A) TCGA data

	Mutation	Expression	Copy Number
Urothelial Bladder Carcinoma (BLCA)	130	211	131
Breast Carcinoma (BRCA)	507	994	823
Colorectal Carcinoma (COAD/READ)	224	413	276
Glioblastoma (GBM)	291	170	574
Chromophobe Renal Cell Carcinoma (KICH)	66	66	66
Clear Cell Renal Carcinoma (KIRC)	424	496	499
Acute Myeloid Leukemia (LAML)	195	123	200
Lung Adenocarcinoma (LUAD)	230	489	230
Lung Squamous Cell Carcinoma (LUSC)	134	482	178
Ovarian Carcinoma (OV)	316	419	557
Gastric Adenocarcinoma (STAD)	289	282	295
Papillary Thyroid Carcinoma (THCA)	401	500	510
Endometrial Carcinoma (UCEC)	248	489	373

Calculate genomic features for each human gene

B) Genomic profile of a gene

Mutation

- **mutation selection score**
- **HiFI / LoFI**
- **splice / LoFI**
- **LOF / LoFI**

- HiFI / silent
- LoFI / silent
- splice / silent
- LOF / silent

- HiFI / benign
- LoFI / benign
- splice / benign
- LOF / benign

- HiFI / total
- LoFI / total
- splice / total
- LOF / total

- (non-syn. mutations / N) / c^d
- (missense mutations / N) / c^d
- (HiFI mutations / N) / c^d
- (LoFI mutations / N) / c^d
- (splice mutations / N) / c^d
- (LOF mutations / N) / c^d
- (total mutations / N) / c^d

d = 1, 2, 3, 4

N : number of samples
c : coding sequence length

Expression

- differential expression score (based on p-value)
- differential expression score (based on fold change)

Copy Number

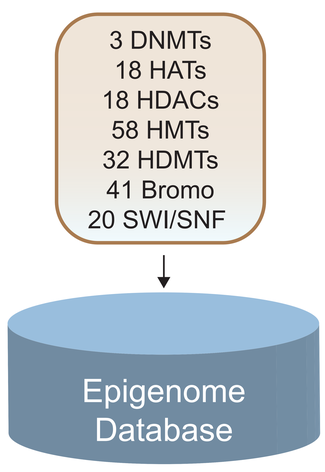
- Proportion of samples with copy number amplificiations
- Proportion of samples with copy number deletions

Features in red: selected for OG prediction
Features in blue: selected for TSG prediction

Definitions

Mutation selection score	=	reflects the occurence of mutation hot spots within one gene
HiFI/LoFI missense mutations	=	missense mutations with high/low functional impact on protein function
Benign mutations	=	silent + LoFI missense mutations
LOF (loss of function) mutations	=	nonsense + frameshift mutations

C) Prediction of cancer genes among ERGs



Train predictors using features of known cancer genes and apply predictors to all ERGs