Whole Genome Sequencing Analysis for Oncology

GeCo
Advanced Genomic Consulting Service

Background

Whole genome sequencing (WGS) provides access to the complete spectrum of driver events in cancer. This includes germline or somatic mutations, coding or non-coding mutations, indels, copy-number alterations, structural rearrangements, and viral insertions. Studying the nature of mutations also allows the exploration of mutational processes occurring within a tumor, particularly genomic instability phenotypes that can be highly relevant to clinical care. Additionally, the clonality distribution of mutations and copy-number alterations enables reconstruction of the clonal architecture of the tumor and the ability to distinguish early clonal from late subclonal mutations.

GeCo's experts have developed state-of-the art modules based on standard pipelines from the International Cancer Genome Consortium's (ICGC) Pan Cancer Analysis of Whole Genomes (PCAWG).[1] The available modules for WGS data analysis are detailed below. Deliverables include high-quality figures and tables as well as a detailed material and methods to support rapid publication of the results.

Module 1 - Global Description of Mutational Burden

This module provides information on the burden of mutations, copy-number alterations (CNAs) and structural variants (SVs) for the sample cohort. Optional data includes the correlation of the above with patient clinical features and comparisons between tumor subgroups. Outputs include tables and figures (barplot graphs) representing the burden of somatic alterations across the cohort, with optional comparison between clinical features and tumor subgroups.

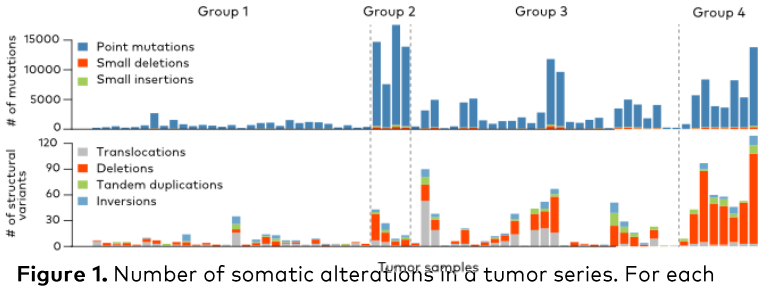


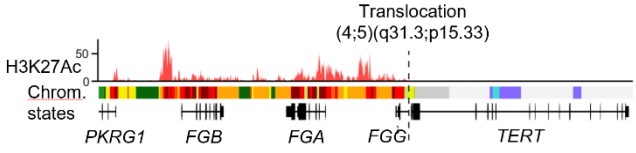
Figure 1. Number of somatic alterations in a tumor series. For each tumor, the number of mutations (top) and structural variants (bottom) is represented, with the type of event represented by a color code. Samples are grouped by subgroups showing different mutation and SV burden.

Module 2 - Integrated Characterization of Driver Events

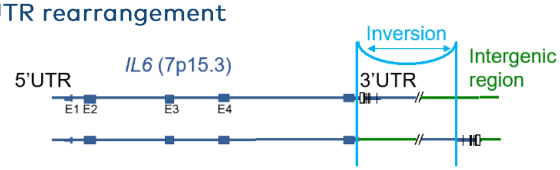
The second module provides an extensive catalog of driver events including mutations and indels affecting cancer genes (coding and non-coding), CNAs (high-level amplifications, homozygous deletions, focal and broad chromosome aberrations) and SVs generating oncogenic gene fusions. The module can highlight original oncogenic mechanisms involving regulatory sequences (e.g., enhancer hijacking or disruption or untranslated region [UTR] sequences),

Deliverables include an OncoPrint representation integrating all driver alterations across the cohort, and the classification of samples based on their driver alterations. Optional data that can be provided includes the comparison of driver frequency between tumor subgroups based on clinical or other molecular features.

a) Enhancer hijacking



b) UTR rearrangement



c) Visualization of driver alterations (OncoPrint)

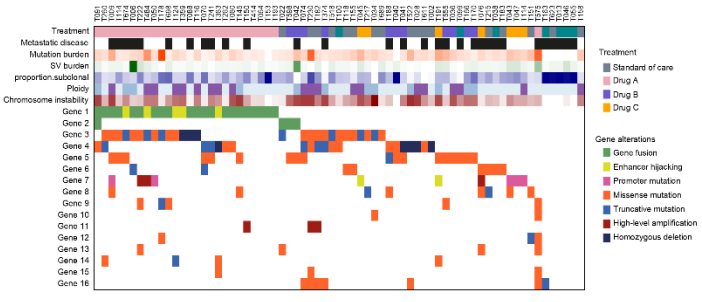


Figure 2. Integrated characterization of driver events: (a) identification of an enhancer hijacking event where a translocation puts TERT oncogene downstream potent enhancers of fibrinogen genes in a hepatocellular carcinoma (adapted from Bayard, et al.[2]); (b) chromosome inversion leading to the loss of regulatory 3' UTR motifs and to the massive overexpression of IL6 in a hepatocellular adenoma (adapted from Calderaro, et al.[3]); (c) OncoPrint representing exhaustive driver events in a tumor series, including coding and non-coding mutations, copy number changes and structural variants, along with clinical and molecular annotations.

Module 3 - Mutational Signatures and Genomic Instability

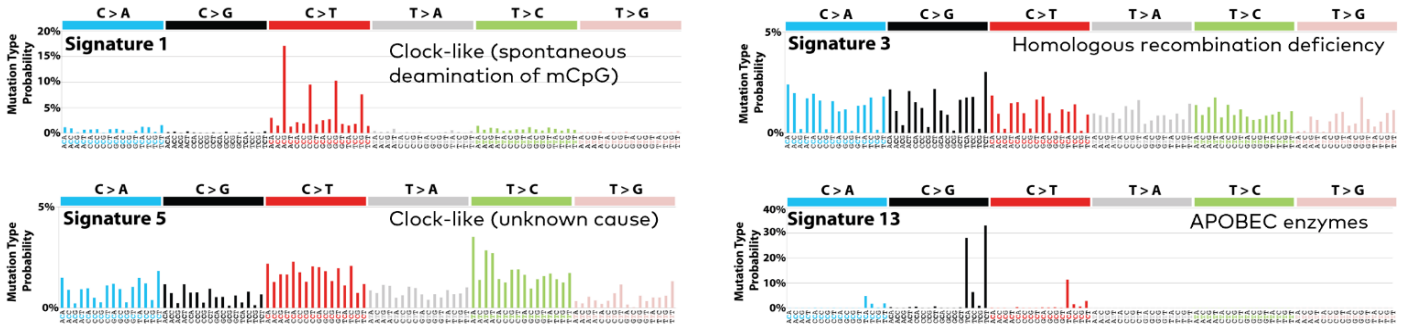
Mutational signature analysis allows exploration of mutational processes occuring in each tumor.[4] This module includes discovery of new mutational signatures in the cohort and characterization of established signatures associated with endogenous mutational processes, carcinogenic exposures and DNA repair defects. The latter step utilzes COSMIC[5] as a reference for known signatures operative in the relevant tumor type.

After quantifying the contribution of each signature to each tumor genome, driver mutations are associated to their most likely causal signatures.

Additionally, tumors are classified based on mutational signature contributions to highlight associations with clinico-molecular alterations.

Genomic instability phenotypes are identified using COSMIC signatures related to specific DNA repair defects (e.g. homologous recombination or DNA mismatch repair) and optionally using dedicated tools like HRDetect or MSIsensor. Other specific focus may include treatment-induced signatures, as well as doublet base substitution and indel signatures in addition to single base substitution signatures.

a) Mutational signatures operative in a tumor genome data set



b) Classification of the tumors based on mutational signature contribution

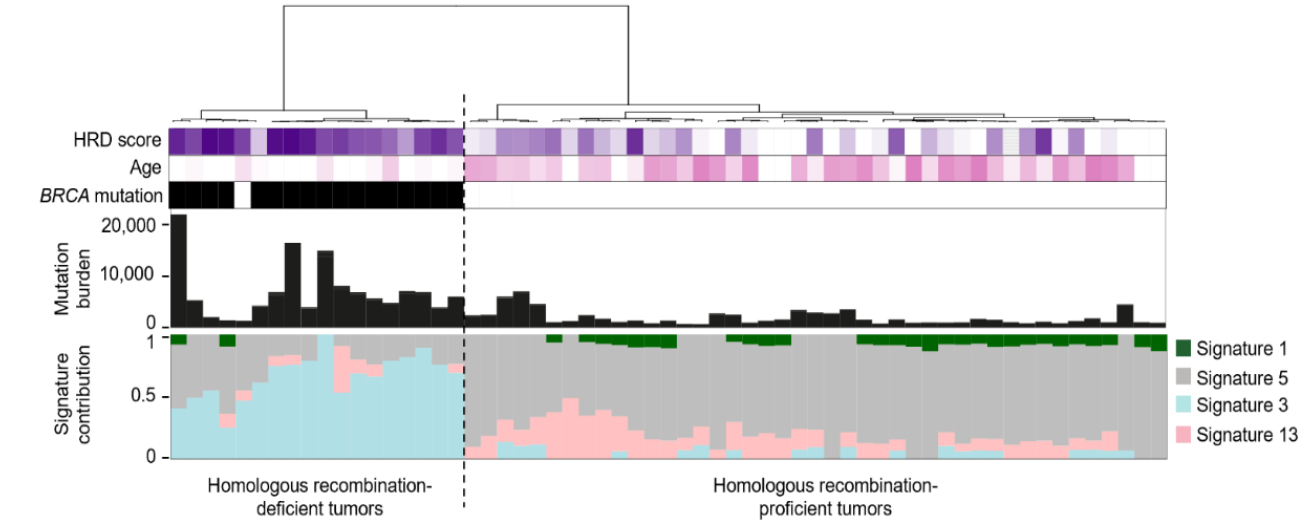
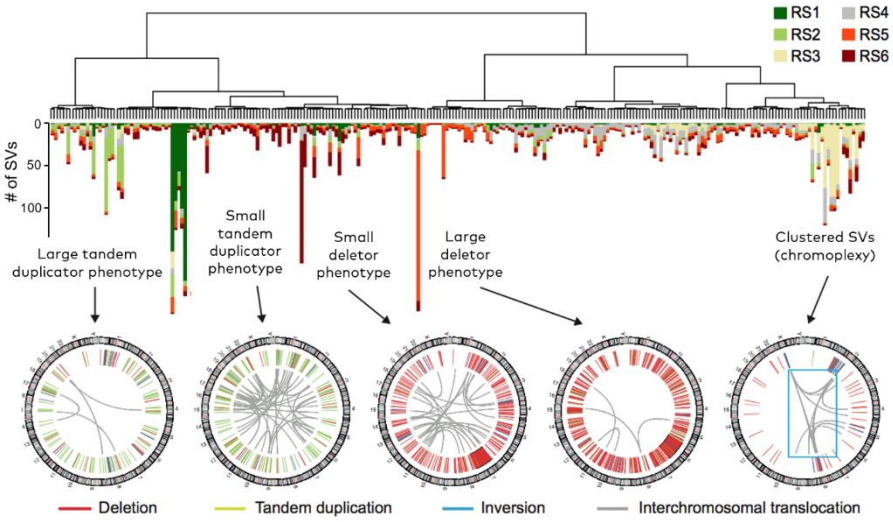


Figure 3. Mutational signature analysis: (a) mutational signatures identified in a tumor genome data set. Each signature displays a specific pattern of substitution types occurring in specific trinucleotide context. The molecular etiology of each signature is indicated; (b) classification of tumor samples based on mutational signatures reveals a cluster of BRCA-mutated tumors with a prevalence of signature 3 related to homologous recombination deficiency.

Module 4 - Complex Structural Variant Phenotypes (chromotripsis, chromoplexy and more)

Structural rearrangements, including catastrophic chromotripsis or chromoplexy events, account for a substantial proportion of driver events, in particular in childhood cancers.[6] In-depth characterization of complex SVs is key to obtaining the full spectrum of driver events and unravel the causal genomic instability mechanisms.[7] This module identifies complex clustered events like chromotripsis or chromoplexy, as well as SV signatures characteristic of new and known SV phenotypes (e.g., related to BRCA1, BRCA2 or CDK12 inactivation). The module also classifies the data set based on SV signatures in order to highlight tumor subgroups sharing common instability phenotypes and their association with clinico-molecular annotations.

a) Structural variant signature analysis



b) Visualization of a chromoplexy event

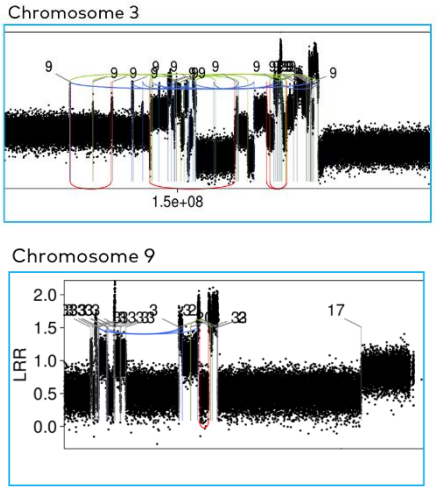


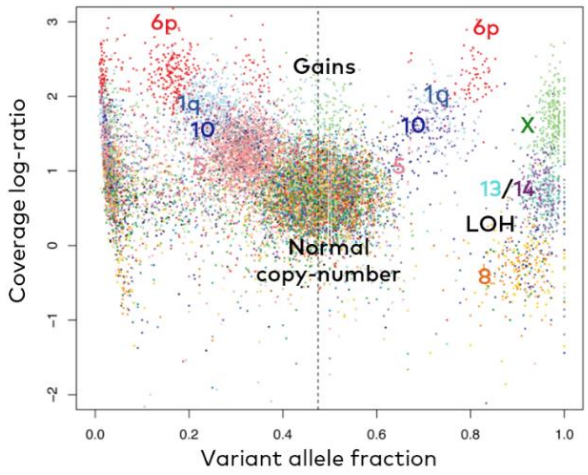
Figure 4. Complex structural variants and signatures: (a) classification of tumor samples based on their SV signatures reveals subgroups with extreme SV phenotypes. The dendogram shows the classification with the number of SVs attributed to each signature (RS1 to RS6) below. CIRCOS plots illustrate typical examples of tumors with each phenotype; (b) Detailed visualization of a chromoplexy event involving two chromosomes interconnected by multiple translocation breakpoints.

Module 5 – Clonal Architecture

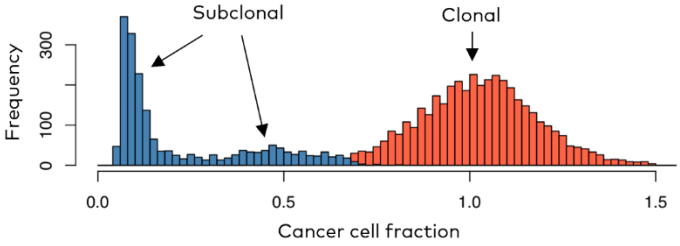
The clonality of mutations allows the ability to distinguish early clonal from late subclonal events and to follow the evolution of driver genes and mutational signatures along tumorigenesis. This module estimates the proportion of tumor cells carrying each mutation (e.g., the cancer cell fraction, CCF), based on the variant allele fraction (VAF), corrected for tumor purity and absolute copy-number. The module identifies subclones based on CCF distribution, classifies each mutation as clonal or subclonal, and reveals the evolution of driver genes and mutational signatures during tumor progression.

This module identifies complex clustered events like chromotripsis or Deliverables include the mutation table annotated with cancer cell fraction (CCF) and its confidence interval, figures representing the distribution of CCF in each tumor with assignment of clonal and subclonal mutations and the prevalence of driver mutations and mutational signatures between early clonal and late subclonal events.

a) Variant allele fraction distribution



b) Cancer cell fraction distribution



c) Evolution of driver genes and signatures

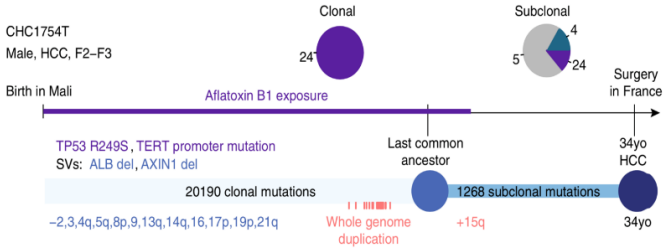


Figure 5. Clonal architecture analysis: (a) Variant allele fraction (VAF) distribution in one tumor sample. Each dot represents a mutation (colored by chromosome) with its VAF on the x axis and coverage log-ratio on the y axis. The VAF of each mutation depends on tumor purity and local copy-number; (b) Cancer cell fraction (CCF) obtained by correcting VAF for tumor purity and copy-number. The distribution clearly shows a peak of clonal mutations (in orange) and two peaks of subclonal mutations (in blue); (c) Evolution of driver genes and mutational signatures (pie charts) between early clonal and late subclonal mutations in a liver cancer patient born in Africa and migrated to France. Signature evolution shows a disappearance of aflatoxin B1 exposure in late tumor progression. Adapted from Letouzé et al.[6]

Module 6 – Molecular Evolution

When several samples from the same patient are analyzed (i.e., primary tumor and relapse), the molecular evolution of the tumor can be reconstructed in terms of driver events and mutational processes. This module calculates the VAF and CCF for each mutation detected in at least one sample for all of the patient samples, allowing the ability to detect mutations present at a very low VAF that may be missed by variant callers. The module then identifies shared and private variants, as well as clonal architecture evolution based on the CCF distribution across samples.

A tumor oncogenetic tree is reconstructed for each patient, annotated with driver events and mutational processes occurring at each step of tumor progression. Oncogenetic trees for several patients may then be integrated to derive general rules regarding the timing of driver events and mutational processes.

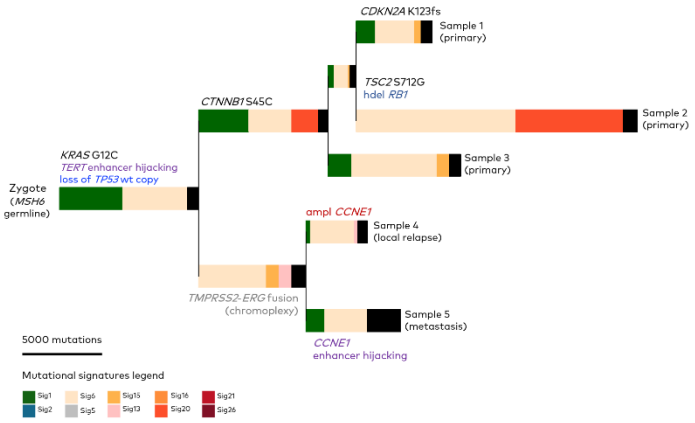


Figure 6. Molecular evolution reconstructed by the analysis of several tumor samples from the same patient. Driver mutations and SVs are indicated on each branch. The length of each branch indicates the number of mutations acquired, with a color code representing the contribution of each mutational signature.

Module 7 – Integration with Matched RNA-Seq Data

Integrating WGS with RNA-seq data enables to refine the characterization of driver events. Examples of integrated analyses available with this module include the identification of regulatory mutations or SVs associated with the overexpression of a neighbor gene (e.g., enhancer hijacking or TAD disruption), or validation of altered transcript structures (i.e. fusion genes) predicted from genomic SVs.

Deliverables include figures and tables reporting somatic alterations associated with the overexpression of neighbor genes, and annotations of SVs with fusion genes identified via RNA-Seq.

Examples of studies using GeCo modules for WGS in oncology:

- Letouzé E, Shinde J, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun.* 2017;8(1):1315.
- Bayard Q, Meunier L, Peneau C, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun.* 2018;9(1):5235.
- Calderaro J, Letouzé E, Bayard Q, et al. Systemic AA Amyloidosis Caused by Inflammatory Hepatocellular Adenoma. *N Engl J Med.* 2018;379(12):1178-1180.

References:

1. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature.* 2020. PMID: 32025007
2. Bayard Q, Meunier L, Peneau C, et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun.* 2018;9(1):5235.
3. Calderaro J, Letouzé E, Bayard Q, et al. Systemic AA Amyloidosis Caused by Inflammatory Hepatocellular Adenoma. *N Engl J Med.* 2018;379(12):1178-1180.
4. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature.* 2020;578(7793):94-101.
5. Tate JG, Bamford S, Jubb HC, et al. COSMIC: The Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 2019;47(D1):D941-D947.
6. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature.* 2018;555(7696):371-376.
7. Li Y, Roberts ND, Wala JA, et al. Patterns of somatic structural variation in human cancer genomes. *Nature.* 2020;578(7793):112-121.
8. Letouzé E, Shinde J, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat Commun.* 2017;8(1):1315.

GeCo Advanced Genomic Consulting Service

A service brought to you by: INTEGRAGEN

To learn more about GeCo or to contact us about your project, email us at services@integragen.com