Multi-factor Data Normalization enables the detection of LOH in amplicon sequencing data

by **Rim ZAAG**[1], Maxime LIÉNARD[1], Sébastien TOFFOLI[2] and Jean-François LAES[1]

1 : OncoDNA, 1 Rue Louis Breguet, 6041 Gosselies, Belgium
2 : Institute of Pathology and Genetics, 25 Avenue Georges Lemaître, 6041 Gosselies, Belgium

Context

Beyond its role as prognosis factor, LOH has recently been reported to play a key role in precision medicine, as it can predict response to immune checkpoint blockade therapies. Current LOH detection methods mainly rely on the comparison of the tumor genotype against its normal counterpart. However their use in routine diagnosis is limited by the frequent lack of paired normal DNA. On the other hand, the exploitation of high-throughput SNP array techniques to detect LOH by genotyping only tumor is not more advantageous for diagnosis as only 30% of SNPs in an individual sample are heterozygous, which makes the remaining 70% non-informative.

AIM OF THE PROJECT:

*Here we address this gap and we present **ALHASCA** (Algorithm for LoH identification in Amplicon Sequencing in CAncer), an algorithm for the detection of large LOH and copy number losses from high-depth tumor-only sequencing.*

ALHASCA principle

- Read coverage normalization to adress intra-library and technology-specific variability : Baseline construction.
- Normalized Amplification ratio calculation (AR).
- B allele frequency (BAF) and homozygosity state computation.
- AR and BAF integration to compute a local satus for each SNP marker as an independent variable :

- Long stretch Identification of successif SNPs with « LOH » local status (at least 3 SNPs)

- Confidence calculation for identified putative regions based on : selected SNPs ratio, curve smoothing, frameshift, damaging variants : High confident /Low confident

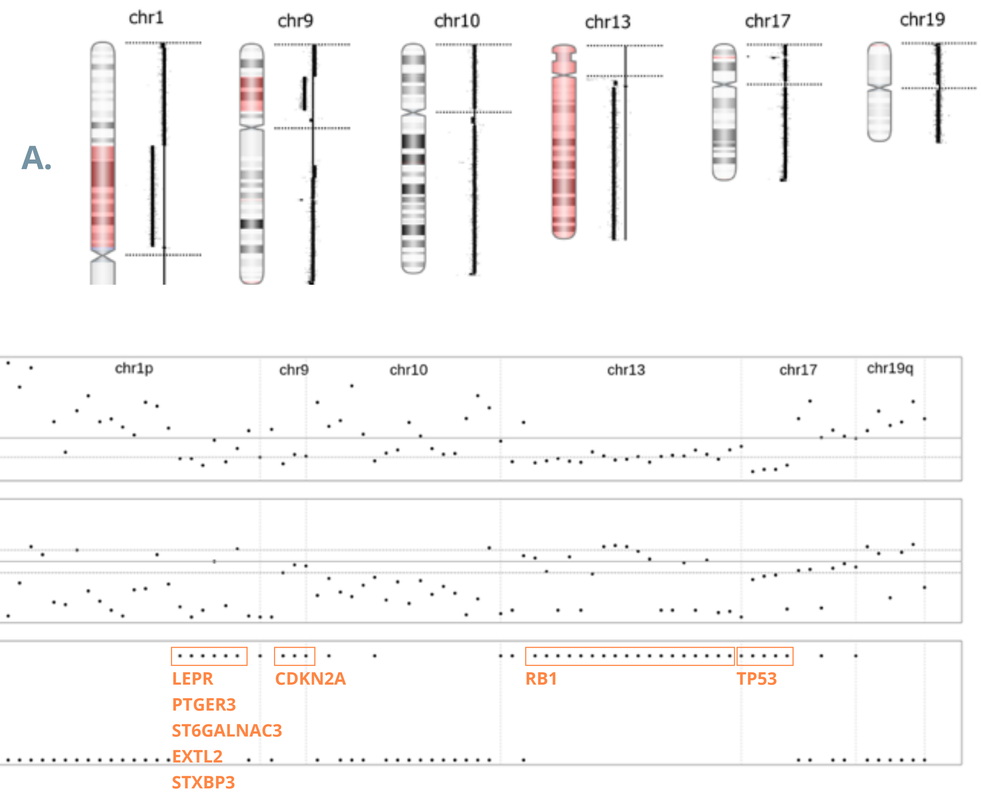
LOCAL STATUS COMPUTATION

A SNP IS MARKED «LOH» IF:

- AR < Threshold → **Homozygous deletion**
- AR < 1 & Homozygous SNP (BAF > 80% or BAF < 20%) → **Homozygous deletion**
- Heterozygous SNP with allelic imbalance (BAF ~ 66 % or 33%) → **Homozygous deletion**

Assay Design

- Assay designed based on the AmpliSeq technology (ThermoFisher Scientific Inc)
- Panel covering :
  - 330 cancer related genes
  - SNPs with a MAF of min 0.30% from dbSNP covering regions susceptible to undergo an LOH in cancer.
- Selected regions were submitted to Ion AmpliSeq Designer to design primer pools.



ALHASCA results :
A: *CGH profiles.*
B: *Amplification ratio.*
C: *Percentage of BAF values.*
D: *LOH events detected by ALHASCA. Boxes= regions detected with High confidence.*

Results

- **ALHASCA predictions were assessed againt CGH profiles for 10 tumor samples. Predictions compare favorably with results obtained from whole exome CGH data :**
  - ✓ Among 22 expected LOH events for the 10 samples : 100 % were correctly identified (20 with High confidence and 2 with Low confidence)
- **Ability to process samples with different tumor impurity levels : learning from positive samples allowed to determine a threshold of AR at 0,5.**
- **AR threshold at 0,5 correctly distinguishes between homozygous and hemizygous deletions even with normal DNA contamination.**

ALHASCA PERFORMANCE

Sensitivity	= **90.9%**
Specificity	= **82.35%**
Precision	= **76.92%**