🧬 Analyse Exomes CLOVES

Dr. Thomas Steimlé

Mars 2023

Laboratoire d'oncohématologie de l'hôpital Necker

Materials & Methods

👥 Matériel, N = 24 paires N/P

Alias	Cases	Patho	Normal
AUZAL	- -	612200191032	612200128560
BASPA	- -	2007N080798	2007N080327
BOUDE-1	- -	2006N080622	2006N080641
BOUDE-2	- -	2006N080656	2006N080641
BRIAN	- -	2105N190483	2105N170599
DEVCL	- -	2010N260761	2010N260754
DUMEL	- -	612200588717	612200588556
FARMO	- -	2105N210788	2105N210782
GOSJU	- -	612120459843	612120454105
KRAMA	- -	1911N040786	1911N040784
LABPH	- -	2006N080633	2006N080640
LASPH	- -	612120335709	612120316267
LECFA	- -	2008N170574	2007N270441
LHYET	- -	2006N080768	2006N080767
MASFA	- -	612100070595	612100070580
MOKRA	- -	1811N280725	1501N300498
MULAY	- -	6612112860	61210092424
OUATI	- -	2104N160342	2104N130615
PLANY	- -	2103N110343	2103N110124
REMMA	- -	2104N160304	2104N150297
ROQRO-1	- -	2006N080629	2006N080644
ROQRO-2	- -	612200572392	6122005723638
SAMAR	- -	1712N040796	612120165753
VERPY	- -	612120515645	612120504794

🧬 Séquençage selon la méthode *Agilent SureSelect Human All Exon V7 panel* sur automates *Illumina Next/NovaSeq*.

💾 Analyse bioinformatique:

Alignement (hg19 bwa mem v0.7.17-r1188) et déduplication UMI (umi_tools v1.1.1)

Callers :

⚙️ **mutect2** (GATK v4.2)
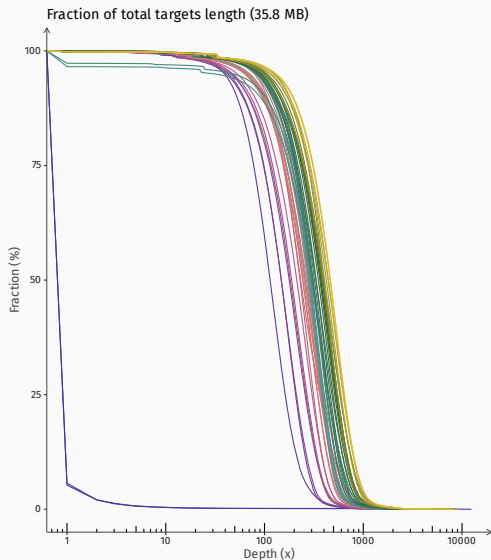⚙️ **strelka** (Illumina v2.9.10)
⚙️ **lancet** (NY Genome Center v1.1.0)

En suivant les modes opératoires displonibles (cf. diapos supplémentaires).
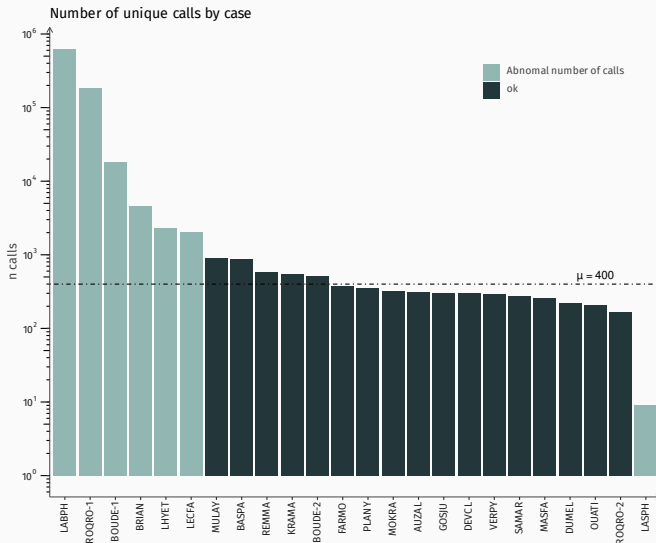
Results

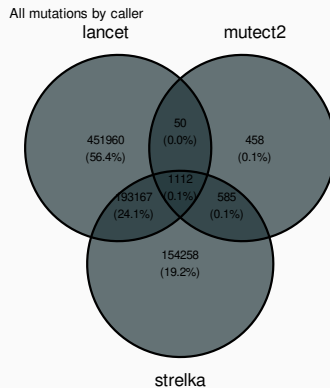## ⟷ Alignment – proportions of targeted sequences at given depth by case

Fraction of total targets length (35.8 MB)

Case	< 100x (%)
— LASPH_N	100
— LASPH_P	100
— BRIAN_P	41
— BOUDE-1_P	26
— KRAMA_N	26
— KRAMA_P	20
— MULAY_P	19
— MULAY_N	18
— DEVCL_N	15
— MASFA_N	11
— DEVCL_P	11
— BOUDE-2_P	11
— LECFA_N	11
— BASPA_P	11
— SAMAR_P	10
— LECFA_P	10
— BRIAN_N	10
— LHYET_N	10
— FARMO_N	9
— LHYET_P	9
— MASFA_P	8
— OUATI_P	7
— PLANY_P	7
— FARMO_P	7

Case	< 100x (%)
— MOKRA_P	7
— OUATI_N	7
— REMMA_N	7
— REMMA_P	6
— LABPH_P	6
— PLANY_N	6
— ROQRO-1_N	6
— MOKRA_N	6
— ROQRO-1_P	5
— BASPA_N	5
— GOSJU_N	5
— ROQRO-2_P	5
— VERPY_P	5
— BOUDE-1_N	5
— BOUDE-2_N	5
— SAMAR_N	5
— VERPY_N	5
— DUMEL_P	5
— DUMEL_N	4
— ROQRO-2_N	4
— GOSJU_P	4
— AUZAL_N	4
— LABPH_N	3
— AUZAL_P	3

4

📍 Variant calling
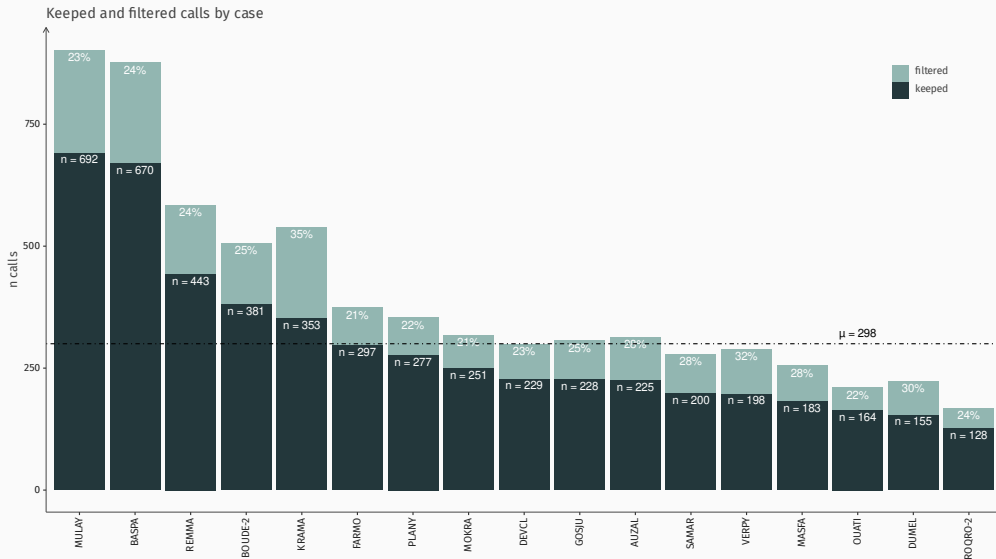

Number of unique calls by case

All mutations by caller

Cases on the left with an abnormal mutionnal burden and LASPH which is has been sequenced as a panel have been filtered out.

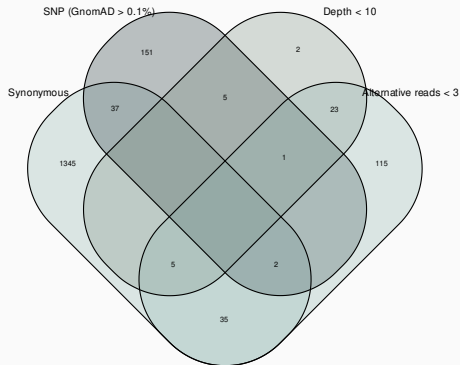# 🗑 Filtering – Proportion of filtered calls for each case

Keeped and filtered calls by case

legend:
- filtered
- keeped

n calls

μ = 298

Cases (x-axis): MULAV, BASPA, REMMA, BOUDE-2, KRAMA, FARMO, PLANY, MOKRA, DEVCL, GOSJU, AUZAL, SAMAR, VERPY, MASFA, OUATI, DUMEL, ROQRO-2

Bar labels:
- MULAV: 23% / n = 692
- BASPA: 24% / n = 670
- REMMA: 24% / n = 443
- BOUDE-2: 25% / n = 381
- KRAMA: 35% / n = 353
- FARMO: 21% / n = 297
- PLANY: 22% / n = 277
- MOKRA: 21% / n = 251
- DEVCL: 23% / n = 229
- GOSJU: 25% / n = 228
- AUZAL: 30% / n = 225
- SAMAR: 28% / n = 200
- VERPY: 32% / n = 198
- MASFA: 28% / n = 183
- OUATI: 22% / n = 164
- DUMEL: 30% / n = 155
- ROQRO-2: 24% / n = 128

Number of calls filtered by each five applied filters

Number of calls for each allelic fraction after filtering
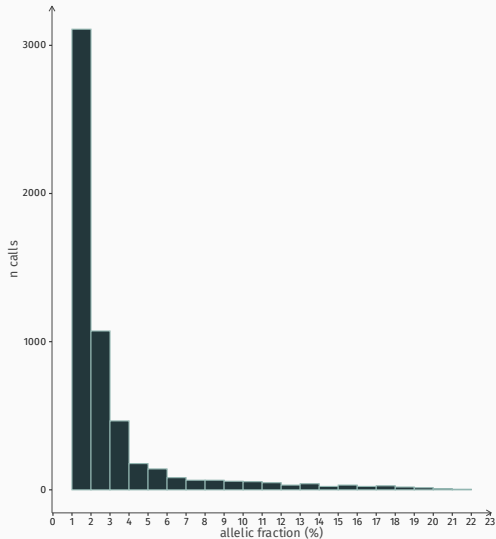
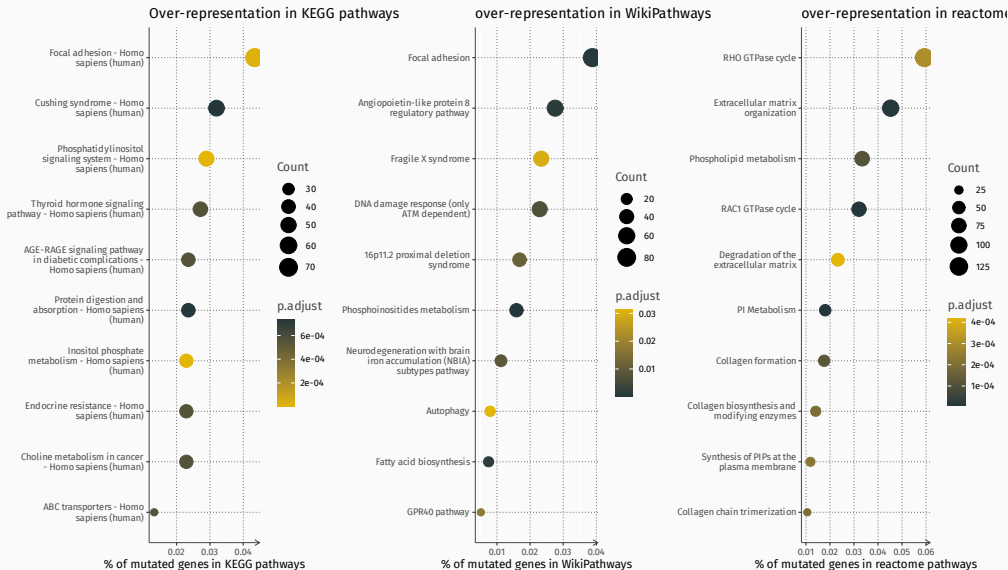**📝 Results – Recurrent genes known as tumor suppressors or proto-oncogenes.**

Alias	Gene	Variant	VAF (%)
BASPA	*ALK*	p.Gly1202GlufsTer56	2.9
REMMA	*ALK*	p.Lys1267Asn	1.4
REMMA	*ARID1A*	p.Thr290Pro	9.3
PLANY	*ARID1A*	p.Arg1223His	1
MULAY	*ARID1A*	p.Met1759Ile	2.4
MULAY	*ARID1A*	p.Ala532Val	5.9
PLANY	*KMT2C*	p.His4278Asn	1.4
FARMO	*KMT2C*	p.Ala117Thr	1
DEVCL	*KMT2C*	intron	0.6
BASPA	*NCOA3*	p.Gln758Ter	2.5
MOKRA	*NCOA3*	p.Pro115Leu	1.1
REMMA	*NCOR1*	p.Arg627Leu	1.7
MULAY	*NCOR1*	intron	3.7
MULAY	*NCOR1*	p.Gly1340Cys	1.3
GOSJU	*NOTCH1*	p.Gln1837Ter	2.7
GOSJU	*NOTCH1*	p.Gly251Asp	1.2
ROQRO-2	*NOTCH1*	p.Ala1418Thr	1.1
BOUDE-2	*NOTCH1*	intron	1.4
ROQRO-2	*NOTCH2*	p.Gln2006Ter	1.2
BASPA	*NOTCH2*	p.His1208Asn	1.5
MOKRA	*PIK3CA*	p.Glu545Lys	3.5
DEVCL	*PIK3CA*	p.Glu542Lys	4.2
SAMAR	*SMARCA4*	p.Met797Ile	2
MASFA	*SMARCA4*	p.Phe939Leu	1.9

Filtered mutated genes over-representation in GO terms

Over-representation in KEGG pathways

over-representation in WikiPathways

over-representation in reactome

PHOSPHATIDYLINOSITOL SIGNALING SYSTEM
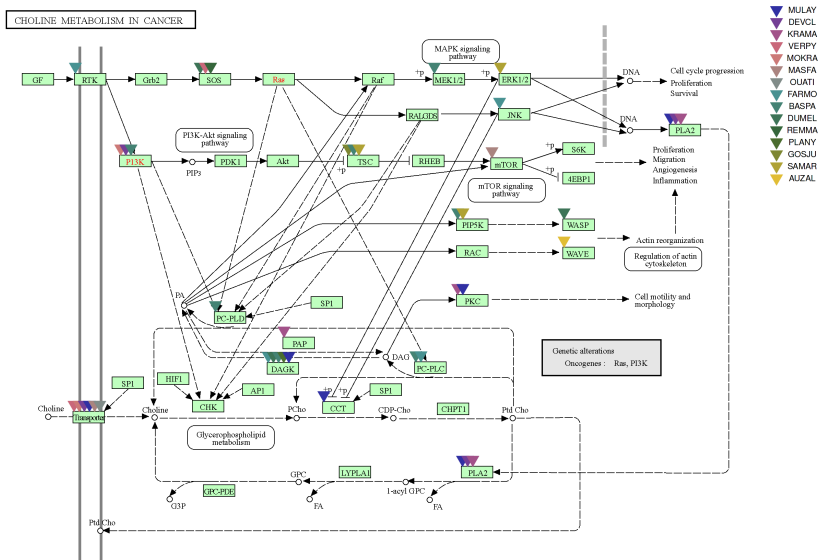
Complexes:

2.7.1.107
diacylglycerol kinase (ATP)
2.7.1.149
1-phosphatidylinositol-5-phosphate 4-kinase
2.7.1.153
phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha/beta/delta
2.7.1.154
phosphatidylinositol-4-phosphate 3-kinase
2.7.1.67
phosphatidylinositol 4-kinase A
2.7.1.68
1-phosphatidylinositol-4-phosphate 5-kinase
2.7.11.13
classical protein kinase C alpha type
2.7.4.21
inositol-hexakisphosphate 5-kinase
2.7.4.24
inositol-hexakisphosphate/diphosphoinositol-pentakisphosphate 1-kinase
3.1.3.-
phosphatidylinositol 4-phosphatase
3.1.3.25
myo-inositol-1(or 4)-monophosphatase
3.1.3.36
inositol polyphosphate 5-phosphatase INPP5B/F
3.1.3.48/ 3.1.3.64/ 3.1.3.95
myotubularin-related protein 3/4
3.1.3.64/ 3.1.3.95
myotubularin-related protein 1/2
3.1.3.66
inositol polyphosphate-4-phosphatase
3.1.3.86
phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1
3.1.4.11
phosphatidylinositol phospholipase C, beta

11

04070 7/7/21

Complexes:
2.7.1.149
1-phosphatidylinositol-5-phosphate 4-kinase
2.7.1.153
phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha/beta/delta
2.7.1.154
phosphatidylinositol-4-phosphate 3-kinase
2.7.1.67
phosphatidylinositol 4-kinase A
2.7.1.68
1-phosphatidylinositol-4-phosphate 5-kinase
3.1.3.-
phosphatidylinositol 4-phosphatase
3.1.3.25
myo-inositol-1(or 4)-monophosphatase
3.1.3.36
inositol polyphosphate 5-phosphatase INPP5B/F
3.1.3.48/3.1.3.64/3.1.3.95
myotubularin-related protein 3/4
3.1.3.56
inositol polyphosphate 5-phosphatase INPP5J/K
3.1.3.64/3.1.3.95
myotubularin-related protein 1/2
3.1.3.66
inositol polyphosphate-4-phosphatase
3.1.3.86
phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1
3.1.4.11
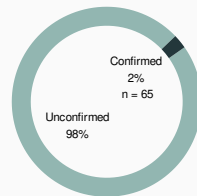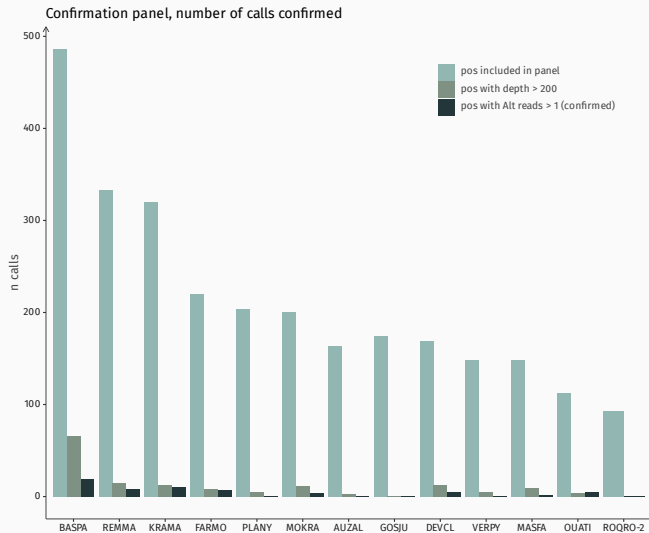phosphatidylinositol phospholipase C, beta

13

CHOLINE METABOLISM IN CANCER

Confirmation panel, number of calls confirmed

- pos included in panel
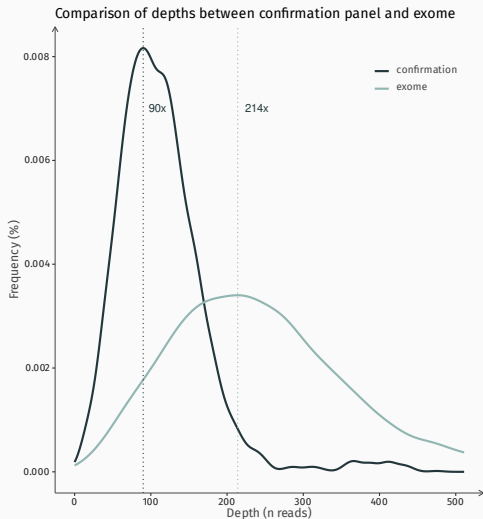- pos with depth > 200
- pos with Alt reads > 1 (confirmed)

Confirmed
2%
n = 65

Unconfirmed
98%

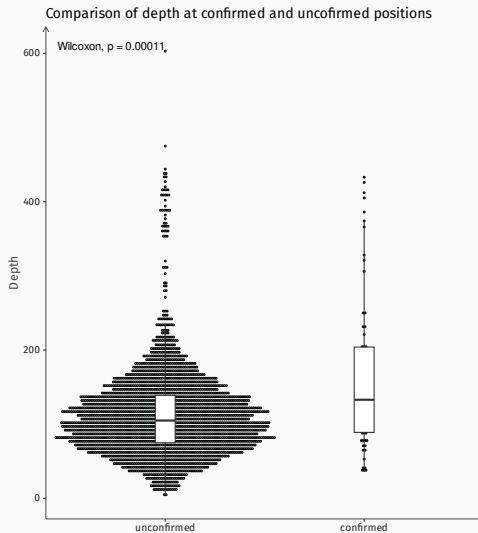Confirmation panel results: 3646 positions have been included in the confirmation panel, 65 have been confirmed.

Comparison of depths between confirmation panel and exome

The exome has a better depth for the variants positions we wanted to confirm

Comparison of depth at confirmed and unconfirmed positions

The sequencing depths at positions we wanted to confirm are deeper in the exome. In the confirmation panel, confirmed calls positions have significantly deeper depth.

16

Conclusions

⇒ 17 éxomes analysés, environ 300 mutations par cas.

⇒ Mutations retrouvées majoritairement à des VAF allant de 1 à 2 %

⇒ Récurrence de mutations sur les gènes connus pour être impliqués dans le cancer: *ALK*, *ARID1A, KMT2C, NCOA3, NCOR1, NOTCH1, NOTCH2, PIK3CA* et *SMARCA4*.

⇒ Net enrichissement de mutations touchant des gènes impliqués dans le métabolisme des Inositol Phosphates.

⇒ Panel de confirmation insuffisamment couvert cependant il confirme 65 mutations.

Supplementaries

Source code : https://github.com/nygenome/lancet

```
1 lancet --ref hg19.fa
2     --tumor  {TUMOR.BAM}
3     --normal {NORMAL.BAM}
4     --bed {AGILENT_REGIONS_V7.BED}
5     --num-threads 31
6     > {OUTPUT.VCF}
```

Listing 1: lancet – bash version

📁 Methods > Calling > Manta puis Strelka

Manta : https://github.com/Illumina/manta

```
1  configManta.py --exome
2      --referenceFasta hg19.fa
3      --tumorBam  {TUMOR.BAM}
4      --normalBam {NORMAL.BAM}
5      --callRegions {AGILENT_REGIONS_V7.BED}
6      --runDir /tmp/...
```

Listing 2: manta – bash version

Strelka : https://github.com/Illumina/strelka

```
1  configureStrelkaSomaticWorkflow.py --exome
2      --referenceFasta hg19.fa
3      --tumorBam  {TUMOR.BAM}
4      --normalBam {NORMAL.BAM}
5      --callRegions {AGILENT_REGIONS_V7.BED}
6      --indelCandidates {MANTA_candidateSmallIndels.vcf.gz}
7      --runDir /tmp/...
```

Listing 3: strelka – bash version

From : https://gatk.broadinstitute.org/hc/en-us/articles/360035531132

```
1  gatk-4.2.1.0/gatk --java-options -Xmx32g
2      Mutect2
3      -R hg19.fa
4      -L {AGILENT_REGIONS_V7.BED}
5      -I {TUMOR.BAM}
6      -I {NORMAL.BAM}
7      --normal-sample {NORMAL}
8      --germline-resource af-only-gnomad.raw.sites.chr.vcf
9      --panel-of-normals pon.vcf.gz
10     --f1r2-tar-gz {TUMOR_tumoral_f1r2.tar.gz}
11     -O {OUTPUT.VCF}
```

Listing 4: Mutect2 – seule étape adaptée – bash version