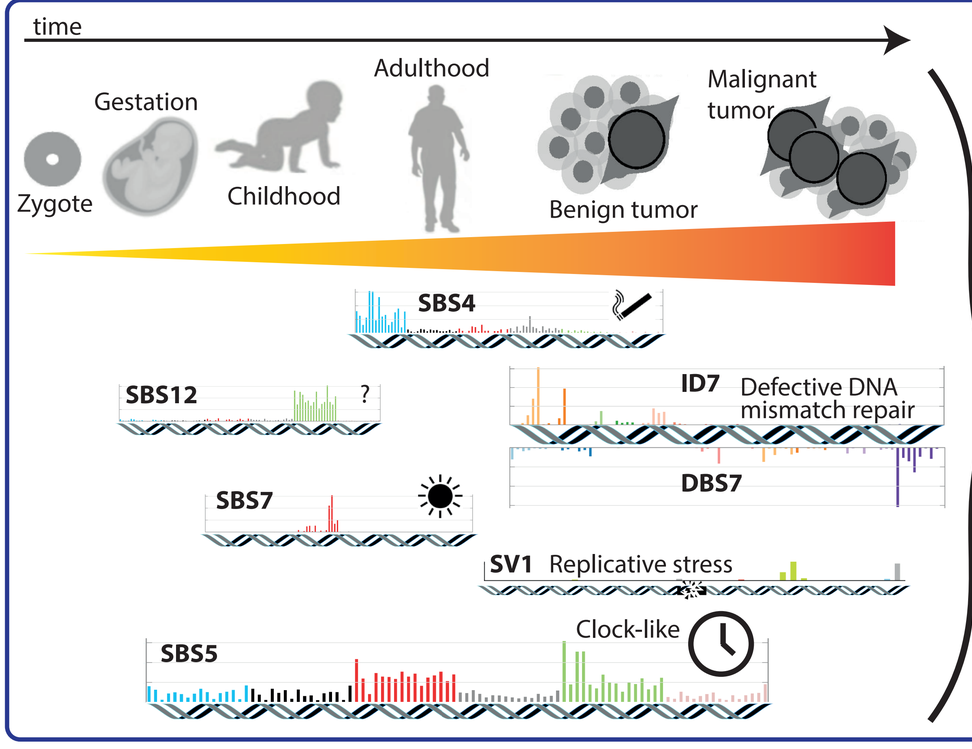
Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer

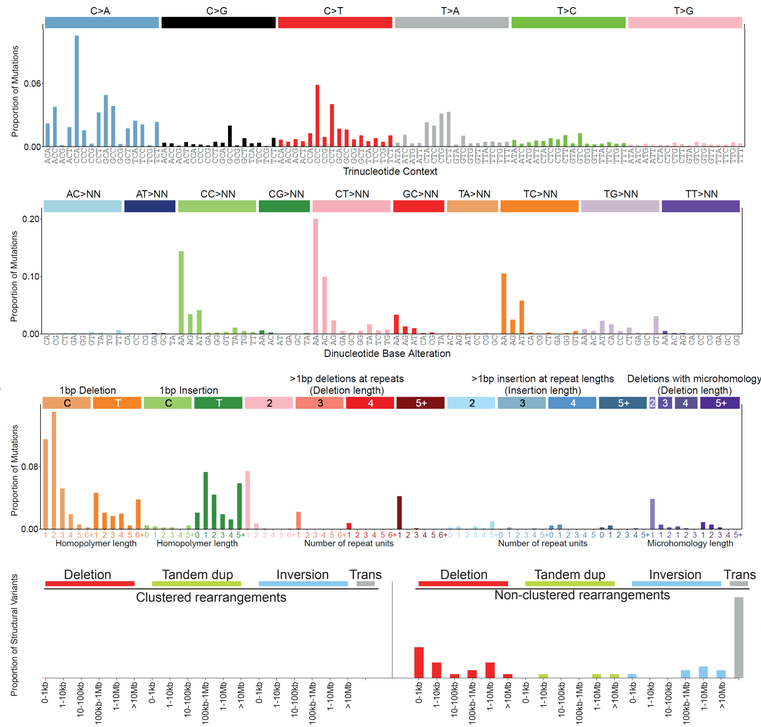
Theo Z HIRSCH[1, 2], Jayendra SHINDE[1, 2], Benedict MONTEIRO[1, 2], Quentin BAYARD[1, 2], Sandrine IMBEAUD[1, 2], Feng LIU[3], Victor RENAULT[3], Jessica ZUCMAN-ROSSI[1, 2] and Eric LETOUZÉ[1, 2]

1 Centre de Recherche des Cordeliers, Functional Genomics of Solid Tumors laboratory, Sorbonne Université, Inserm, USPC, Université Paris Descartes, Université Paris Diderot, Paris, France
2 Labex OncoImmunology, Equipe labellisée Ligue Contre le Cancer, Centre de Recherche des Cordeliers, Paris, France 3 Laboratory for Bioinformatics, Fondation Jean Dausset – CEPH, Paris F-75010, France

Cancer genomes are altered by various **mutational processes** and, like palimpsests, bear the signatures of these different processes.
Mutational signature analysis is a powerful approach to **decipher the origin of somatic mutations in cancer** [1].
The **Palimpsest R package** [2] allows to **integrate mutational signature and clonality analyses** in order to reconstruct the **natural history of a tumor**.

Cancer genomes bear the signatures of different mutational processes





Single base substitution
SBS

Double base substitution
DBS

Small insertions and deletions
ID

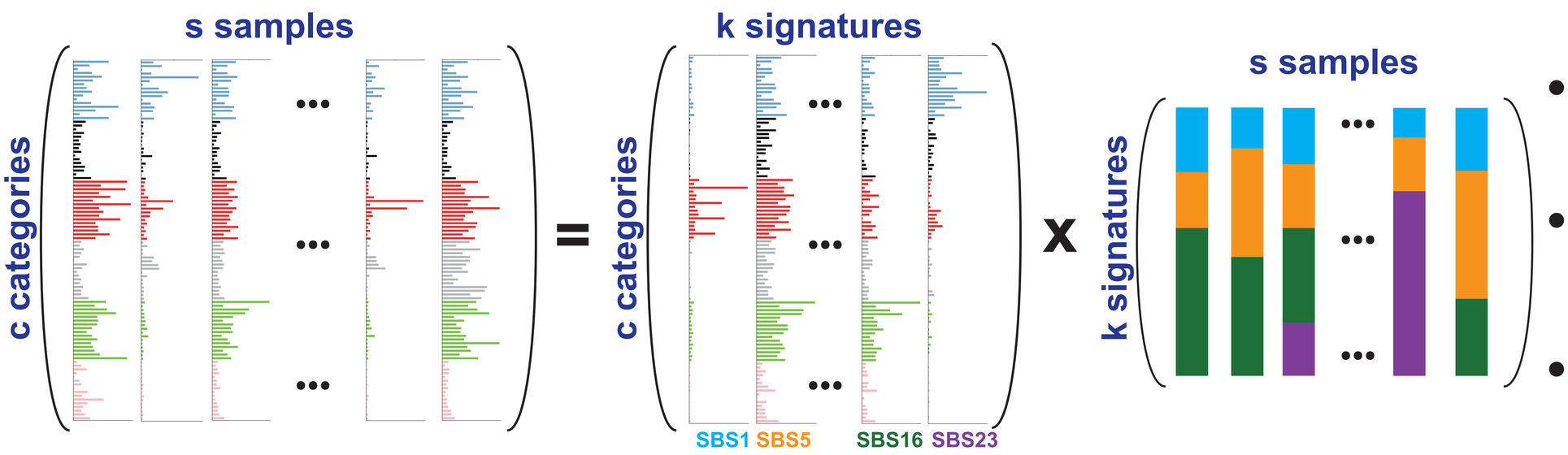
Structural variants
SV

Somatic alterations are **classified into categories** [3,4] according to:
- **Type** (T>A, Inversion, Deletion ...)
- **Context** (trinucleotide, repeats ...)
- **Size** (for ID and RS)
- **...**

Identification of **recurrent patterns** reveals the **signatures of each mutational processes**

These can occur **from the zygote through adulthood**, and also **throughout the lifespan of the tumor**

Palimpsest package: extracting signatures of known and new mutational processes



Mutation catalog = Processes × Exposures

- Blind source separation with **NMF** (non-negative matrix factorization)
- Can be applied to any type of somatic alteration (**SBS**, **DBS**, **ID** or **SV**)
- *De novo* extracted signatures can be **compared to Cosmic** signatures with cosine similarities

Processes
- ***De novo* extraction of novel signatures**
- **Already known signatures (Cosmic database)**

NMF: finding the solution to

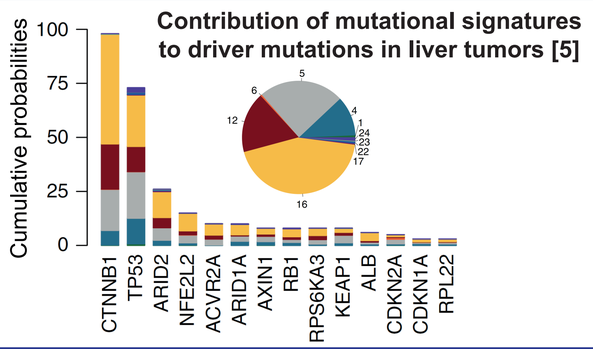
$$\min_{P \in M_{R^+}^{(c,k)}, E \in M_{R^+}^{(k,s)}} \|M - P \times E\|_F$$

Assigning the causal process of each alteration

For each variant in each sample:

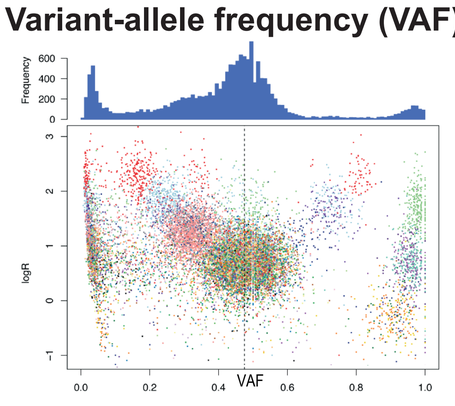
- probability that it was due to each identified signatures
- assign the most likely signature

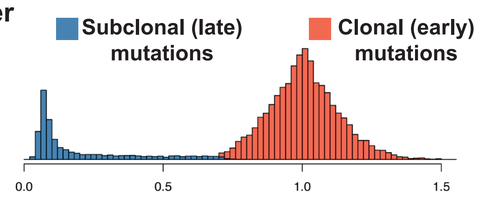
Drivers ⟷ Etiologies



Clonal architecture of tumors - single sample

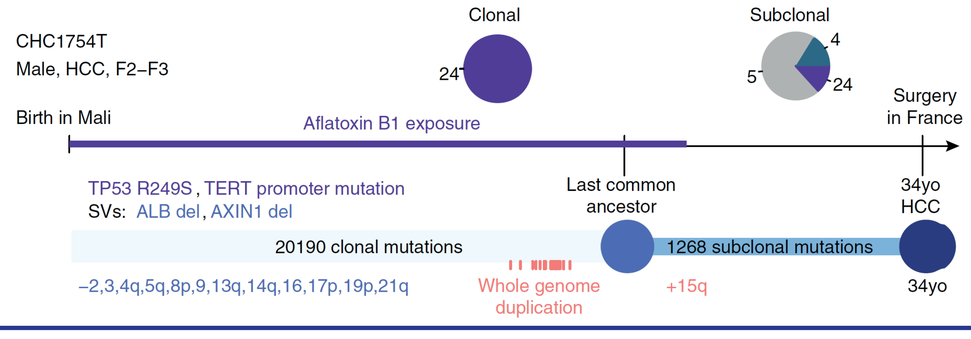
Variant-allele frequency (VAF) → **Cancer Cell Fraction (CCF)**
+ Tumor purity
+ Copy-number





$$CCF = VAF \times \frac{\rho N_t + (1-\rho) N_n}{\rho n_{chr}}$$

Comparison of mutational processes between clonal and subclonal mutations:



[5]

Clonal architecture of tumors - multiple samples

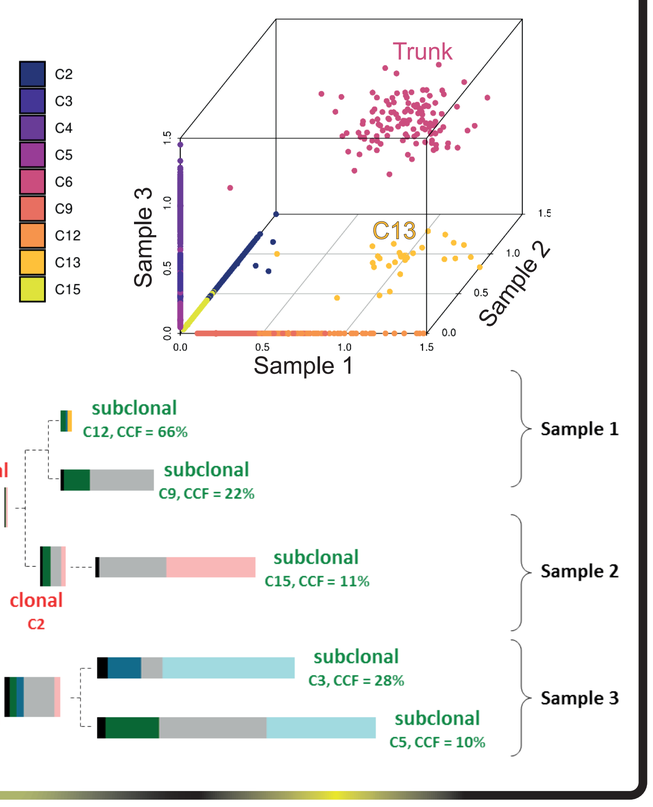
AVAILABLE SOON

For each mutation, compute the CCF in each sample

Bayesian Dirichlet process to identify clusters of CCF

Reconstruct the evolutionary tree of the tumor

Analyze mutational signatures in each branch



Corresponding Authors: TZ. Hirsch (theo.hirsch@inserm.fr); E. Letouzé (eric.letouze@inserm.fr)

Our GitHub

Scan me

https://github.com/FunGeST/Palimpsest

Get the poster

Scan me

[1] Alexandrov *et al*, Nature 2013. PMID: 23945592
[2] Shinde *et al*, Bioinformatics 2018. PMID: 29771315
[3] Bayard *et al*, Nature Communications 2018. PMID: 30531861
[4] Alexandrov *et al*, Preprint 2018. bioRxiv 322859
[5] Letouzé*, Shinde* *et al*, Nature Communications 2017. PMID: 29101368