Tool		Sigflow	SigProfiler	SomaticSignatures	MutationalPatterns	deconstructSigs
Support signature type	SBS	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
	DBS	**Yes**	**Yes**	No	No	**Yes**
	INDEL	**Yes**	**Yes**	No	No	No
	CN	**Yes**	No	No	No	No
Has CLI		**Yes**	No	No	No	No
Platform		R	Python	R	R	R
Analysis features	Manual extraction	**Yes**	No	**Yes**	**Yes**	No
	Auto extraction	**Yes**	**Yes**	No	No	No
	Refit after extraction	**Yes**	**Yes**	No	No	No
	Reference signature fitting (efficiency)	**Yes (fast)**	No	No	**Yes (fast)**	**Yes (slow)**
	Batch signature fitting	**Yes**	No	No	**Yes**	No
	Signature stability analysis	**Yes**	No	No	No	No
	Visualization	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Extensible		**Yes**	**Yes**	No	No	No
Core methods		NMF/Bayesian NMF/QP/NNLS	NMF/NNLS	NMF/PCA	NMF/NNLS	NNLS
Signature quantification	Relative exposure	**Yes**	No	No	**Yes**	**Yes**
	Absolute exposure	**Yes**	**Yes**	No	Yes	No
Input format		VCF/MAF/CSV/EXCEL	VCF/MAF/CSV	VCF/R matrix/R Vranges object	VCF/R matrix	Custom mutation file
Software URL		https://github.com/ShixiangWang/sigflow	https://github.com/AlexandrovLab/SigProfilerExtractor	https://github.com/juliangehring/SomaticSignatures	https://github.com/UMCUGenetics/MutationalPatterns	https://github.com/raerose01/deconstructSigs
Abbr.: SBS for single base substitution; DBS for doublet base substitution; INDEL for short insertions and deletions; CN for copy number; SV for structure variation; CLI for command line interface; NMF for non-negative factorization; QP for quadratic programming; NNLS for nonnegative least square; PCA for principal component analysis; VCF for variant call format; MAF for mutation annotation format						

Table S1. Comparisons between Sigflow and other mutational signature analysis tools. Note: "Refit after extraction" means signature fitting with signatures from de-novo extraction, and this could optimize the exposures of extracted signatures. "Batch signature fitting" means signature fitting in multiple samples simultaneously. "Relative exposure" means relative mutation proportion contributed by a signature in a tumor. "Absolute exposure" means mutation counts contributed by a signature in a tumor.
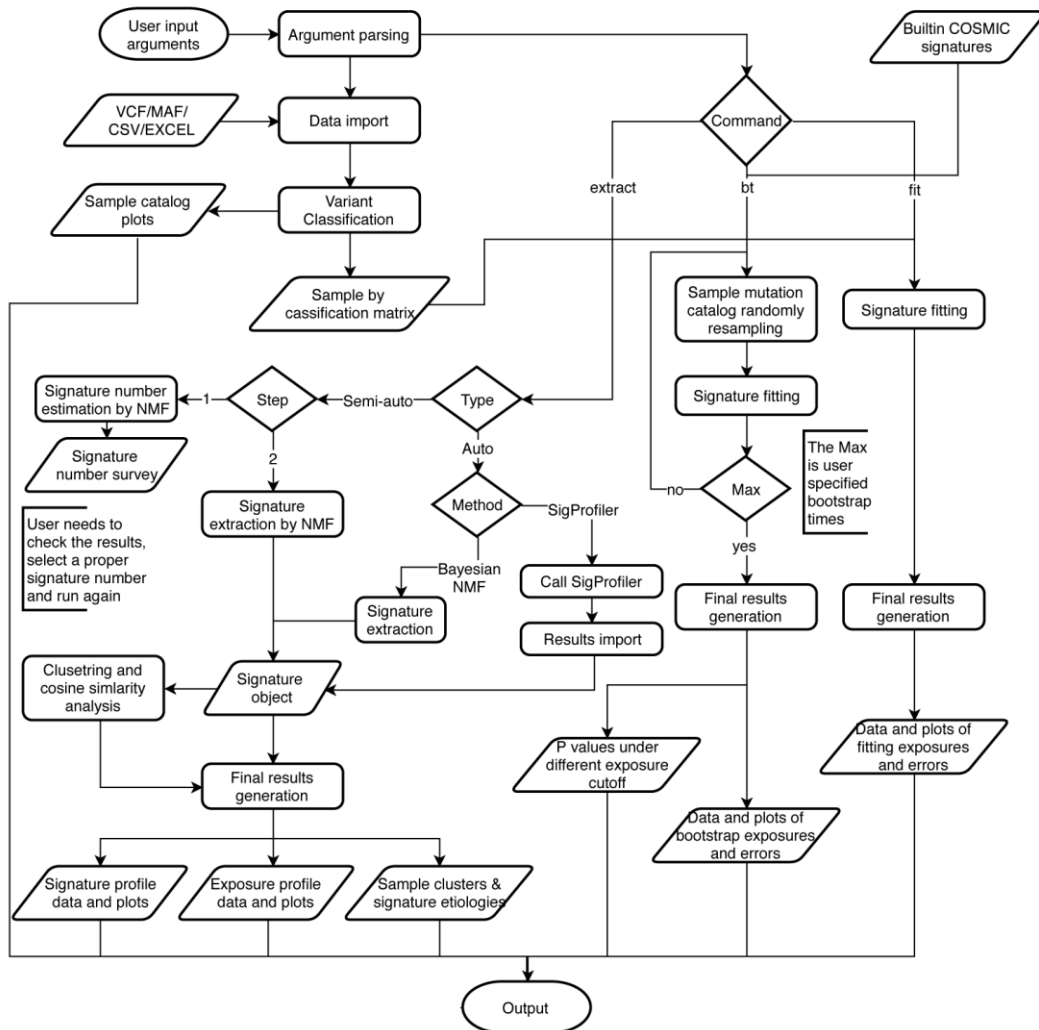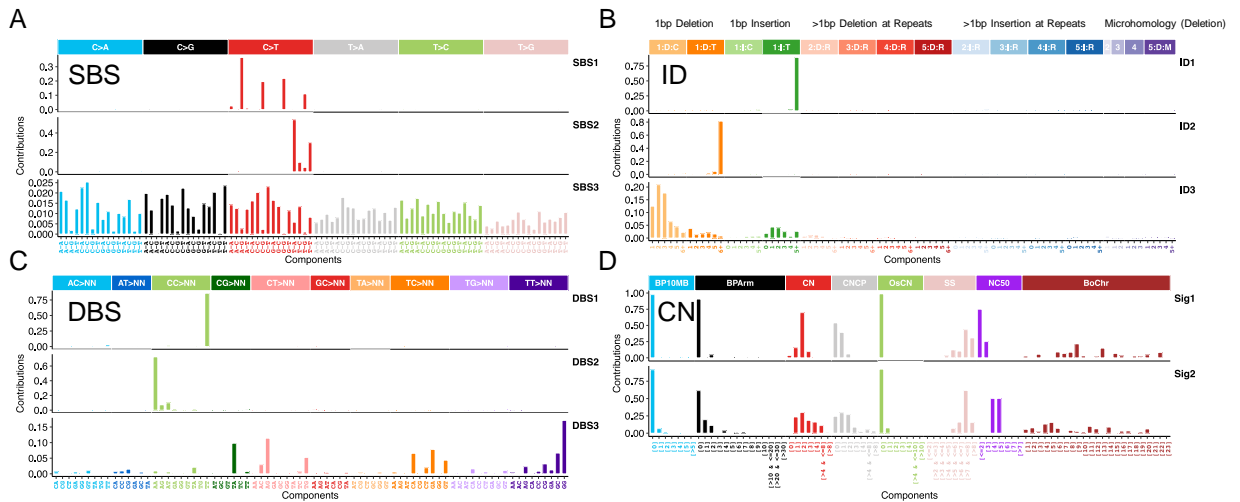
Fig. S1. Detailed flow chart of Sigflow.

Fig. S2. Example signature profiles generated by Sigflow. First 3 signatures for mutational types including SBS (A), ID (B) and DBS (C) catalogued in COSMIC signature v3 database (https://cancer.sanger.ac.uk/cosmic/signatures) are shown. (D) Profile of two copy number signatures extracted from 10 TCGA samples is shown. Abbr.: SBS, single base substitution; DBS, doublet base substitution; ID, insertion and deletion; CN, copy number.
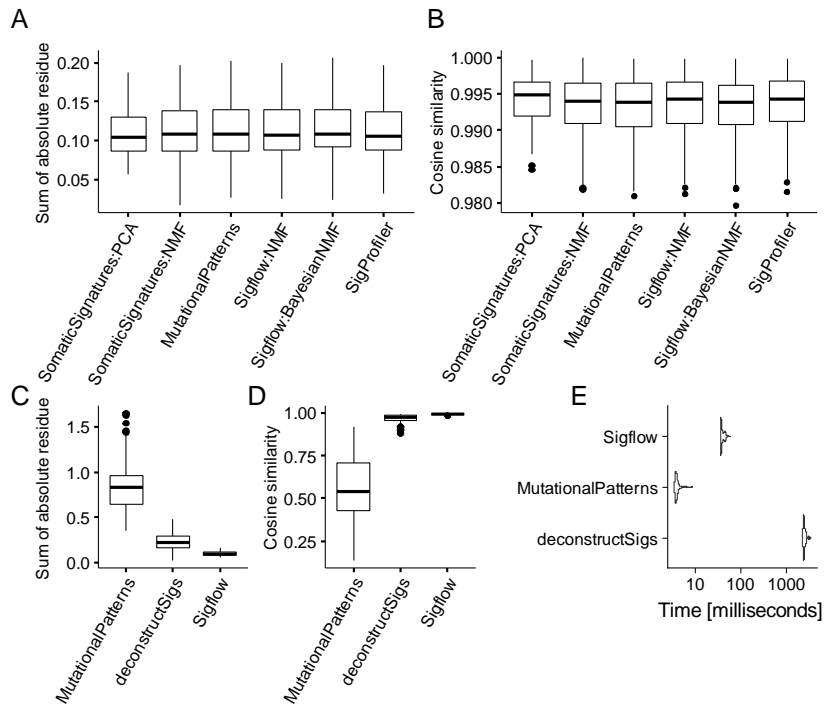
Fig. S3. Performance comparison in SBS signature extraction and fitting between Sigflow and other commonly used approaches with 214 PCAWG breast tumors. Distribution of error (A) and cosine similarity (B) between observed and reconstructed mutation spectrum in *de novo* signature extraction. Distribution of error (C) and cosine similarity (D) between observed and reconstructed mutation spectrum in reference signature fitting. Reference signatures are SBS signatures from COSMIC signature v3 database (https://cancer.sanger.ac.uk/cosmic/signatures). (E) Distribution of fitting time for single sample. The computation was repeated by 100 times. Here, Sigflow takes extra time to print the detailed information about program progress, so the speed is slower than MutationalPatterns.

Figure S4

A

| Run	Signature number	Posterior (log(e) transformed)
20	13	-13481.43071975890
14	13	-13482.900913046700
3	14	-13879.256323790000
8	14	-13980.186397149100
18	15	-14208.742754524600
12	15	-14212.198381532200
13	15	-14218.327880046800
9	15	-14221.33654585160
10	15	-14221.74220450850
2	15	-14229.04225343580
11	15	-14232.707494554600
5	15	-14235.023136448400
6	15	-14238.73360582180
1	15	-14259.25965695080
7	15	-14334.513048630400
15	16	-14603.115175310500
19	16	-14712.99807645120
4	16	-14732.01230457900
17	17	-15098.230327010300
16	17	-15122.315211541100

B
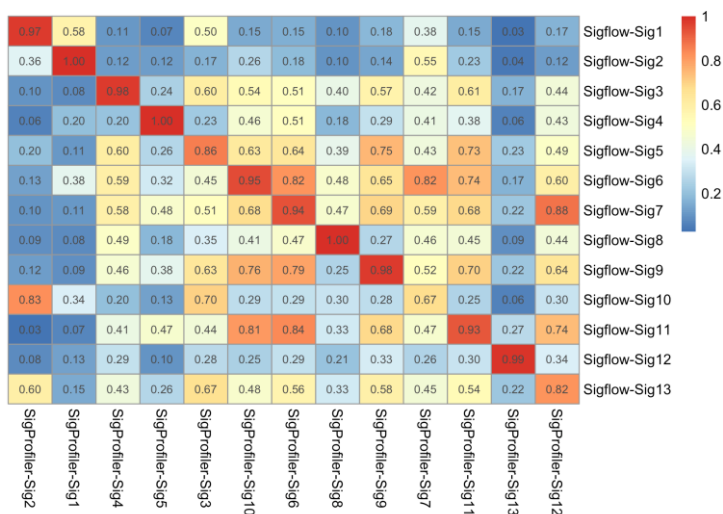
Fig. S4. Sigflow Bayesian NMF approach automatically extract 13 signatures from 214 PCAWG breast tumors with 20 runs. (A) Bayesian NMF run summary. The No. 20 run is selected as the optimal solution due to its maximum posterior probability. (B) The cosine similarity analysis between the 13 signatures extracted by Sigflow and the 13 signatures extracted by SigProfiler.

A

B

SP117933

Fig. S5. Signature stability evaluation through bootstrapping analysis using SBS mutations of 214 PCAWG breast cancer samples. (A) Comparison of signature exposure stability of SBS signatures measured as root mean squared error (RMSE) between exposures in 1000 bootstrapping samples and exposures in the original samples for each tumor. The metric RMSE indicates how many mutation differences for a signature may be introduced by resampling. (B) Distribution of bootstrapping absolute signature exposure of SBS signatures for SP117933, a selected tumor sample (SBS mutation n=1203) from the 214 PCAWG breast cohort.