decompTumor2Sig

R package for the identification of active mutational signatures in individual tumors

Sandra Krüger[(1)], *Rosario M. Piro*[(1,2,3)]

(1) Institute of Computer Science and Institute of Bioinformatics, Freie Universität Berlin, Germany
(2) Institute of Medical Genetics and Human Genetics, Charité-Universitätsmedizin Berlin, Germany
(3) German Cancer Consortium (DKTK) partner site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany

The mutational processes responsible for the somatic mutations observed in tumor samples can be modeled by "**mutational signatures**" [1-3] which reflect the occurrences of base changes within their sequence contexts (i.e., in dependence on their flanking bases). The age-related mutations initiated by spontaneous deamination of 5-methylcytosine, for example, regard C>T transitions in the context of CpGs (because those are methylated; see Figures).

The first signature model described by **Alexandrov et al** [1,2] separately counts all possible nucleotide triplets whose central base is mutated (total of 6*4*4 = 96 possible triplet mutations, i.e., parameters).

The second model by **Shiraishi et al** [3] describes probabilistic mutational signatures in analogy to the way we model transcription factor binding motifs, i.e., considering the single bases of the motif as independent (much fewer parameters).

We developed a user-friendly R package, called **decompTumor2Sig** [4], that can be used to evaluate the contributions ("exposures") of both Alexandrov- and Shiraishi-type signatures to the somatic mutations found in an individual tumor.

Alexandrov-type mutational signatures

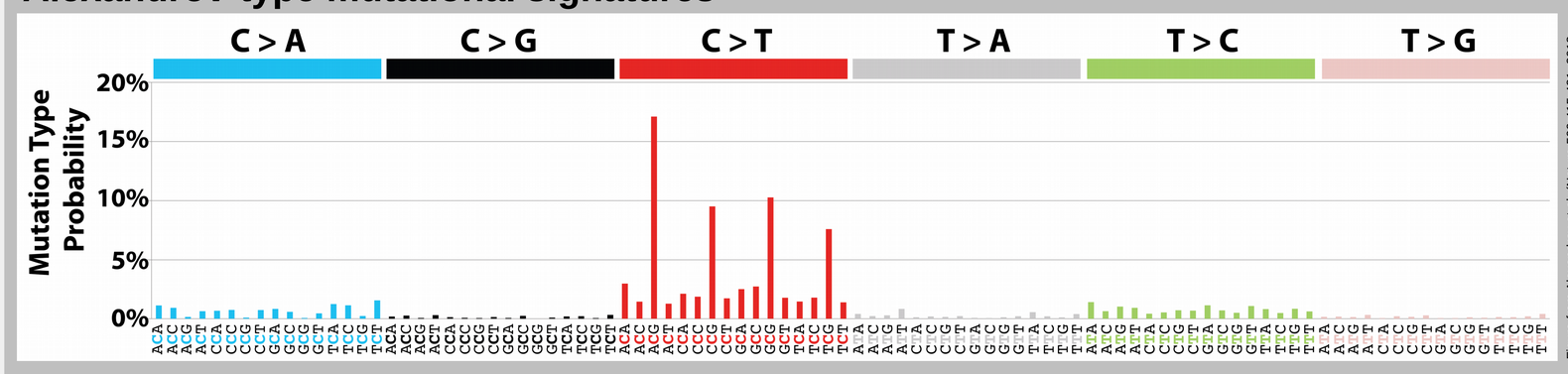
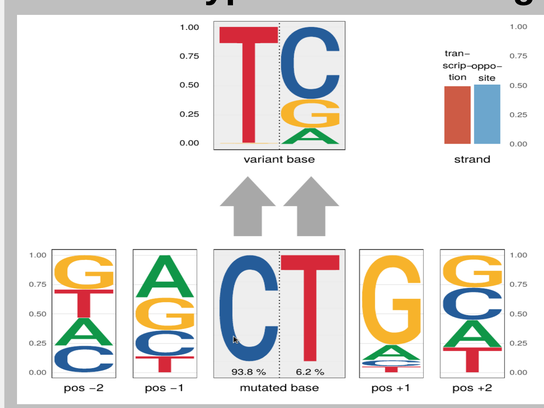


Figure from Alexandrov et al. (Nature 500:415-421, 2013; https://cancer.sanger.ac.uk/cosmic/signatures)

Shiraishi-type mutational signatures



Nucleotide change (central base)

C>A	C>G	C>T	T>A	T>C	T>G
0.004	0.006	0.928	0.009	0.038	0.015

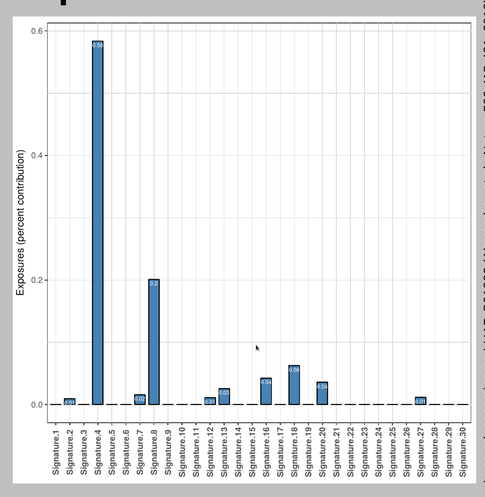
Flanking bases

Position	A	C	G	T
-2	0.237	0.228	0.293	0.242
-1	0.362	0.220	0.279	0.139
+1	0.131	0.053	0.764	0.052
+2	0.232	0.277	0.277	0.214

Transcription strand

plus strand	minus strand
0.493	0.507

Exposures / contributions



Lung adenocarcinoma LUAD-S01302 (Alexandrov et al., Nature 500:415-421, 2013)
Alexandrov / COSMIC signature 4 = smoking; signature 8 = unknown aetiology

Cross validation of predicted exposures with two datasets

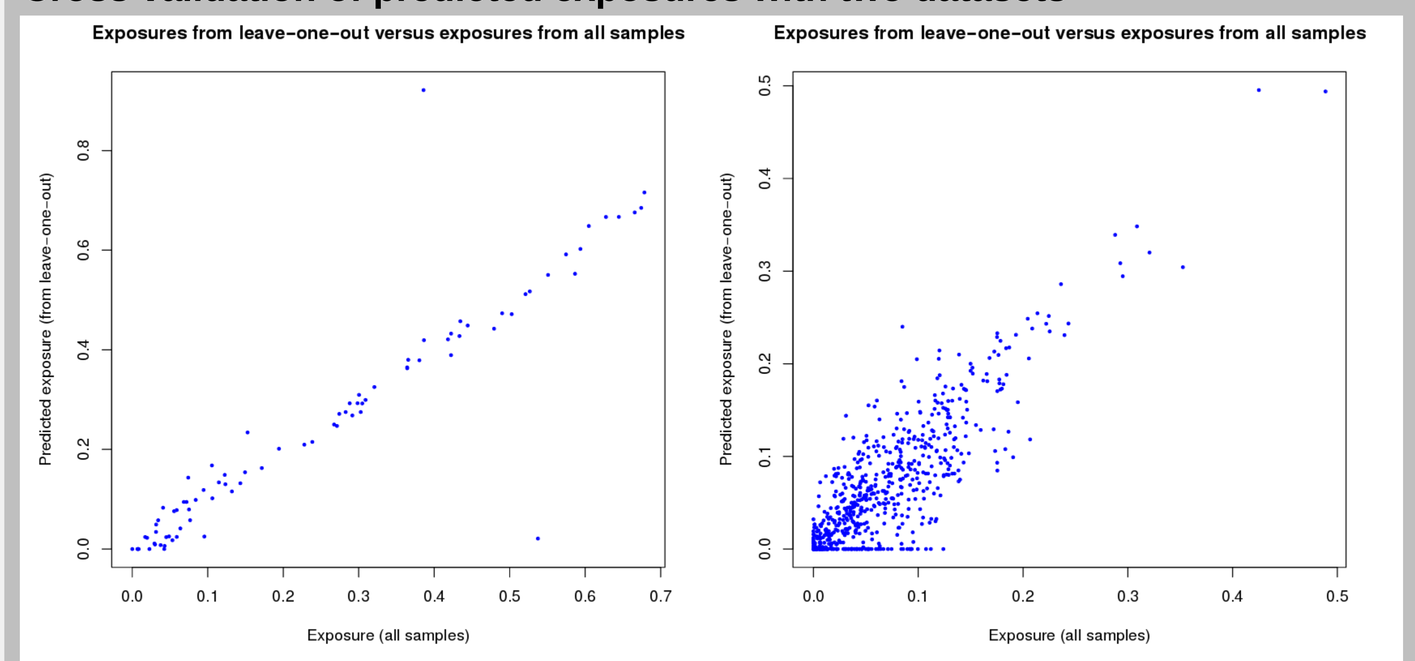


Figure 2: Comparison of contributions/weights ("exposures") predicted for individual tumors (`decompTumor2Sig`; y-axis) and collectively computed (`pmsignature`; x-axis). Left: leave-one-out test on 21 breast cancers ($r = 0.923$). Right: test set of 44 out of 435 tumors ($r = 0.807$).

Usage example

```
### decompTumor2Sig example for Shiraishi signatures ###

library(decompTumor2Sig)

# load signatures (or convert from pmsignature)
signatures <- loadShiraishiSignatures(sigfiles)

# load reference genome and transcript annotation
rG <- BSgenome.Hsapiens.UCSC.hg19:: [...]
tA <- TxDb.Hsapiens.UCSC.hg19.knownGene:: [...]

# read mutations of the individual tumor(s)
genomes <- loadGenomesFromVCF("tumor.vcf", numBases=5,
                      type="Shiraishi", trDir=TRUE,
                      refGenome=rG, transcriptAnno=tA)

# predict exposures for tumor(s)
exposures <- decomposeTumorGenomes(genomes, signatures)

exposures
#$PD4072a
#[1] 0.41464965 0.52231228 0.06303806 0.00000000
```

Mathematical approach to compute exposures (=weights, vector *w*) to signatures (matrix ***S***) for tumor genome (vector *g*):

The approach is based on classical quadratic programming because we can solve the following problem:

$$\text{minimize} \quad (g - \mathbf{S}w)^T(g - \mathbf{S}w) = g^T g - g^T\mathbf{S}w - (\mathbf{S}w)^T g + (\mathbf{S}w)^T\mathbf{S}w$$
$$\text{subject to} \sum_{s=1}^{K} w_s = 1, w_s \geq 0$$

or simplified:

$$\text{minimize} \quad -g^T\mathbf{S}w + \tfrac{1}{2}w^T\mathbf{S}^T\mathbf{S}w \qquad \text{subject to} \quad \sum_{s=1}^{K} w_s = 1, w_s \geq 0$$

[1] Alexandrov et al. (2013) Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3**(1):246-259.
[2] Alexandrov et al. (2013) Signatures of mutational processes in human cancer. *Nature* **500**:415-421.
[3] Shiraishi et al. (2015) A simple model-based approach to inferring and visualizing cancer mutation signatures. *PLoS Genet* **11**(12):1005657.
[4] Krüger, Piro (2018) decompTumor2Sig: Identification of mutational signatures active in individual tumors. *BMC Bioinformatics* (accepted).