Glossary

APOBEC apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like

ASCAT allele-specific copy number analysis of tumors

BAF B-allele frequency

CAF Carcinoma-associated fibroblasts

CCF Cancer cell fraction - The fraction of tumour cells that carry a particular mutation, or the fraction of tumour cells in the sequencing sample that is represented by a mutation cluster

CGH Comparative genomic hybridization

Clonal Is used to refer to mutations carried by all tumour cells. A cluster of these mutations is referred to as clone

CNA Copy number alteration - Somatic copy number change

CNLOH Copy neutral loss of heterozygosity

COSMIC Catalogue of somatic mutations in cancer

CP Cellular proportion/prevalence - The proportion of cells in the sequencing sample that a mutation is carried by or is represented by a mutation cluster. CP is CCF multiplied by the tumour purity: CP=CCF*ρ

ctDNA Cell free tumour DNA

dbSNP Single Nucleotide Polymorphism Database

DNA Deoxyribonucleic acid

Driver mutation A somatic mutation that is thought to convey any selective advantage

ECM Extra cellular matrix

EMD Earth movers distance

EMT Epithelial-mesenchymal transition

FDR False discovery rate

GC content The proportion of bases within a range of the types C or G

Haplotype phasing The process of estimating which SNP alleles appear on the same chromosome within an individual organism

Haplotype A set of SNPs on the same chromosome

hg19 Human reference genome build 19

HMM Hidden Markov model

ICGC International Cancer Genome Consortium

IGH locus Immunoglobulin heavy locus

Indel Somatic short insertion or deletion

Infinite sites assumption The assumption that mutations occur only once during the life time of the tumour

ITH Intra-tumour heterogeneity

K-S Kolmogorov-Smirnov test

Kataegis Localised somatic hypermutation

logR Quantification of the amount of DNA available for a certain locus. In sequencing data, logR represents the ratio of the coverage of the tumour over that in the matched normal

LOH Loss of heterozygosity

MCMC Markov chain Monte Carlo

MPEAR Maximal posterior expected Rand index

Multiplicity The number of chromosome copies that carry a somatic mutation

Passenger mutation A somatic mutation that is thought to not convey any selective advantage

PCAWG Pan-Cancer Analysis of Whole Genomes

PCF Piecewise constant fitting

Ploidy The average number of chromosome copies of a cell

Purity The proportion of tumour cells available in the data

QC Quality control

RMSE Root mean squared error

RNA Ribonucleic acid

SMC-het Somatic Mutation Calling heterogeneity - Challenge to benchmark subclonal reconstruction methods

SNP Single nucleotide polymorphism - A germline single base difference from the reference genome

SNV Single nucleotide variant - A somatic single base substitution

Subclonal reconstruction Estimation of the number of subclonal cell populations within a tumour sequencing sample, the number of mutations in each population and the size of each population (fraction of tumour cells)

Subclonal Is used to refer to mutations carried by a subset of tumour cells. A cluster of these mutations is referred to as a subclone

SV Somatic structural variant

TAM Tumour-associated macrophages

TCGA The Cancer Genome Project

TIL Tumour-infiltrating lymphocytes

TME Tumour micro-environment

TSG Tumour suppressor gene

UV Ultraviolet

VAF Variant allele frequency

WABCS West African Breast Cancer Study

WGS Whole genome sequencing

WXS Whole exome sequencing