ID 01.3

Estimating the mutation risks conferred by mutational processes

A.K.M. Rosendahl Huber[1]*, F. Muiños[1,2], A. González-Pérez[1,2,3], & N. López Bigas[1,2,3]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain
[2]Centro de Investigación Biomédica en Red en Cáncer (CIBERONC), Instituto de Salud Carlos III, Madrid, Spain
[3]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
** axel.rosendahl@irbbarcelona.org*

DNA mutations can disrupt vital cellular functions leading to disase. One of the most striking examples is cancer, which requires mutations disturbing the activity of cancer driver genes. The risk at which these genes are mutated is not equal across mutational processes, as certain positions in the genome have been shown to exhibit considerable variation in mutation rates. This variation in mutation risk is the result of chromatin and sequence differences at multiple scales, the mode of action of the process, and repair of DNA damage. Thus, mutagenic processes may confer varying risks to induce cancer driver mutations. Accurate determination of these mutation risks is important to understand the induction of cancer driver mutations.

We aim to compute the mutation risks for mutational processes across the genome using computational modeling approaches. Currently, we have explored these mutation risks using different modeling strategies, using linear regression and decision tree based machine learning methods such as gradient boosting. As input for modeling, we use mutation probabilities from 33 mutational processes, defined as mutational signatures. Mutations are distributed in genomic bins ranging from 1mb to exonic regions of genes. As determinants for the mutation rates we use the mean scores of epigenetic, replication timing and transcription factor binding site data, which have been known to correlate with absolute mutation counts. Currently, we are able to predict genomic mutation risks accurately for most mutagenic processes (R2 values ranging between 0.6 and 0.8), with tree-based regression performing better at predicting mutational loads. Decomposing the contribution of covariates enables the identification of the determinants behind the variance in mutation rates. The currently generated models provide the basis to calculate relative risks scores for mutagenic processes for specific mutations across the genome. In the future, this will enable the quantification of risk scores of various mutagenic processes to induce cancer driver mutations and neoantigens.

Keywords:
Mutational Signatures; Statistical Learning; Risk Estimation; Carcinogenesis; keywords.