

# **GALAXY PIPELINE FOR FASTER WHOLE GENOME GENOTYPE CALLING ON THE GENETITAN PLATFORM**

**Oleksiy Karpenko, Neil J. Bahroos**

**Galaxy Community Conference  
Chicago  
July 26, 2012**

# OVERVIEW

- What is Affymetrix GeneTitan?
- Axiom® Genome-Wide Population-Optimized Human Arrays
- Affymetrix Power Tools (APT) vs Affymetrix Genotyping Console
- APT in Galaxy at UIC
- Galaxy workflow for Axiom® arrays
- Future work

# AFFYMETRIX GENETITAN



- Automated high-throughput solution for monitoring gene expression and genome-wide SNP genotyping
- 16-, 24- and 96-format array plates
- On campus at RRC Core Genomics Facility, College of Medicine West (CMW), Room A-301

# AXIOM® GENOME-WIDE POPULATION-OPTIMIZED HUMAN ARRAYS

- One Axiom Genome-Wide PanAFR Plate for 96 samples
- ~700K SNPs per plate
- One Axiom GW (~2.1M) PanAFR Array Set is a set of 3 plates
- Also available CEU (European), ASI (Asian), CHB (Chinese Beijing), and custom myDesign plates

# POWER TOOLS VS GENOTYPING CONSOLE

```
okarpe2@nike:~/apt-1144-amd64-pc-linux/bin$ ls
apt-annotation-converter      apt-rt-cel-transformer      CreateDOWDocument          probe-set-summarize-quick-test
apt-calvin-equivalent         apt-rt-chp-to-txt          createTav                  probe-set-summarize-quick-test.valgrind
apt-calvinlite-test          apt-rt-copynumber-cyto     dmet-genotype-test        probe-set-summarize-test
apt-calvinlite-util          apt-rt-dmet-genotype       dmet-translation-test     PSVIMWriter
apt-canary                    apt-rt-geno-gc             dm-platform-issue        Redirect
apt-cdf-export                apt-rt-matrix-diff         DOMCount                  SAM2Count
apt-cdf-to-spf                apt-rt-probeset-genotype   DOMPrint                  SAMPrint
apt-cel-convert              apt-rt-probeset-summarize  dump-guid                 SAMCount
apt-cel-extract              apt-rt-summary-genotype   dump-pgf-test            SAMPrint
apt-cel-transformer          apt-rt-tsv-join           engine-runner             SCMPrint
apt-check-calvinchp          apt-run-pnode              EnumVal                   SEnumVal
apt-check-cel                apt-snp-compare           err-example               simple-file5-example
apt-check-chp                apt-snp-model-converter   example-file5-1          simpleRmaExample
apt-check-matrix             apt-snp-model-converter-quick-test  file5-equivalent-test    snp-clustere-score-test
apt-check-mixedfile          apt-snp-summary           file-example              snp-summary-quick-test
apt-check-textfile           apt-socket-run            geno-gc-quick-test       snp-summary-test
apt-chp-to-table             apt-summary-genotype      geno-gc-test              socket-listener-demo
apt-chp-to-txt               apt-summary-vis           gif2h5                    StdInParse
apt-convert-spf              apt-table-to-impute       h52gif                    summary-genotype-quick-test
apt-copynumber-cyto          apt-tsv-join              h5cc                      summary-genotype-quick-test.valgrind
apt-copynumber-wave          birdseed                   h5debug                   summary-genotype-test
apt-copynumber-workflow      birdseed-priors-util      h5diff                    summary-vis-test
apt-data-step                birdseed-priors-util-v1   h5dump                    test-affy-random-sample
apt-data-subset              birdseed-v1                h5import                   test-tboard
apt-dmet-copynumber          calvin-compare-cel        h5jam                      test-dabg
apt-dmet-genotype            calvin-converter-cel      h5ls                       test-files
apt-dmet-translation         calvin-converter-chp      h5redploy                 test-regression-copynumber-cyto
apt-dump-pgf                 calvin-data-file-extractor  h5repack                  test-regression-copynumber-cyto-quick
apt-engine-wrapper           calvin-equivalent-test    h5repart                   test-regression-copynumber-cyto-quick.valgrind
apt-file5-equivalent         canary-quick-test          h5sunjam                   test-regression-copynumber-wave
apt-file5-to-sqlite          canary-test                mas5-test                  test-regression-copynumber-workflow
apt-file5-util               cc-chp-compare            matrix-diff-test           test-regression-copynumber-workflow-quick
apt-geno-gc                  cdf-export-quick-test     mem-info                   test-regression-copynumber
apt-mas5                     cel-compare                MemParse                   test-tat-int16
apt-matrix-diff              cel-convert-quick-test    multiChannelExample       test-templatefactory
apt-midas                    cel-convert-test          pcre-config                 test-toy-norm
apt-probeset-genotype         cel-extract-quick-test    pcregrep                    test-ToySelect
apt-probeset-summarize       cel-to-data-store         pcretest                    test-tsv
apt-probeset-summarize-stages  cel-transformer-quick-test  plier-example              test-tsv-example
apt-rr-canary                cel-transformer-test       PParse                      test-tsv-join-test
apt-rr-cdf-export            chp-to-txt-quick-test     probe-set-genotype-quick-test  tsv-util
apt-rt-cel-convert           chp-to-txt-test           probe-set-genotype-quick-test.valgrind  util-meminfo
apt-rt-cel-extract           convert-priors              probe-set-genotype-stages-quick-test  xda-chp-compare
                                                         probe-set-genotype-test     XInclude
```

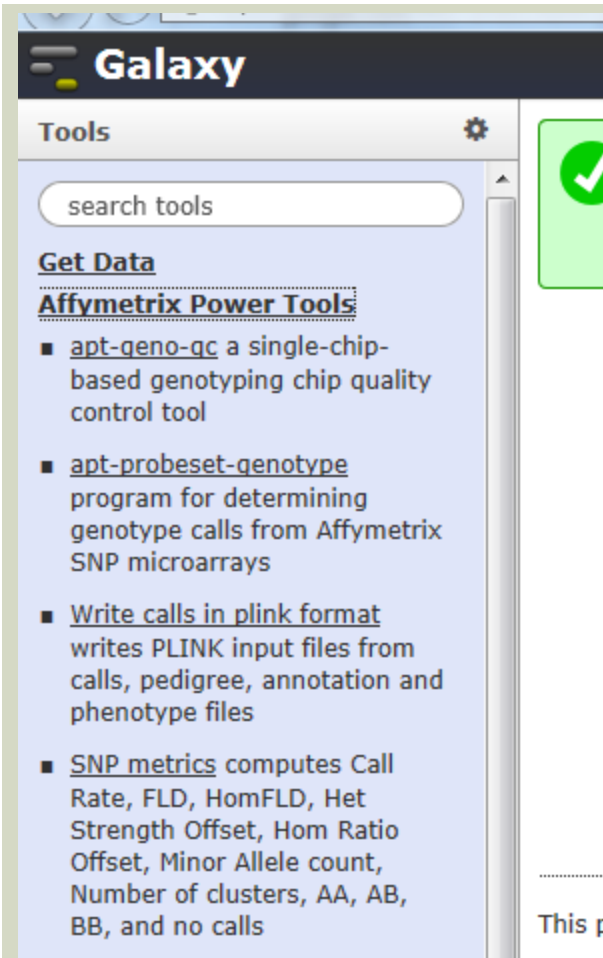
Affymetrix  
Power Tools

Genotyping  
Console

The screenshot shows the Affymetrix Genotyping Console interface. A menu is open, listing various analysis options such as 'Perform QC', 'Perform Genotyping...', 'Copy Number/LOH Analysis...', 'Copy Number Variation Analysis...', and 'Remove Intensity Data Group...'. Below the menu, a table displays genotyping results for multiple samples.

T	Channel	GC Channel	File ID	Computed Gender	#	CHP,CEL	File Date			
1	5	5	5.23	male	2	3/5/2012 1:23 PM				
2	51	51	5.73	female	2	3/5/2012 1:23 PM				
3	37	37	5.71	female	2	3/5/2012 1:23 PM				
4	25	25	5.53	male	2	3/5/2012 1:23 PM				
5	17	17	5.36	male	2	3/5/2012 1:23 PM				
6	16	16	5.70	male	2	3/5/2012 1:23 PM				
7	gC12_ACEL	in	2	0.950	3.77	3.61	5.44	male	2	3/5/2012 1:23 PM
8	gC1_ACEL	in	2	0.976	4.15	4.98	6.19	male	2	3/5/2012 1:23 PM
9	gC2_ACEL	in	2	0.890	3.28	2.70	4.67	female	1	3/5/2012 1:23 PM
10	gC3_ACEL	in	2	0.953	3.73	3.73	5.69	male	2	3/5/2012 1:23 PM
11	gC4_ACEL	in	2	0.953	3.84	3.75	5.49	female	2	3/5/2012 1:23 PM
12	gC5_ACEL	in	2	0.932	3.62	3.19	5.45	female	2	3/5/2012 1:23 PM
13	gC6_ACEL	in	2	0.918	3.51	3.17	5.39	male	2	3/5/2012 1:23 PM
14	gC7_ACEL	in	2	0.963	3.89	3.83	5.63	male	2	3/5/2012 1:23 PM
15	gC8_ACEL	in	2	0.956	3.87	3.91	5.49	male	2	3/5/2012 1:23 PM
16	gC9_ACEL	in	2	0.951	3.80	3.81	5.74	female	2	3/5/2012 1:23 PM
17	gD10_ACEL	in	2	0.958	3.84	3.90	5.66	female	2	3/5/2012 1:23 PM
18	gD11_ACEL	in	2	0.962	3.93	3.90	5.95	male	2	3/5/2012 1:23 PM
19	gD12_ACEL	in	2	0.958	3.90	4.28	5.73	male	2	3/5/2012 1:23 PM
20	gD1_ACEL	in	2	0.959	3.93	4.12	5.79	male	2	3/5/2012 1:23 PM
21	gD2_ACEL	in	2	0.941	3.64	3.34	5.42	male	2	3/5/2012 1:23 PM
22	gD3_ACEL	in	2	0.907	3.43	3.10	5.20	female	2	3/5/2012 1:23 PM
23	gD4_ACEL	in	2	0.947	3.81	3.73	5.72	male	2	3/5/2012 1:23 PM
24	gD5_ACEL	in	2	0.962	3.92	4.09	5.66	male	2	3/5/2012 1:23 PM
25	gD6_ACEL	in	2	0.932	3.61	3.27	5.35	male	2	3/5/2012 1:23 PM
26	gD7_ACEL	in	2	0.934	3.63	3.36	5.51	female	2	3/5/2012 1:23 PM
27	gD8_ACEL	in	2	0.961	3.97	4.14	5.78	male	2	3/5/2012 1:23 PM
28	gD9_ACEL	in	2	0.942	3.74	3.57	5.36	male	2	3/5/2012 1:23 PM
29	NC_ACEL	Out	2	0.955	0.08	0.12	0.25	male	0	3/5/2012 1:23 PM
30	P001_ACEL	in	2	0.955	3.44	3.05	5.18	male	2	3/5/2012 1:23 PM
31	P002_ACEL	in	2	0.915	3.44	3.43	5.47	male	2	3/5/2012 1:23 PM
32	P003_ACEL	in	2	0.939	3.60	3.44	5.49	female	2	3/5/2012 1:23 PM
33	P006_ACEL	in	2	0.957	3.88	4.43	5.60	female	2	3/5/2012 1:23 PM
34	P062_ACEL	in	2	0.910	3.43	3.19	5.04	female	2	3/5/2012 1:23 PM
35	P064_ACEL	in	2	0.943	3.70	3.86	5.39	female	2	3/5/2012 1:23 PM
36	P068_ACEL	in	2	0.970	4.06	4.57	6.09	female	2	3/5/2012 1:23 PM
37	P069_ACEL	in	2	0.902	3.40	2.77	5.12	male	2	3/5/2012 1:23 PM
38	P075_ACEL	in	2	0.910	3.42	3.11	5.14	male	2	3/5/2012 1:23 PM

# AFFYMETRIX POWER TOOLS IN GALAXY



The screenshot shows the Galaxy web interface. At the top, the 'Galaxy' logo is visible. Below it, the 'Tools' panel is open, featuring a search bar labeled 'search tools'. Under the 'Get Data' section, the 'Affymetrix Power Tools' category is expanded, listing several tools with their descriptions:

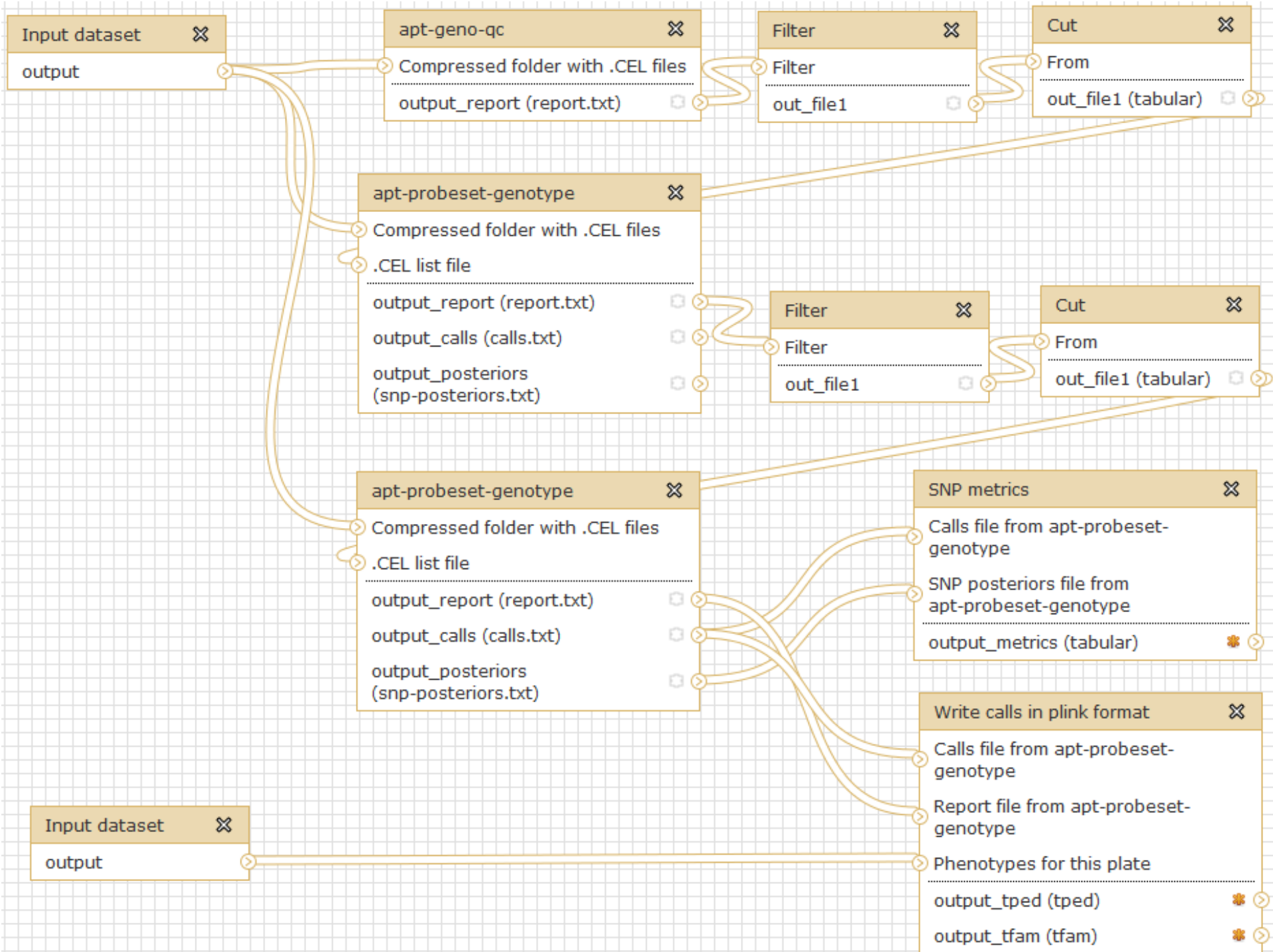
- [apt-geno-qc](#) a single-chip-based genotyping chip quality control tool
- [apt-probeset-genotype](#) program for determining genotype calls from Affymetrix SNP microarrays
- [Write calls in plink format](#) writes PLINK input files from calls, pedigree, annotation and phenotype files
- [SNP metrics](#) computes Call Rate, FLD, HomFLD, Het Strength Offset, Hom Ratio Offset, Minor Allele count, Number of clusters, AA, AB, BB, and no calls

At the bottom of the panel, the text 'This p' is partially visible.

- **apt-geno-qc** for quality control
- **apt-probeset-genotype** for genotype calling
- Write calls in plink format
- SNP metrics

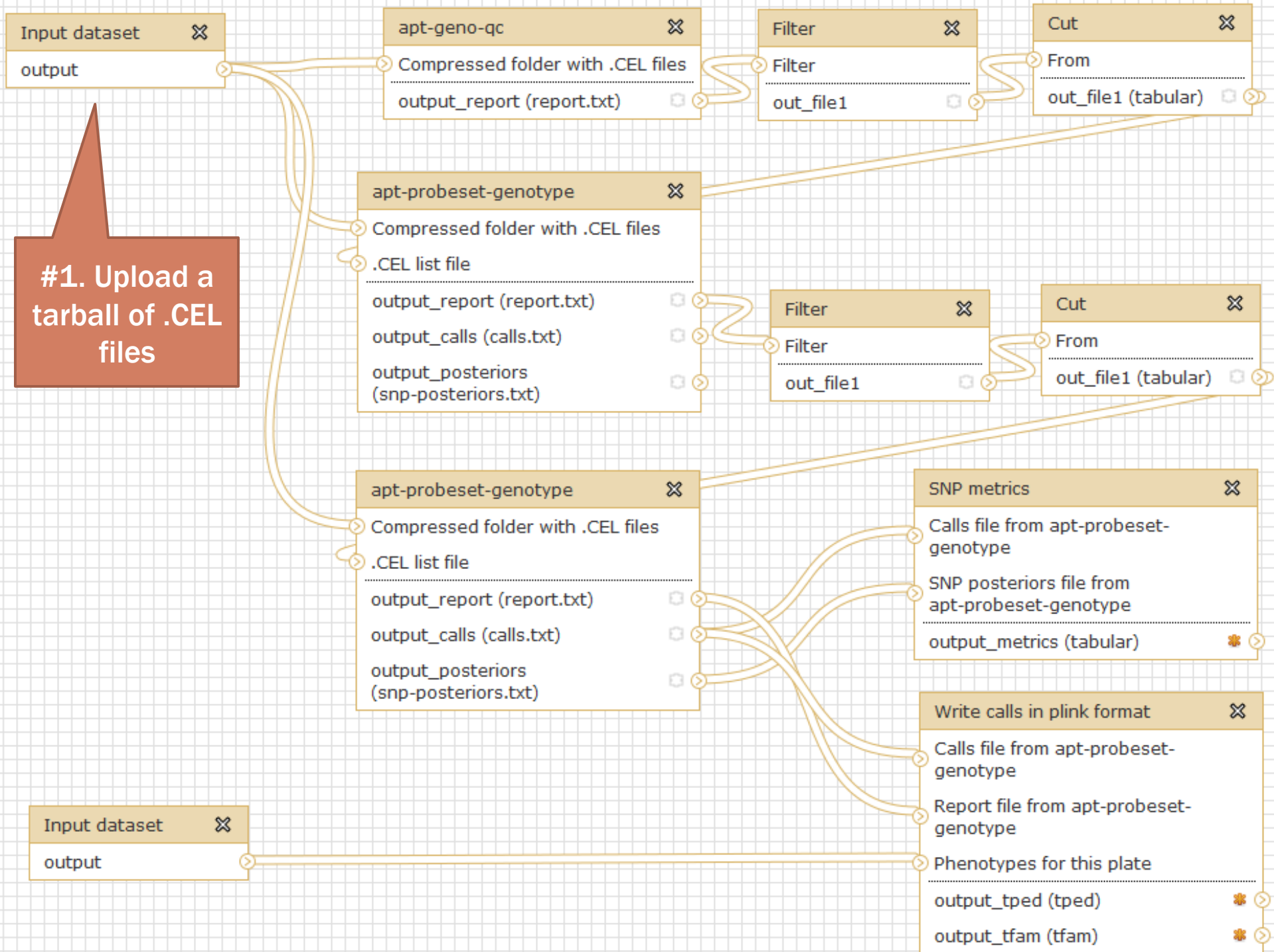
# MINOR CUSTOMIZATIONS

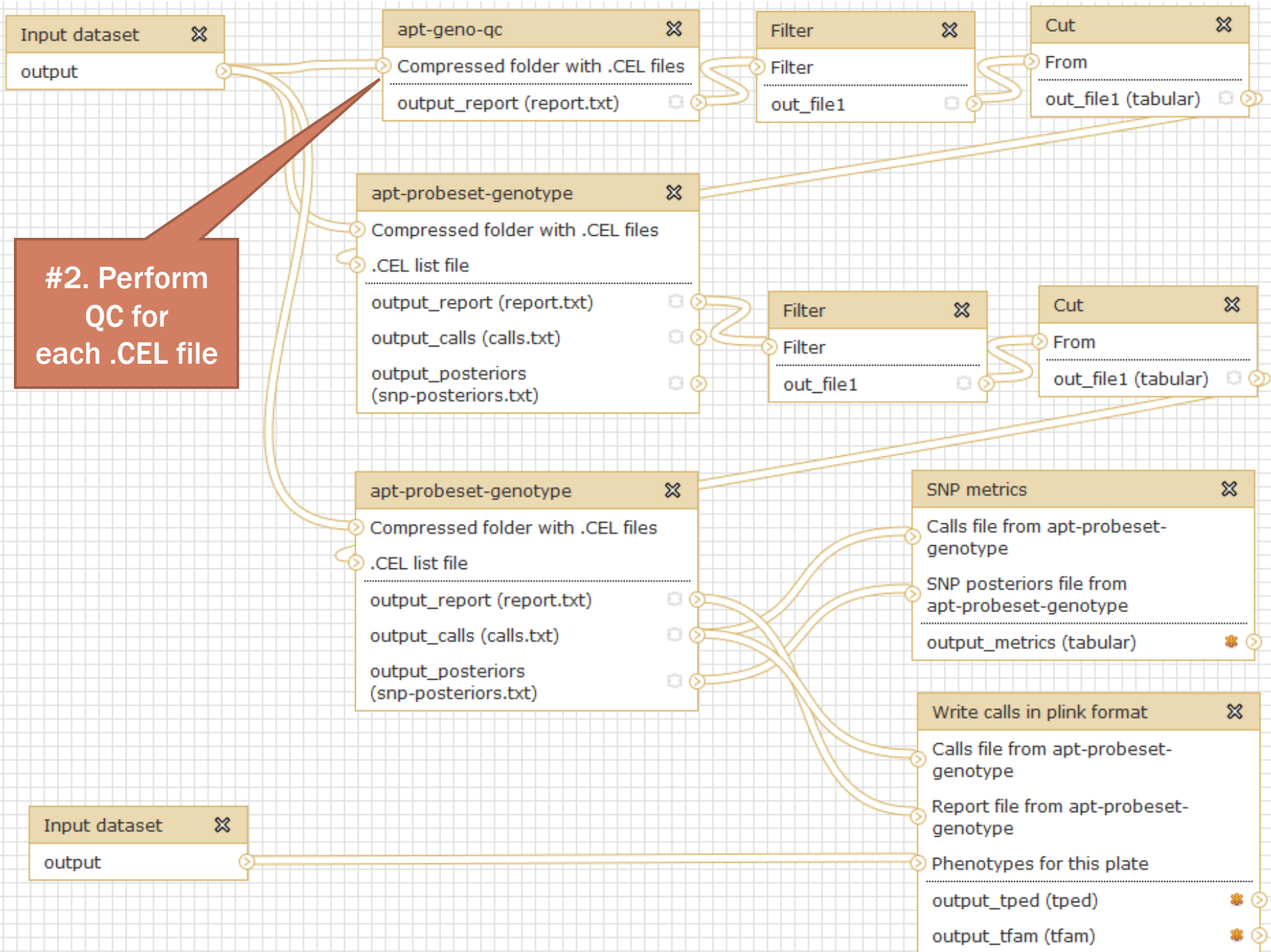
- Registered ".CEL.TAR" file formats for batch uploads
- Modified Galaxy upload tool for better handling custom binary files
- Registered APT ".calls.txt", ".report.txt", "snp-posteriors.txt" file formats
- Registered PLINK formats ".tped", "tmap"

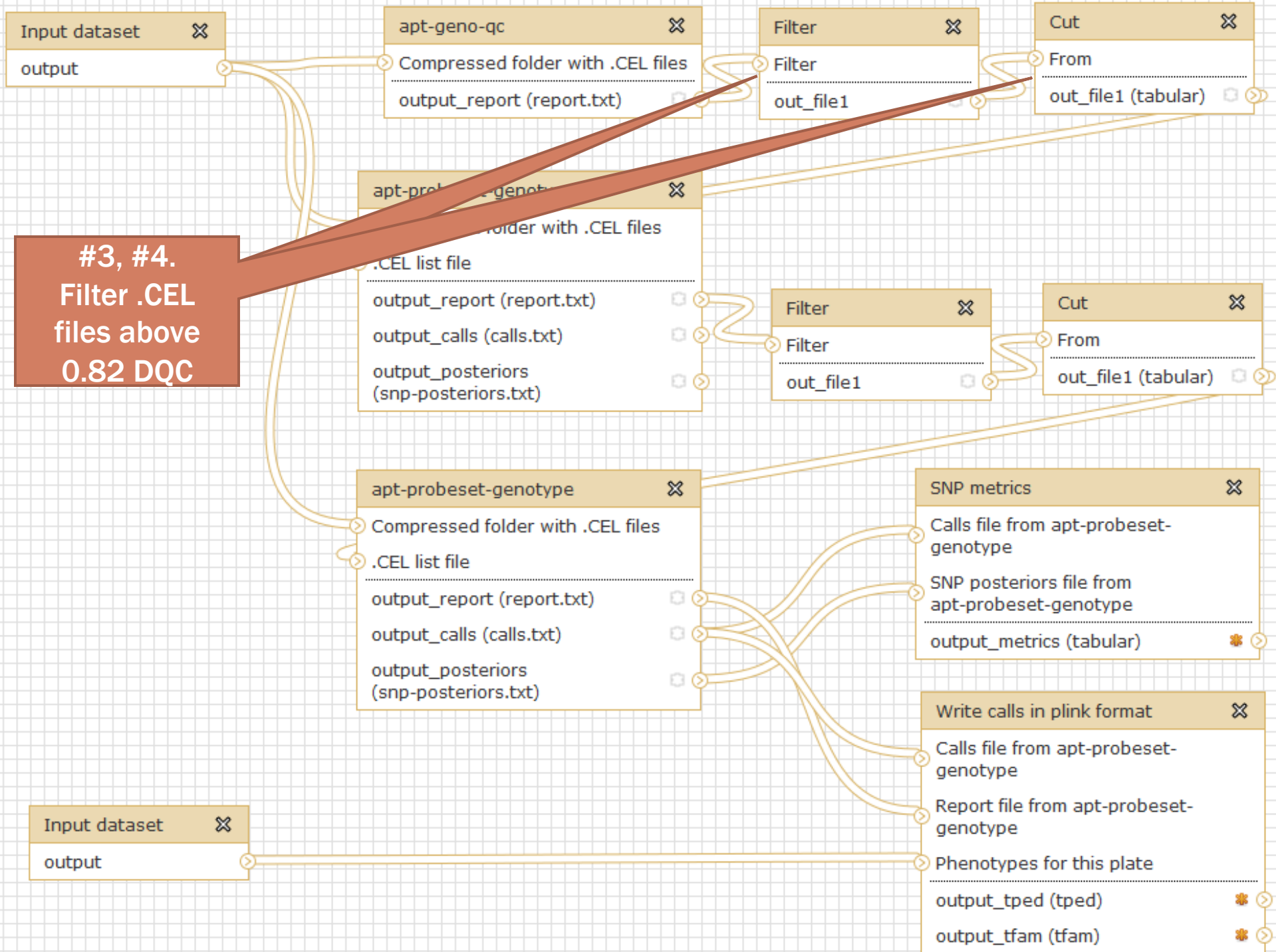


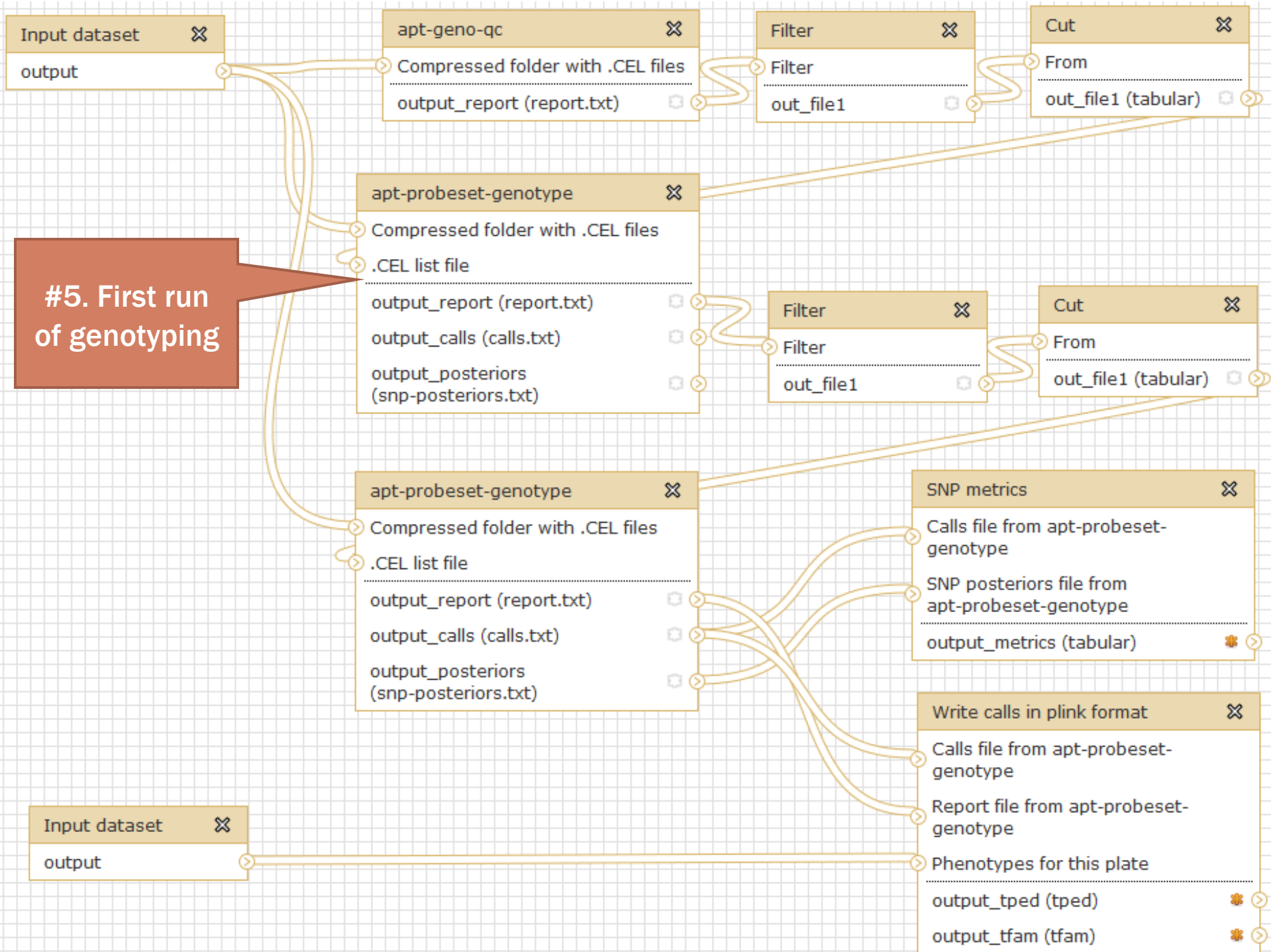


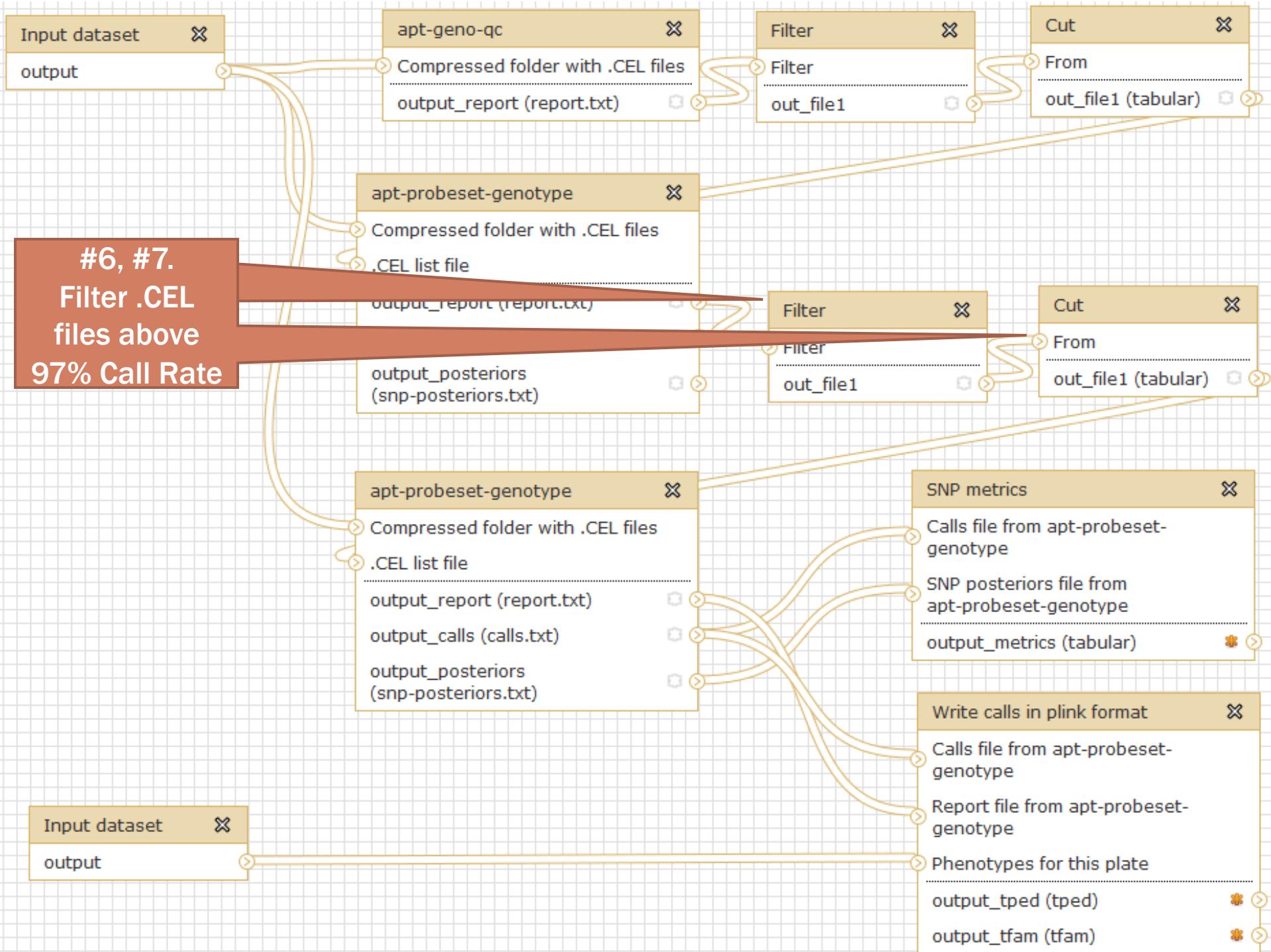
#1. Upload a tarball of .CEL files



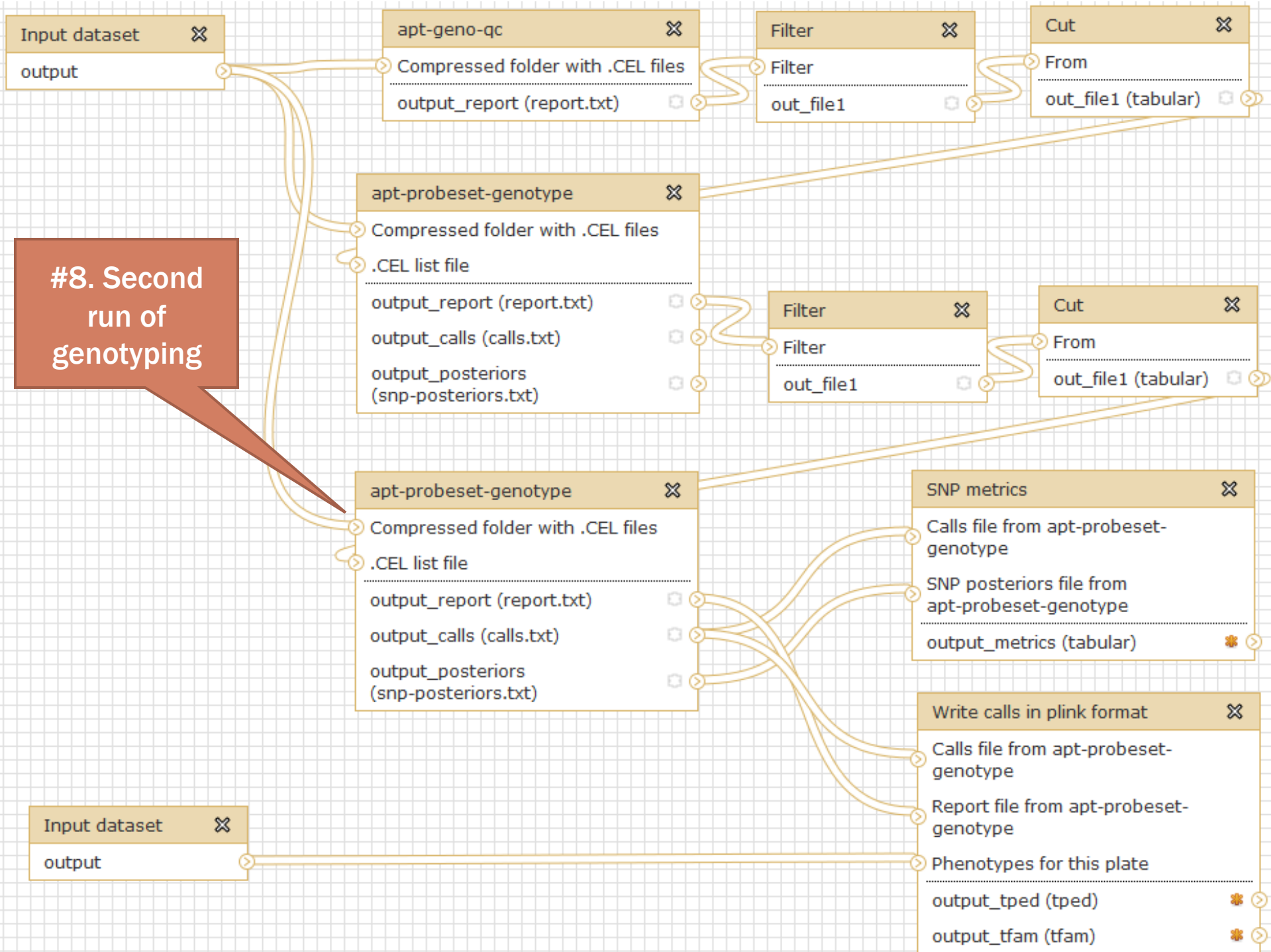






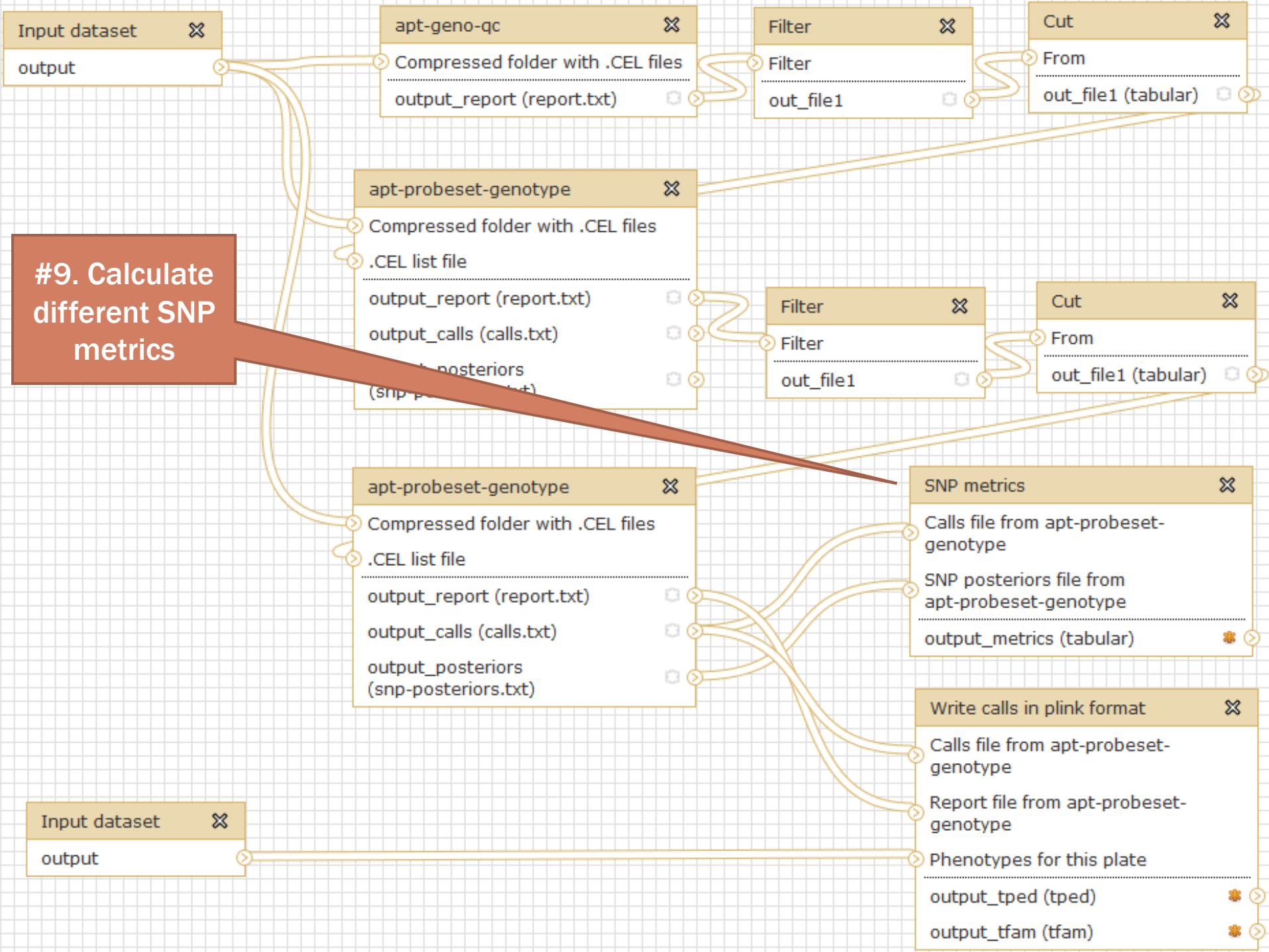


#6, #7.  
Filter .CEL files above 97% Call Rate

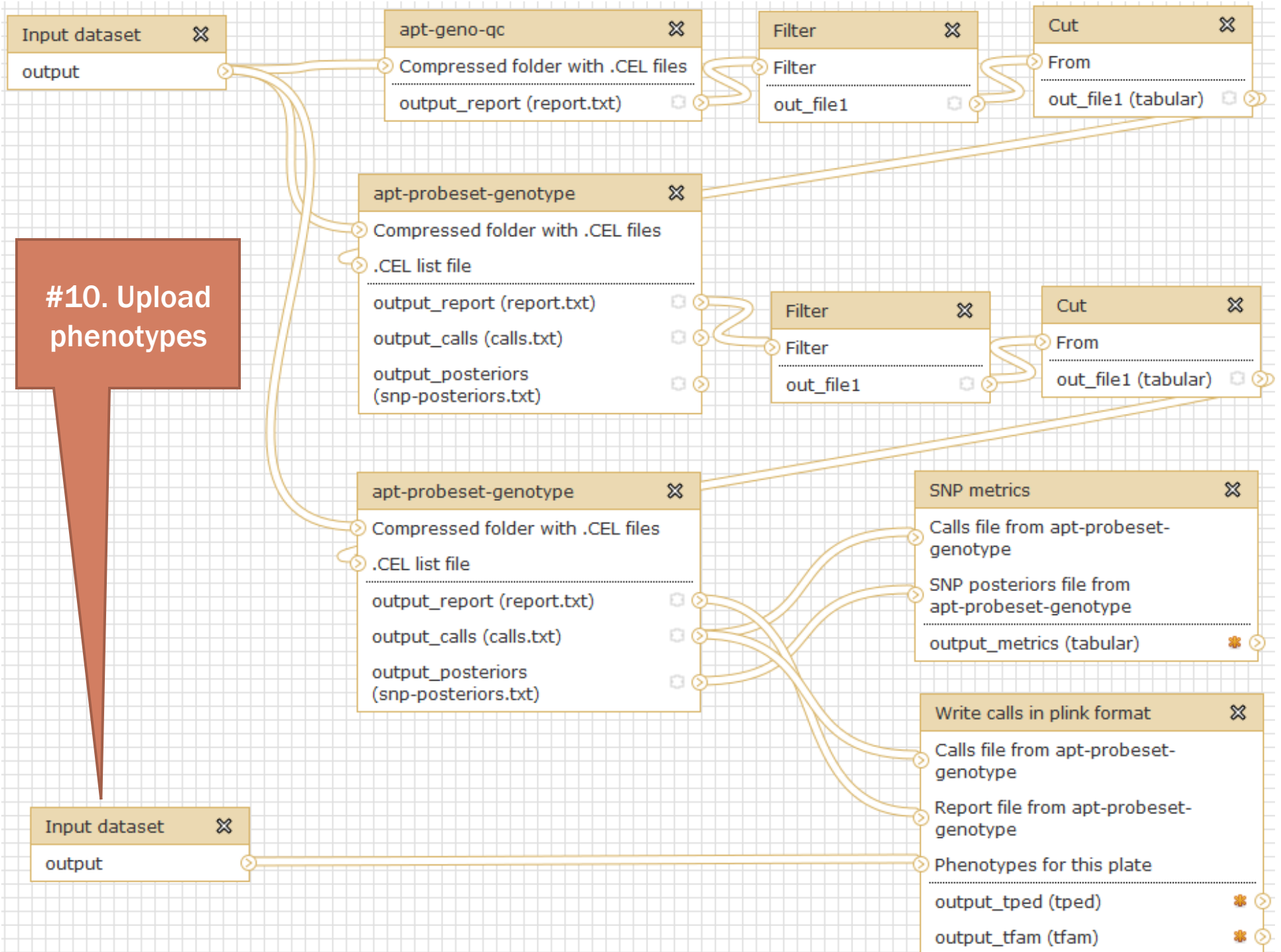




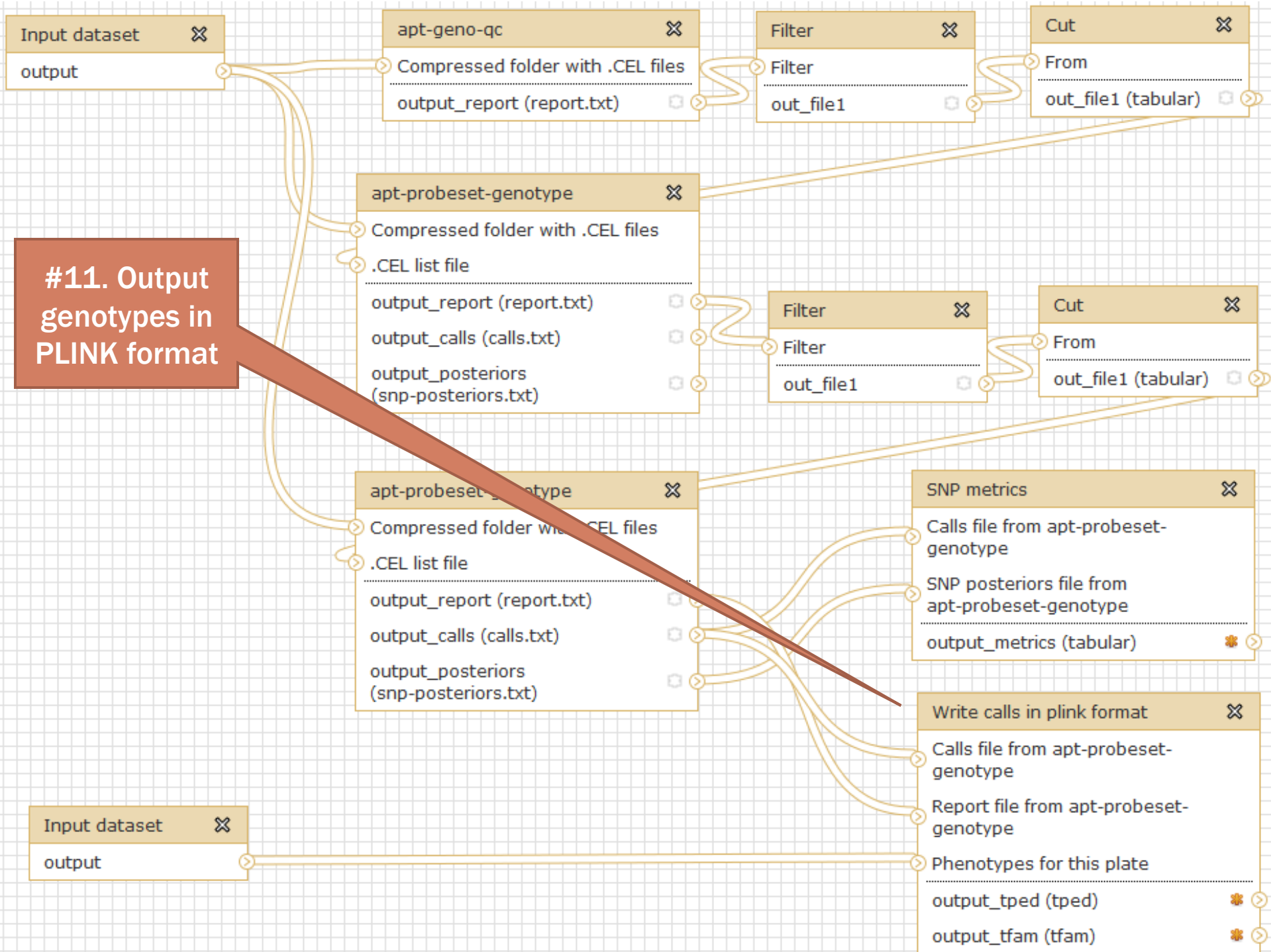
#9. Calculate different SNP metrics



## #10. Upload phenotypes







#11. Output genotypes in PLINK format

# FUTURE WORK

- Integrate a few more Affymetrix Power Tools into Galaxy
- Publish the APT tools in Galaxy Tool Shed
- Better support for binary/compressed data files
- Publish the APT genotyping workflow

# ACKNOWLEDGEMENTS

- Research Resources Center and Center for Clinical and Translational Science at UIC
- Dr. Damir Herman, Affymetrix
- Morris Chukhman, Tommie Jackson, and the team at the Center for Research Informatics at UIC
- Dr. Rick Kittles, UIC Institute of Human Genetics
- Dr. Zarema Arbieva, UIC Core Genomics Facility