

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      CreatedOMDocument          probeSet-summarize-quick-test
apt-calvin-equivalent          apt-rt-chp-to-txt           createTav                   probeSet-summarize-quick-test.valgrind
apt-calvinlite-test           apt-rt-copynumber-cyto     dmet-genotype-test         probeSet-summarize-test
apt-calvinlite-util           apt-rt-dmet-genotype       dmet-translation-test     PSVIMriter
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-cluster-core-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-board
apt-dmet-copynumber           calvin-compare-cel         h5jam                        test-dagb
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-workflow-quick-test
apt-ma5                        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-stages  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        tsv-join-test                 tsv-join-test
apt-rr-cdf-export             chp-to-txt-quick-test      tsv-util                       tsv-util
apt-rt-cel-convert            chp-to-txt-test            util-meminfo                    test-tsv-util
apt-rt-cel-extract            convert-priors               xda-chp-compare                test-tsv-util
okarpe2@nike bin]$
```

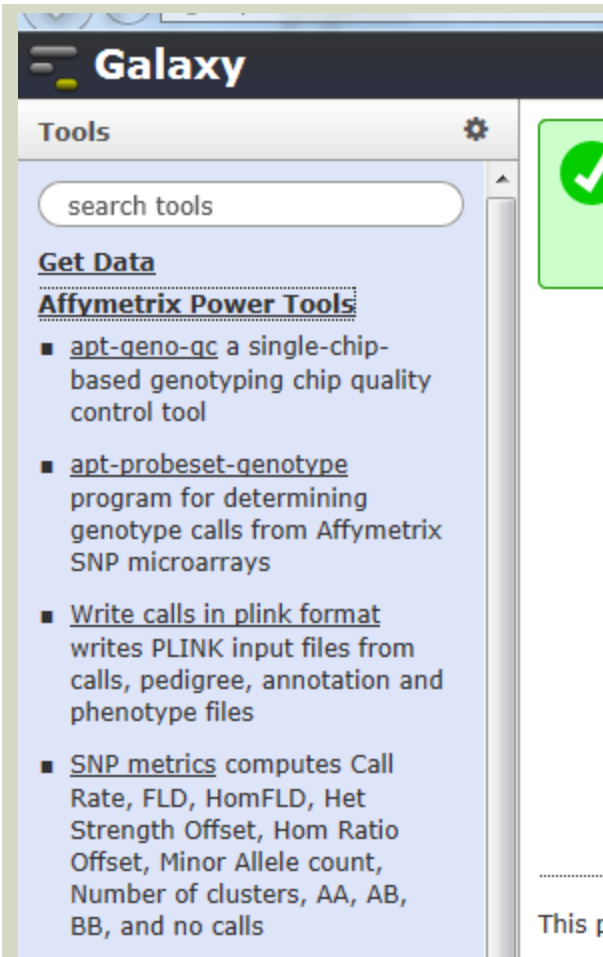
Affymetrix
Power Tools

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 Results', and 'Remove Intensity Data Group...'. Below the menu, a table displays genotyping results for multiple samples.

T	Channel	GC Channel	File	Gender	#	CHP	CEL	File Date
6	50	5.23		male	2			3/5/2012 1:23 PM
7	51	5.73		female	2			3/5/2012 1:23 PM
8	52	5.71		female	2			3/5/2012 1:23 PM
9	53	5.53		male	2			3/5/2012 1:23 PM
10	54	5.36		male	2			3/5/2012 1:23 PM
11	55	5.70		male	2			3/5/2012 1:23 PM
12	56	5.44		male	2			3/5/2012 1:23 PM
13	57	6.19		male	2			3/5/2012 1:23 PM
14	58	4.67		female	1			3/5/2012 1:23 PM
15	59	5.69		male	2			3/5/2012 1:23 PM
16	60	3.75		female	2			3/5/2012 1:23 PM
17	61	3.19		female	2			3/5/2012 1:23 PM
18	62	5.38		male	2			3/5/2012 1:23 PM
19	63	5.63		male	2			3/5/2012 1:23 PM
20	64	5.49		male	2			3/5/2012 1:23 PM
21	65	3.81		female	2			3/5/2012 1:23 PM
22	66	5.74		female	2			3/5/2012 1:23 PM
23	67	3.80		female	2			3/5/2012 1:23 PM
24	68	3.90		male	2			3/5/2012 1:23 PM
25	69	5.95		male	2			3/5/2012 1:23 PM
26	70	4.28		male	2			3/5/2012 1:23 PM
27	71	4.12		male	2			3/5/2012 1:23 PM
28	72	3.34		male	2			3/5/2012 1:23 PM
29	73	5.42		male	2			3/5/2012 1:23 PM
30	74	3.10		female	2			3/5/2012 1:23 PM
31	75	5.72		male	2			3/5/2012 1:23 PM
32	76	3.73		male	2			3/5/2012 1:23 PM
33	77	4.09		male	2			3/5/2012 1:23 PM
34	78	3.27		male	2			3/5/2012 1:23 PM
35	79	3.36		female	2			3/5/2012 1:23 PM
36	80	5.51		female	2			3/5/2012 1:23 PM
37	81	4.14		male	2			3/5/2012 1:23 PM
38	82	3.57		male	2			3/5/2012 1:23 PM
39	83	0.08		male	0			3/5/2012 1:23 PM
40	84	3.05		male	2			3/5/2012 1:23 PM
41	85	5.18		male	2			3/5/2012 1:23 PM
42	86	3.43		male	2			3/5/2012 1:23 PM
43	87	3.44		male	2			3/5/2012 1:23 PM
44	88	3.47		male	2			3/5/2012 1:23 PM
45	89	3.44		female	2			3/5/2012 1:23 PM
46	90	5.36		male	2			3/5/2012 1:23 PM
47	91	5.04		female	2			3/5/2012 1:23 PM
48	92	5.04		female	2			3/5/2012 1:23 PM
49	93	3.86		female	2			3/5/2012 1:23 PM
50	94	6.09		female	2			3/5/2012 1:23 PM
51	95	2.77		male	2			3/5/2012 1:23 PM
52	96	5.12		male	2			3/5/2012 1:23 PM
53	97	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:

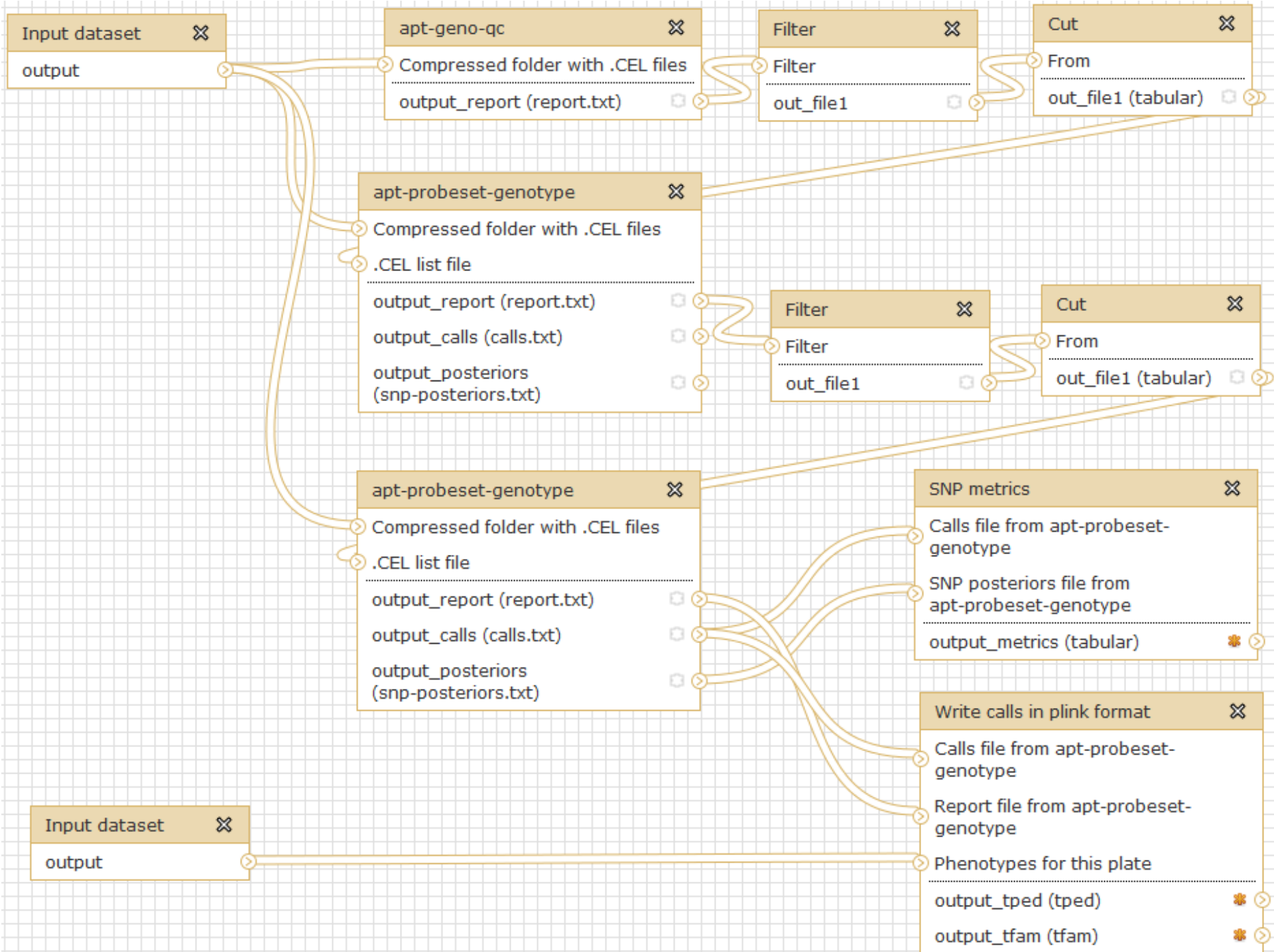
- [apt-geno-qc](#) a single-chip-based genotyping chip quality control tool
- [apt-probeset-genotype](#) program for determining genotype calls from Affymetrix SNP microarrays
- [Write calls in plink format](#) writes PLINK input files from calls, pedigree, annotation and phenotype files
- [SNP metrics](#) 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.

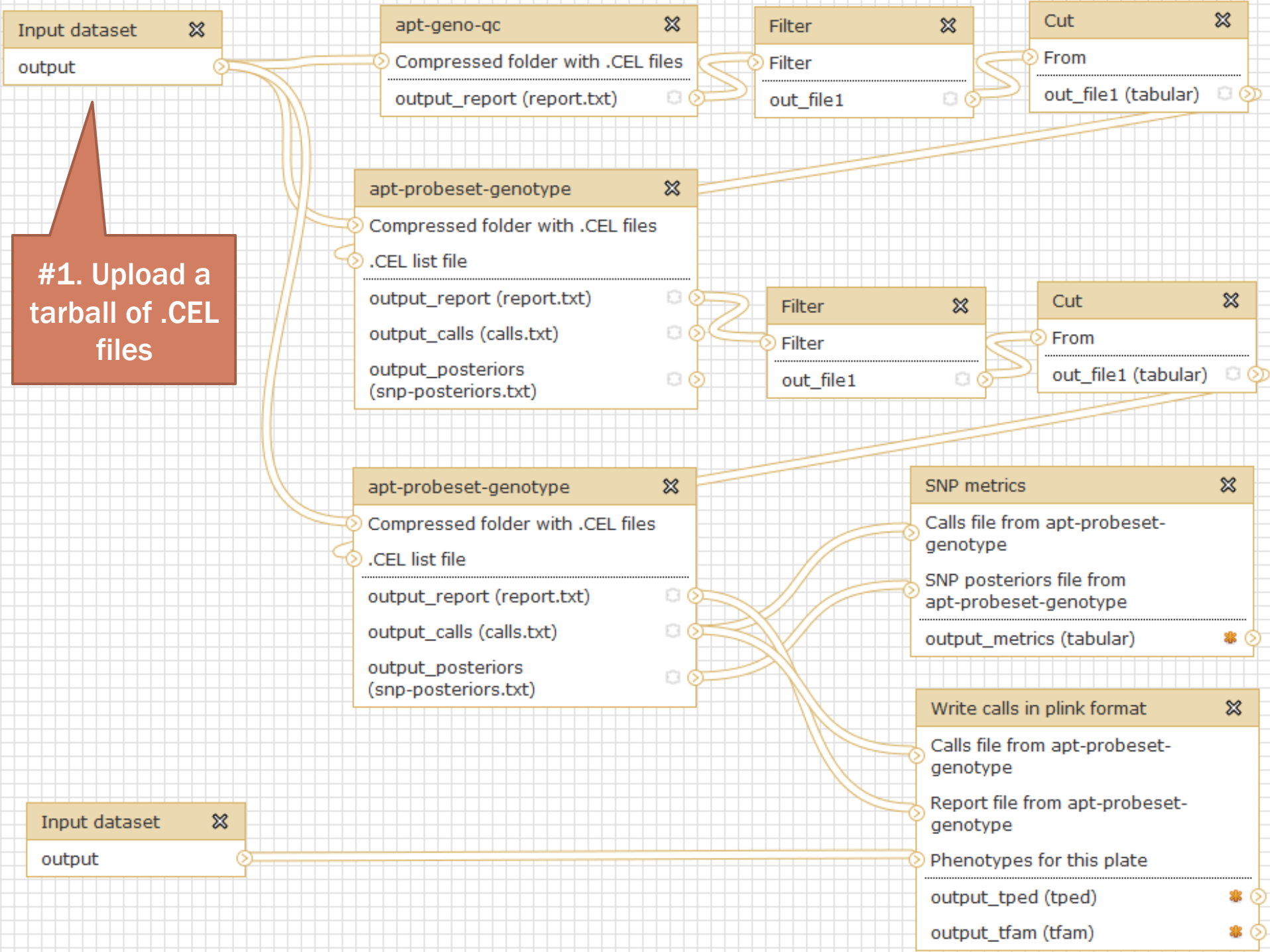
- **apt-geno-qc** for quality control
- **apt-probeset-genotype** for genotype calling
- Write calls in plink format
- SNP metrics

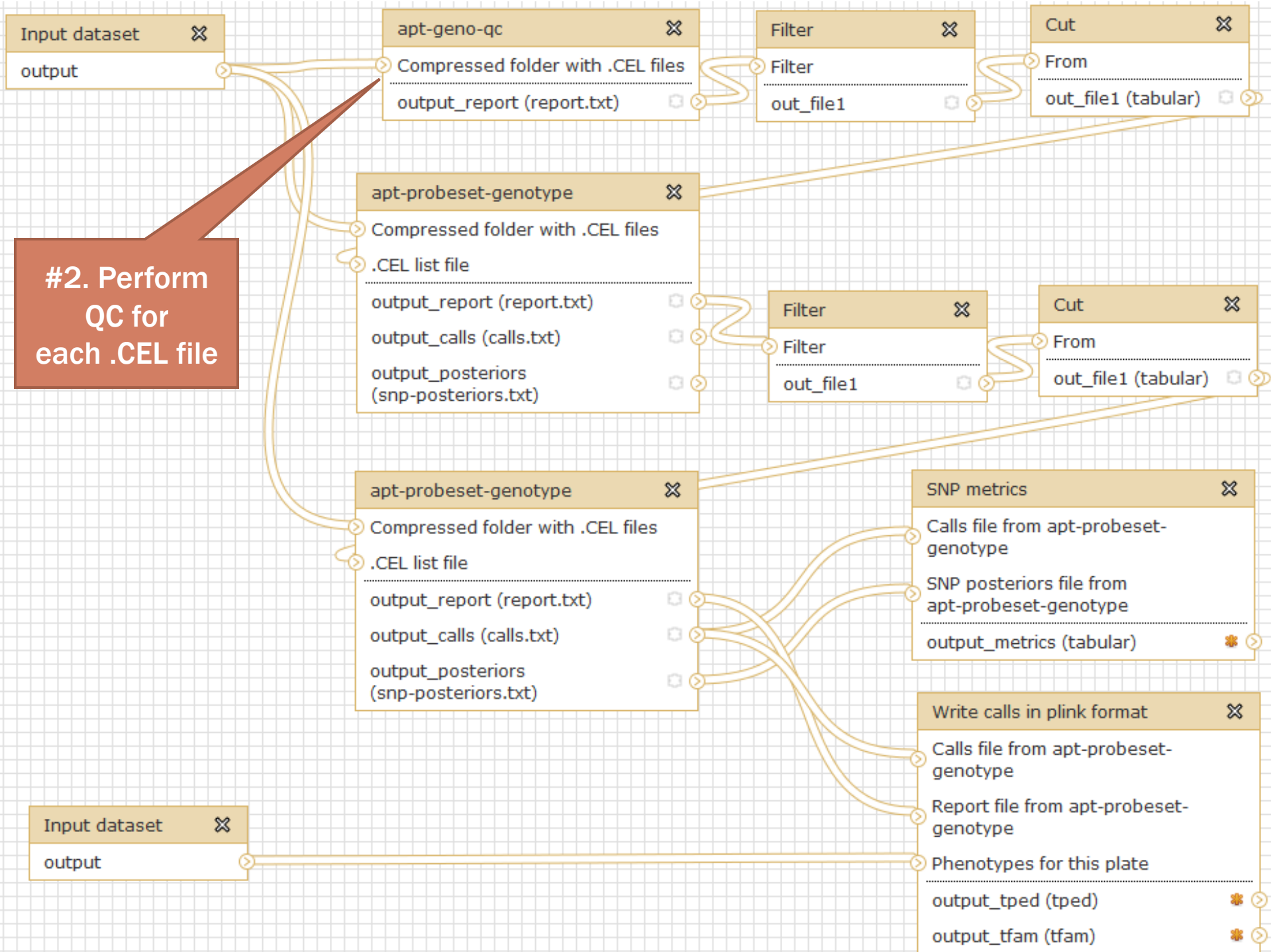
MINOR CUSTOMIZATIONS

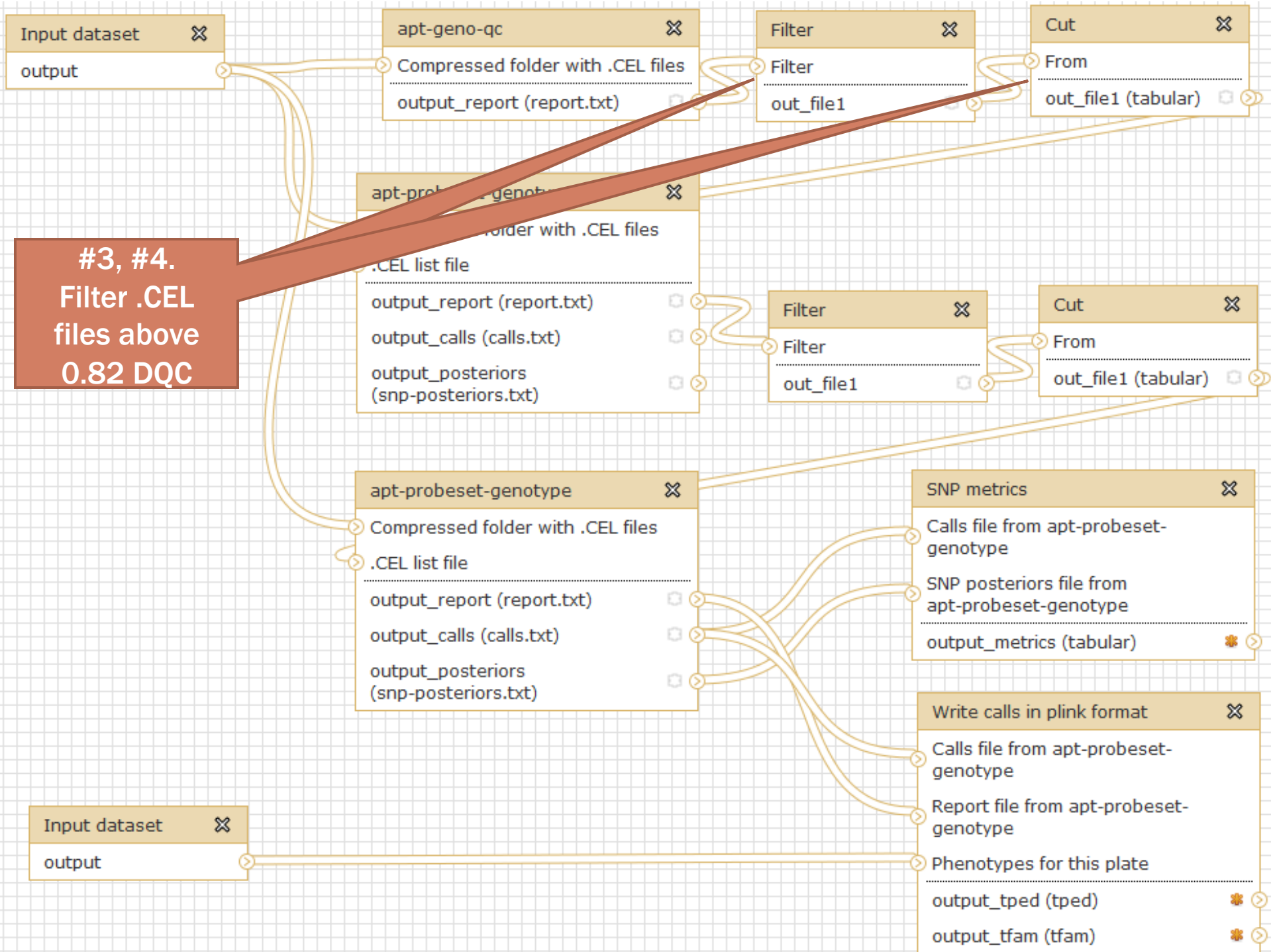
- Registered ".CEL.TAR" file formats for batch uploads
- Modified Galaxy upload tool for better handling custom binary files
- Registered APT ".calls.txt", ".report.txt", "snp-posteriors.txt" file formats
- Registered PLINK formats ".tped", "tmap"

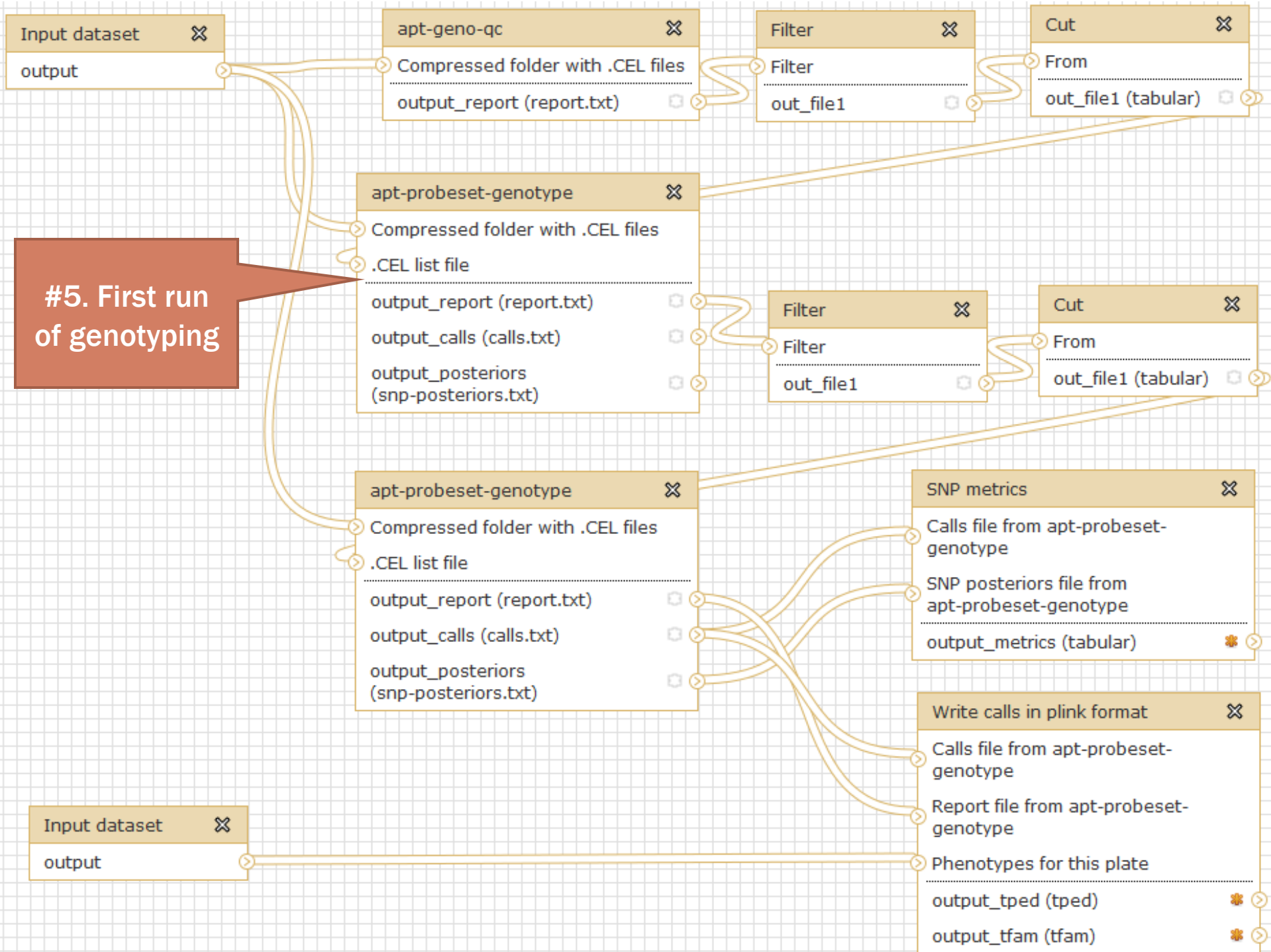


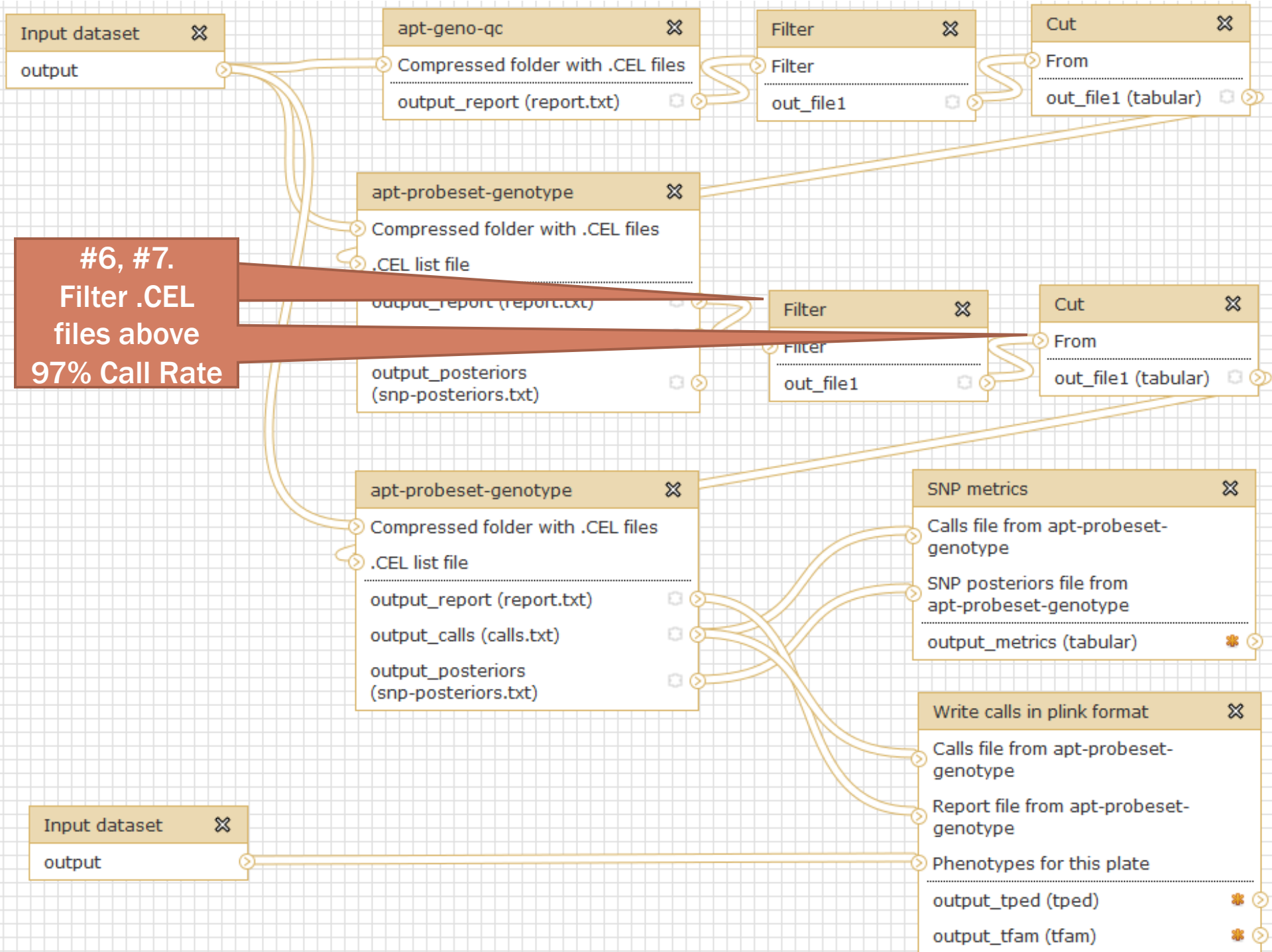
#1. Upload a tarball of .CEL files

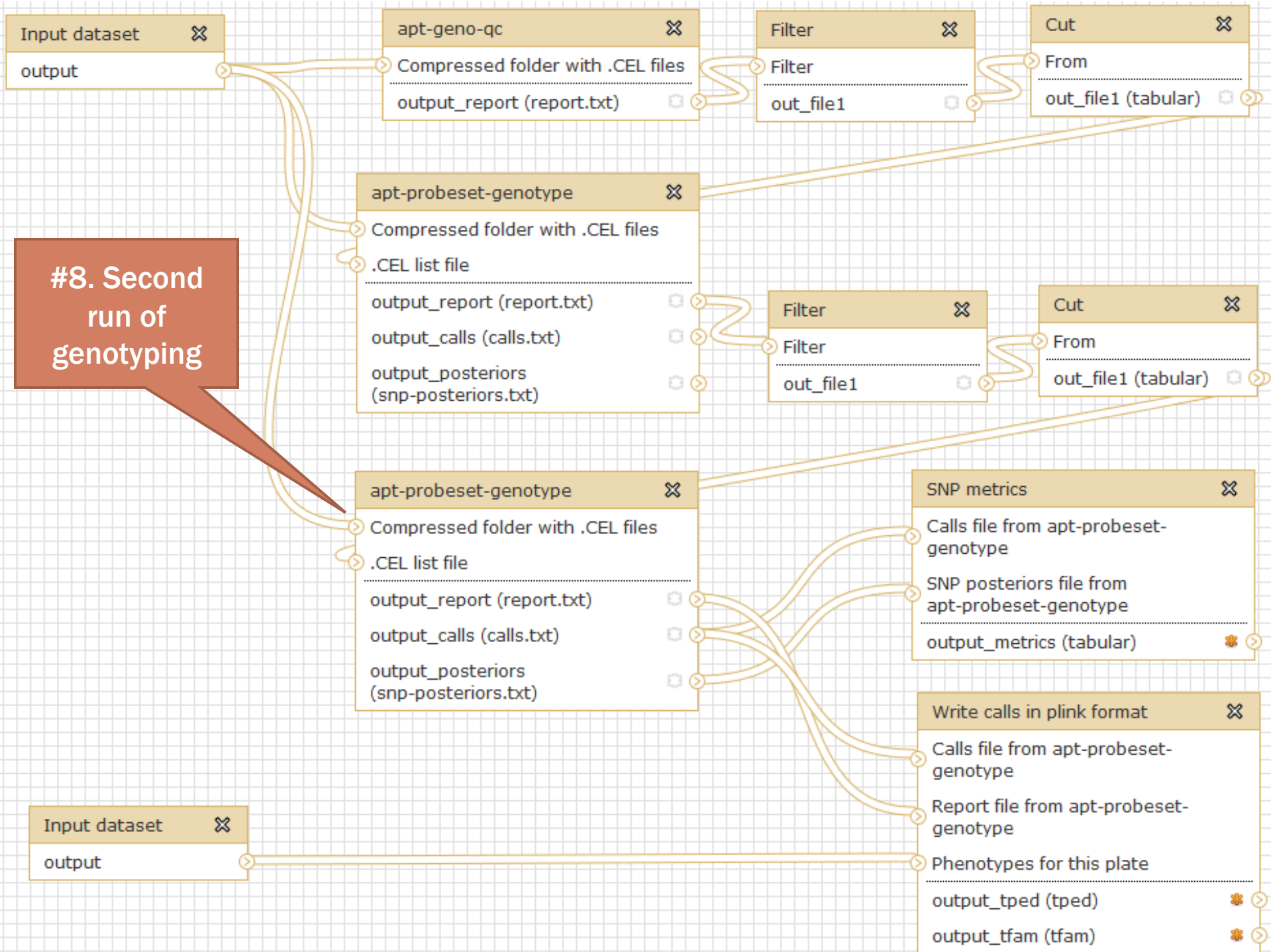


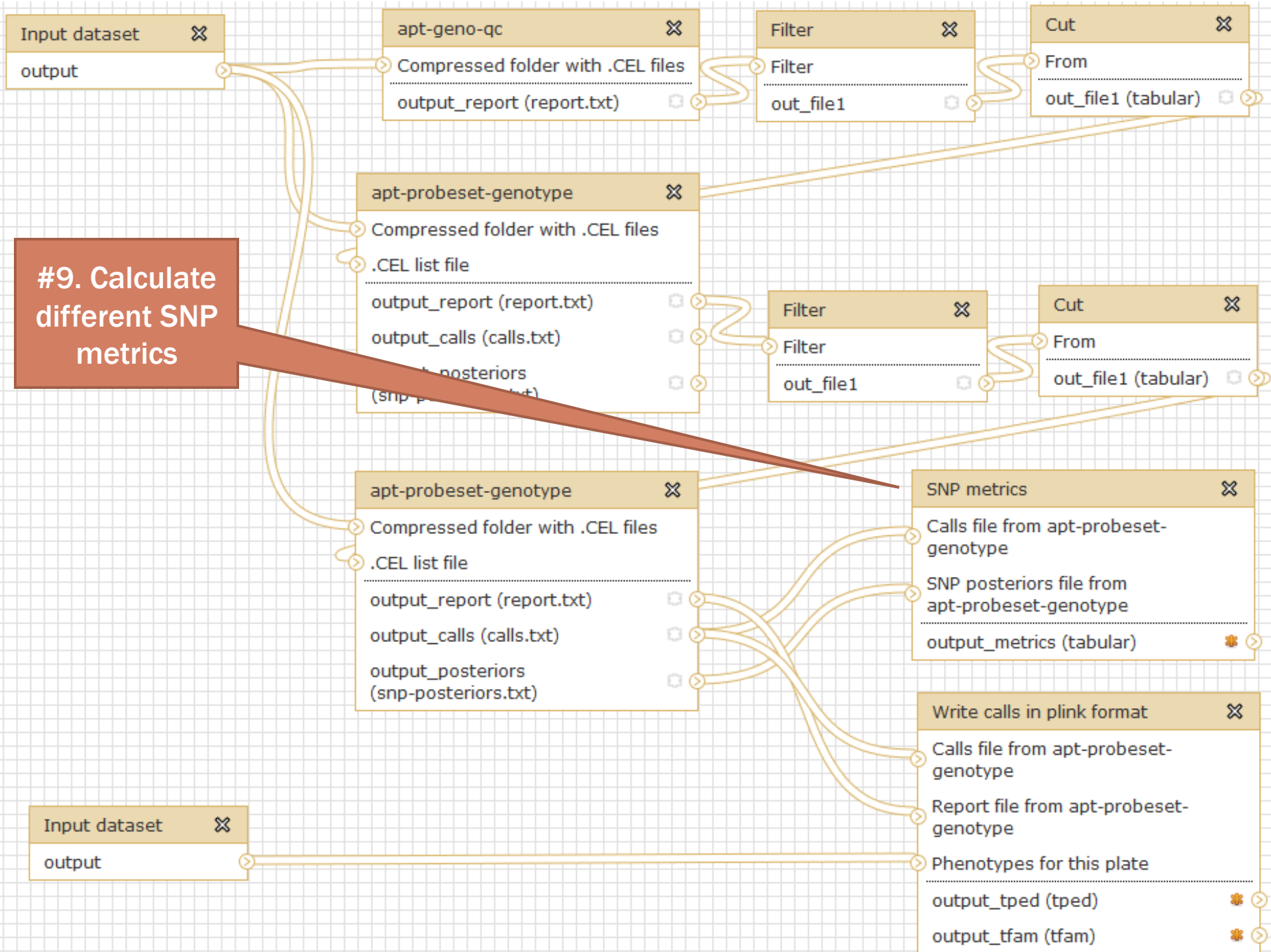












Input dataset ✕
output

apt-geno-qc ✕
Compressed folder with .CEL files
output_report (report.txt)

Filter ✕
Filter
out_file1

Cut ✕
From
out_file1 (tabular)

#10. Upload phenotypes

apt-probeset-genotype ✕
Compressed folder with .CEL files
.CEL list file
output_report (report.txt)
output_calls (calls.txt)
output_posteriors (snp-posteriors.txt)

Filter ✕
Filter
out_file1

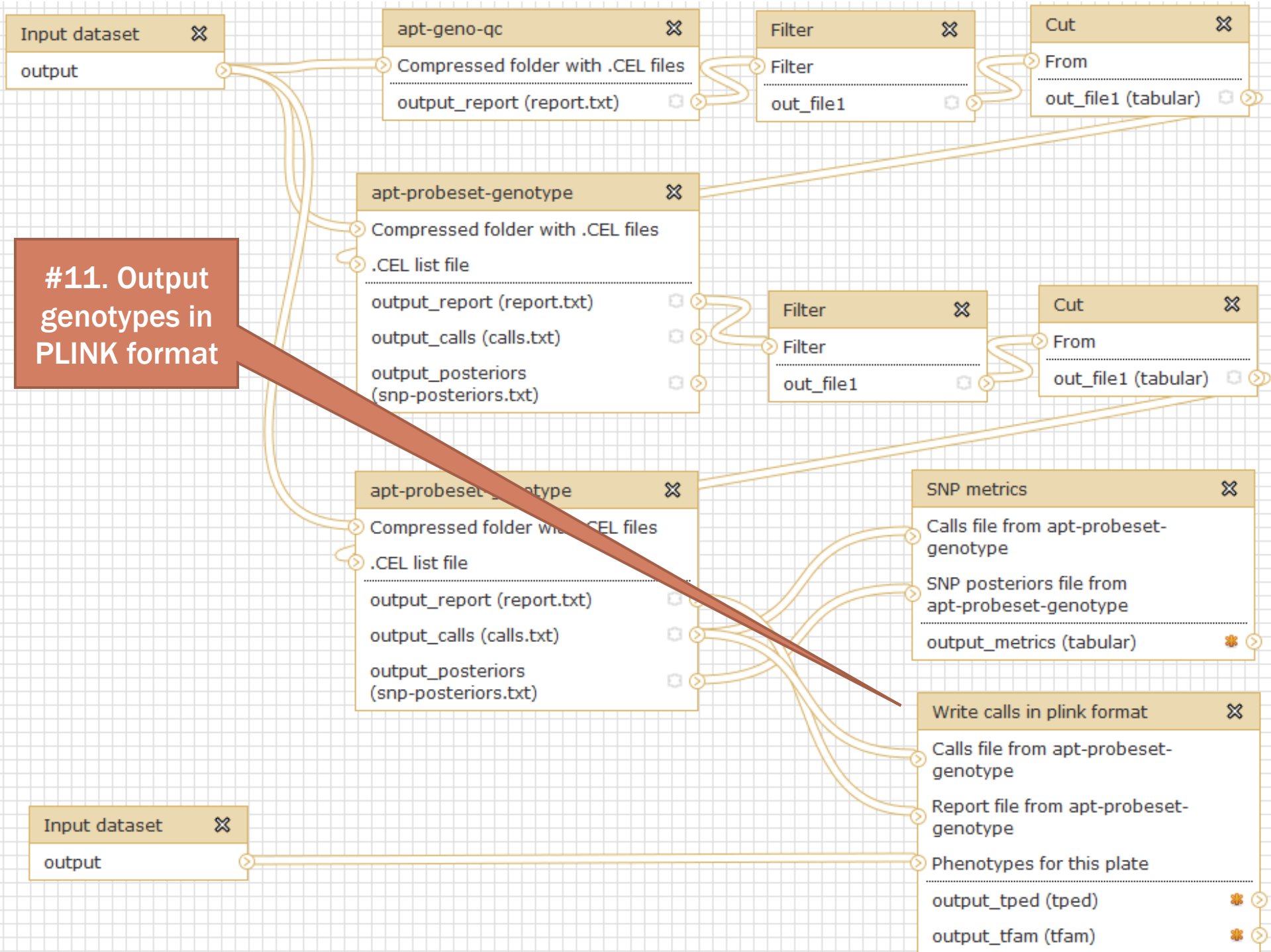
Cut ✕
From
out_file1 (tabular)

apt-probeset-genotype ✕
Compressed folder with .CEL files
.CEL list file
output_report (report.txt)
output_calls (calls.txt)
output_posteriors (snp-posteriors.txt)

SNP metrics ✕
Calls file from apt-probeset-genotype
SNP posteriors file from apt-probeset-genotype
output_metrics (tabular)

Write calls in plink format ✕
Calls file from apt-probeset-genotype
Report file from apt-probeset-genotype
Phenotypes for this plate
output_tped (tped)
output_tfam (tfam)

Input dataset ✕
output



#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