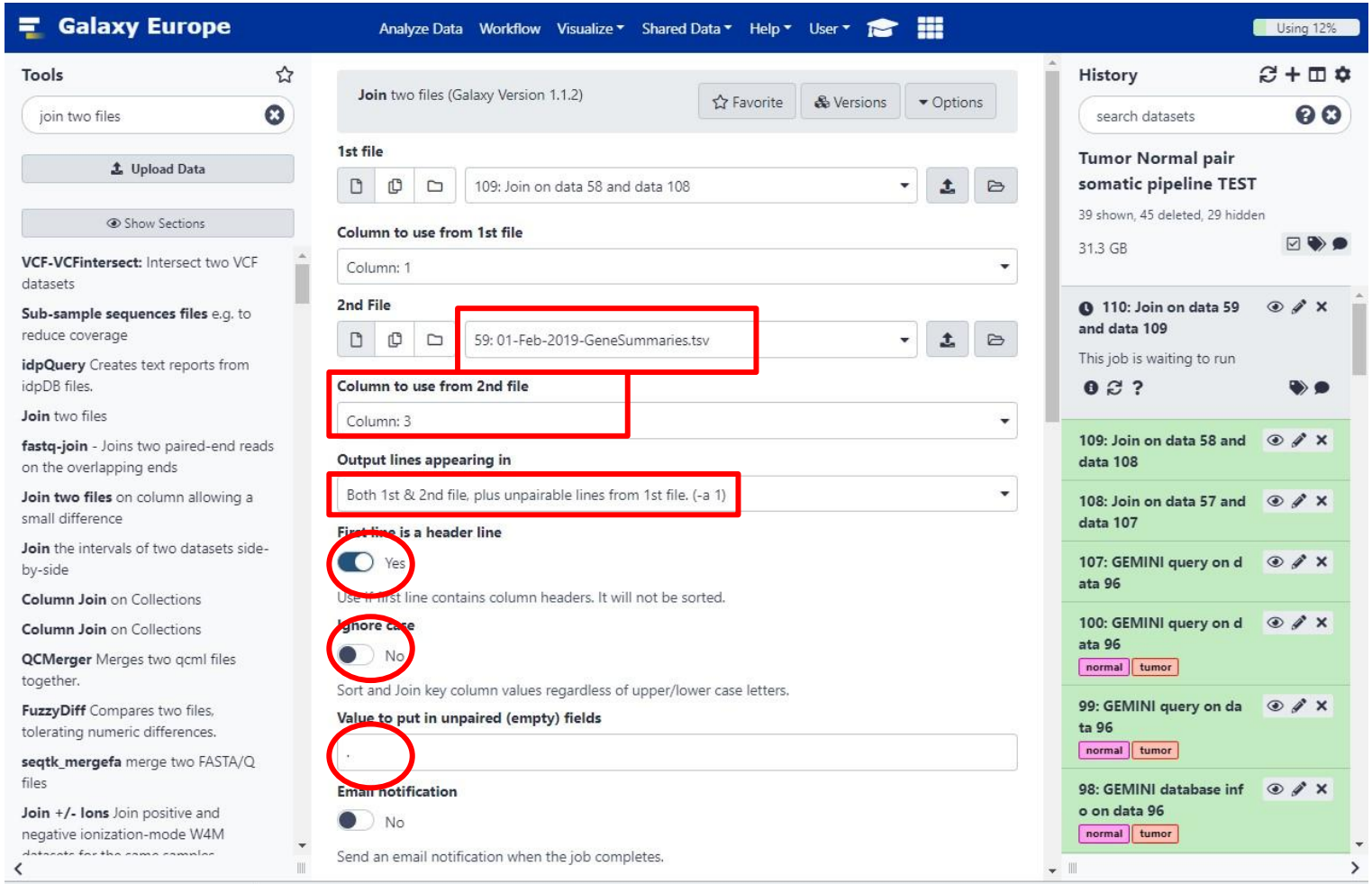
Adding UniProt cancer genes information



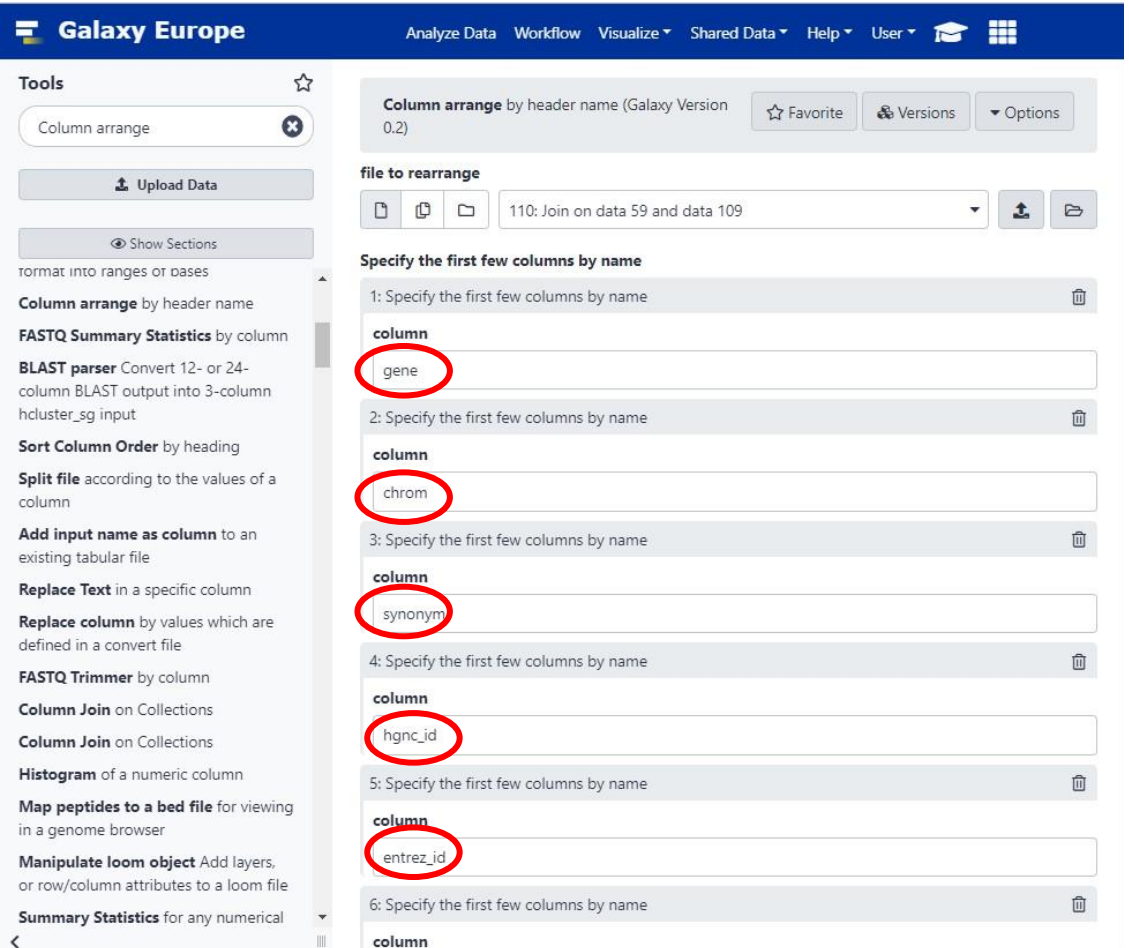
Adding CGI biomarkers information



Adding gene information from CIViC



Rearrange to get a fully annotated gene report



Inspecting fully annotated gene report

gene	chrom	synonym	hgnc_id	entrez_id	rvis_pct	is_OG	is_TS	in_cgi_biom	clinvar_gene	gene_civic_	description	gene_id	entrez_id	last_review_date
A2ML1	chr12	CPAMD9,FLJ2517	23336	144568	98.52559566	0	0	0	nonsyndrom					
ANKDD1B	chr5	None	32525	728780	None	0	0	0	None					
APC	chr5	DP3,PPP1R46,DP2	583	324	0.902335456	0	1	1	apc-associa	https://civic		66	324	2017-02-09 21:58:08 UTC
ARHGAP9	chr12	10C,MGC1295	14130	64333	30.90351498	0	0	0	coronary_ar					
C2CD5	chr12	CDP138,KIAA0528	29062	9847	None	0	0	0	None					
CCDC62	chr12	ERAP75,CT109,FL	30723	84660	82.29535268	0	0	0	None					
CDH12	chr5	Br-cadherin,CDH	1751	1010	29.48808681	0	0	0	None					
CDH18	chr5	EY-CADHERIN,CD	1757	1016	5.602736494	0	0	0	None					
CLEC4C	chr12	DLEC,CLECSF7,CD	13258	170482	86.47676339	0	0	0	None					
CLEC6A	chr12	dectin-2,CLECSF1	14556	93978	89.95635763	0	0	0	None					
COX7C	chr5	None	2292	1350	62.38499646	0	0	0	None					
DDX51	chr12	None	20082	317781	64.96225525	0	0	0	None					
ELAC2	chr17	FLJ10530,HPC2	14198	60528	10.12031139	1	0	0	combined_c					
ERN1	chr17	IRE1P,IRE1	3449	2081	24.63434772	0	0	0	None					
ESM1	chr5	None	3466	11082	56.64071715	0	0	0	None					
FBXO21	chr12	FBX21,KIAA0875	13592	23014	12.77423921	0	0	0	None					
HAPLN1	chr5	CRTL1	2380	1404	20.53550366	0	0	0	None					
ITGB7	chr12	None	6162	3695	4.62962963	0	0	0	None					
KRAS	chr12	KRAS1,KRAS2	6407	3845	42.87567823	1	0	1	acute_myel	https://civic	Mutations i	30	3845	2017-02-09 21:59:28 UTC
KRBA2	chr17	None	26989	124751	57.31304553	0	0	0	None					
LINC01019	chr5	None	27742	285577	None	0	0	0	None					
LYRM7	chr5	FLJ20796,C5orf31,	28072	90624	56.2514744	0	0	0	mitochondri					
METTL2A	chr17	METTL2,FLJ12760	25755	339175	67.03231894	0	0	0	None					
MROH2B	chr5	FLJ40243,DKFZp7	26857	133558	None	0	0	0	None					
PCDHB9	chr5	PCDH-BETA9,PCDI	8694	None	None	0	0	0	None					
PCDHGB1	chr5	PCDH-GAMMA-B1	8708	56104	10.92238736	0	0	0	None					
PCDHGB7	chr5	PCDH-GAMMA-B7	8714	56099	48.90894079	0	0	0	None					
PDE3A	chr12	CGI-PDE	8778	5139	6.894314697	0	0	0	brachydacty					
RACK1	chr5	GNB2L1,H12,3,Gn	None	10399	46.20193442	0	0	0	None					
RNF213	chr17	NET57,KIAA1618,	14539	57674	97.65274829	1	0	0	moyamoya_					
SLC16A5	chr17	MCT5,MCT6	10926	9121	36.22906346	0	0	0	None					
SMARCC2	chr12	Rsc8,CRACC2,BAF	11105	6601	2.707006369	0	0	0	malignant_t					
SOX5	chr12	L-SOX5,MGC3515	11201	6660	9.088228356	0	0	0	aplasia/hyp					
SPEF2	chr5	KPL2,FLJ23577,CT	26293	79925	99.06817646	0	0	0	None					
SYNPO	chr5	KIAA1029	30672	11346	40.67586695	0	0	0	None					
TENM2	chr5	Ten-M2,KIAA1127	29943	57451	None	0	0	0	None					
TMEM132B	chr12	KIAA1906,KIAA17	29397	114795	2.677518283	0	0	0	None					
TP53	chr17	LFS1,p53	11998	7157	35.98726115	0	1	1	acute_mega	https://civic	TP53 mutati	45	7157	2018-03-30 15:05:39 UTC
TTC23L	chr5	FLJ25439	26355	153657	96.55579146	0	0	0	None					
TTC37	chr5	KIAA0372	23639	9652	58.00896438	0	0	0	malignant_t					
USP22	chr17	KIAA1063,USP3L	12621	23326	27.41802312	0	0	0	None					
VCAN	chr5	CSPG2,PG-M	2464	1462	19.95753715	0	0	0	malignant_t					