## Real datasets

### Benchmarking cancer datasets

All cancer (colon, glioblastoma, kidney and lung) datasets used for the benchmarking analyses were obtained from <http://compbio.cs.toronto.edu/SNF/SNF/Software.html> (Wang *et al.*1). For the mRNA datasets, all transcripts with the same gene symbol were averaged.

### Breast cancer multi-omics study

*Datasets accession:* The level 3 TCGA data (version 2015\_11\_01) were retrieved from firebrowse.org hosted by the Broad Institute. The clinical data file (Merge\_Clinical) was downloaded from the Primary tab of the BRCA Clinical Archives. The mRNA RSEM normalized dataset (illuminahiseq\_rnaseqv2-RSEM\_genes\_normalized) was downloaded from the Primary tab of the BRCA mRNASeq Archives. The miRNA datasets (illuminahiseq\_mirnaseq-miR\_gene\_expression and illuminaga\_mirnaseq-miR\_gene\_expression) were downloaded from the Primary tab of the BRCA miRSeq Archives. The reverse phase protein array dataset (mda\_rppa\_core-protein\_normalization) was downloaded from the Primary tab of the BRCA RPPA Archives. The beta values for the methylation datasets (humanmethylation27-within\_bioassay\_data\_set\_function and humanmethylation450-within\_bioassay\_data\_set\_function MD5) were downloaded from the Primary tab of the BRCA Methylation Archives.

*Data processing:*Clinical data were present for 1,098 subjects for 3,703 variables. 29 unannotated transcripts were removed from the mRNA dataset composed resulting in 20,502 genes x 1212 samples. Two transcripts corresponded to *SLC35E2*, therefore one of the transcripts was re-labelled *SLC35E2.rep*. The miRNA datasets (1,046 miRNA x 1190 samples) was derived using two different Illumina technologies, the Illumina Genome Analyzer (341 samples) and the Illumina HiSeq (849 samples). The read counts instead of the reads\_per\_million\_miRNA\_mapped were used. The proteomics dataset obtained using a reverse phase protein array consisted of 142 proteins for 410 samples. The methylation data was derived from two different platforms, the Illumina Methylation 27 (27,578 CpG probes x 343 subjects) and the Illumina 450K (485,577 CpG probes x 885 subjects). There were 25,978 CpG probes in common between the platforms. The PAM50 labels for 1,182 samples were obtained from the TCGA staff.All datasets were restricted to samples coming from the primary solid tumor (sample type code 01) and to the first vial (vial code A).

*Normalization and pre-filtering:*The count data for the mRNA dataset, Xcounts was normalized to log2-counts per million (logCPM), Xnorm, similar to limma voom 25:

**

After library size (lib.size = total number of reads per sample) normalization, genes with counts less than 0 in more than 70% of samples were removed. The PAM50 genes were also removed from the mRNA dataset prior to analyses. Similarly, the miRNA count data was normalized to logCPM and miRNA transcripts with counts less than 0 in more than 70% of the samples were also removed.

### Asthma multi-omics study

*Datasets accession:* Paired blood samples were obtained from 14 asthmatic individuals undergoing allergen inhalation challenge as previously described 26. Cell counts were obtained from a hematolyzer (percentage of Neutrophils, Lymphocytes, Monocytes, Eosinophils and Basophils) and DNA methylation analysis (percentage of T regulatory cells, T cells, B cells and Th17 cells). Gene expression profiling was performed using Affymetrix Human Gene 1.0 ST (GSE40240). Metabolite profiling was performed by Metabolon Inc. (Durham, North Carolina, USA). All asthma data have been published as part of previous studies 27,28.

*Normalization:* Microarray data was normalized using Robust MultiArray Average (RMA), consisting of background correction, quantile normalization and probe summarization using median polish. Preprocessing of mass spectrometry data including data extraction, peak-identification and data preprocessing for quality control and compound identification was performed by Metabolon Inc. (Durham, North Carolina, USA).