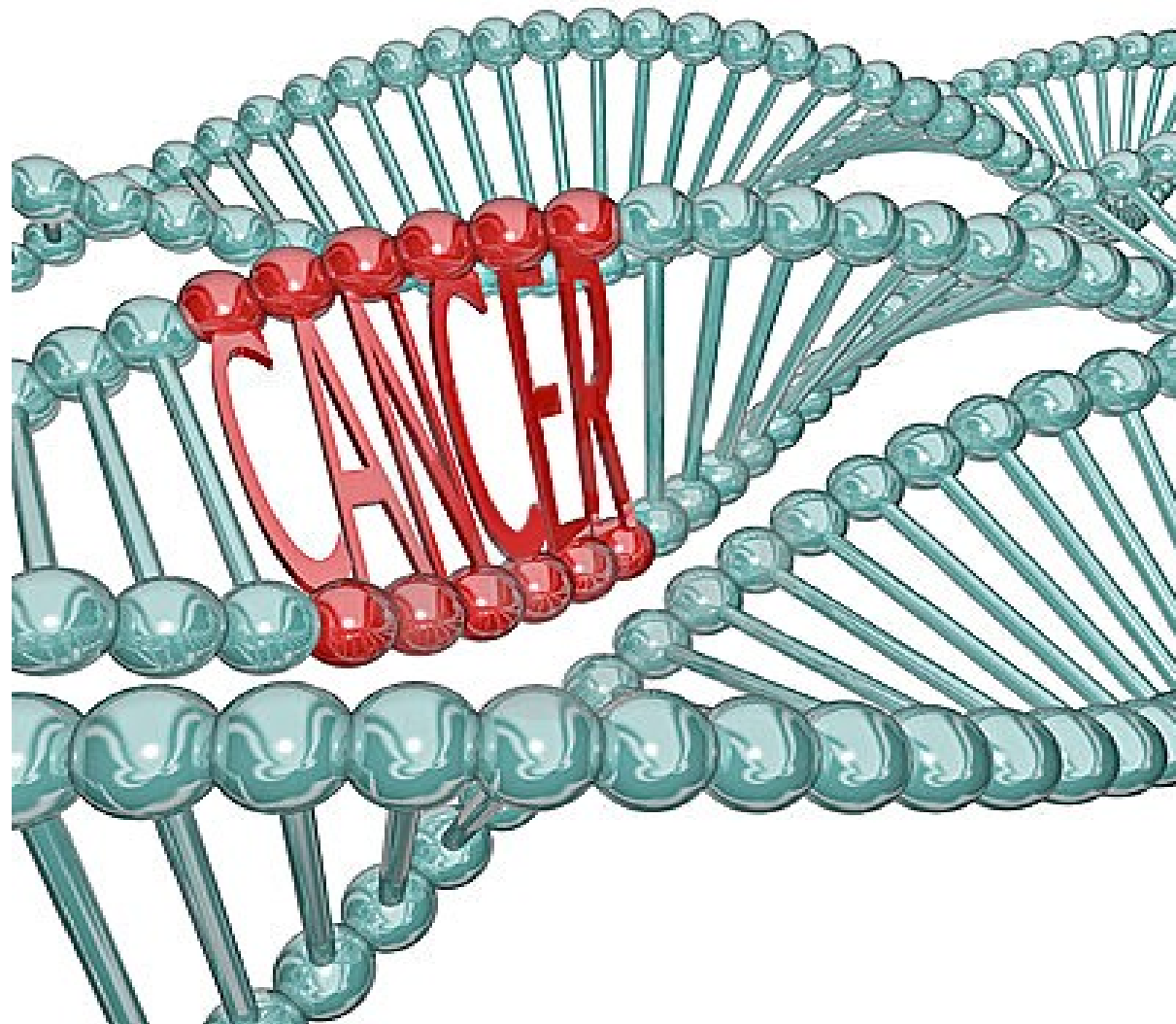


**Sylvain Mareschal** [[@etu.univ-rouen.fr](mailto:@etu.univ-rouen.fr)]

# Stratégies de séquençage à haut débit en oncologie



## Stratégies de séquençage à haut débit en oncologie

### *Continuité du cours de Gaël Nicolas*

#### **Génétique somatique ≠ constitutionnelle**

Cellularité, clonalité → VAF faibles

Polymorphismes à différencier des événements somatiques

#### **Recherche d'événements récurrents**

Mécanismes biologiques qui puissent expliquer les symptômes

Identifier des cibles thérapeutiques

Identifier des sous-types tumoraux

Identifier des bio-marqueurs

#### **Plan du cours (1h)**

- Introduction
- Un cas d'école : BRAF & HCL
- Quelle technique ?
- Quelle profondeur ?
- Combien de patients ?
- Conclusion

# 1. Un cas d'école : *BRAF* & HCL

## Hairy-Cell Leukemia

### Forme rare de leucémie

Accumulation de « tricholeucocytes » (lymphocytes B)  
0,2 cas pour 100 000 personnes / an (USA)

### Bien prise en charge

Analogues des purines  
DFS 10 ans ~ 70%

→ Quelles causes génétiques ?

## Whole-Exome Sequencing (WES)

### Cellules normales et tumorales d'un patient

CD19 : marqueur des lymphocytes B (normaux ou tumoraux)

Tri cellulaire du sang périphérique

- CD19+ avant traitement : tumeur ++
- CD19- après traitement : normal ++

### Séquençages indépendants

Capture de 38 Mb d'exons (Agilent SureSelect V1)

Séquençage paired-ends Illumina GAIIx 2x108 pb, 70x

Alignement au génome de référence hg18

### Détection des variants somatiques

Calling séparés

Élimination (drastique) des variants :

- synonymes
- également détectés dans les cellules normales (FDR < 10<sup>-6</sup>)
- présents dans moins de 25% des reads

### 5 variants somatiques retenus

*BRAF*, *CSMD3*, *SLC5A1*, *CNTN6* et *OR8J1*

Sanger : présents dans la tumeur, absents du normal

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

## *BRAF* Mutations in Hairy-Cell Leukemia

Enrico Tiacci, M.D., Vladimir Trifonov, Ph.D., Gianluca Schiavoni, Ph.D., Antony Holmes, Ph.D., Wolfgang Kern, M.D., Maria Paola Martelli, M.D., Alessandra Pucciarini, Ph.D., Barbara Bigerna, B.Sc., Roberta Pacini, B.Sc., Victoria A. Wells, B.Sc., Paolo Sportoletti, M.D., Valentina Pettirossi, Ph.D., Roberta Mannucci, Ph.D., Oliver Elliott, M.Sc., Arcangelo Liso, M.D., Achille Ambrosetti, M.D., Alessandro Pulsoni, M.D., Francesco Forconi, M.D., Livio Trentin, M.D., Gianpietro Semenzato, M.D., Giorgio Inghirami, M.D., Monia Capponi, M.D., Francesco Di Raimondo, M.D., Caterina Patti, M.D., Luca Arcaïni, M.D., Pellegrino Musto, M.D., Stefano Pileri, M.D., Claudia Haferlach, M.D., Susanne Schnittger, Ph.D., Giovanni Pizzolo, M.D., Robin Foà, M.D., Laurent Farinelli, Ph.D., Torsten Haferlach, M.D., Laura Pasqualucci, M.D., Raul Rabadan, Ph.D., and Brunangelo Falini, M.D.

N ENGL J MED 364:24 NEJM.ORG JUNE 16, 2011

## Focus sur *BRAF*

### Bon candidat *a priori*

Hotspot connu dans les mélanomes et cancers thyroïdiens  
Implication dans la voie MAPK, connue dans les HCL

### Séquençage Sanger de 47 patients

Séquençage ADN limité à l'exon 15 de *BRAF*

47/47 présentent la mutation V600E

10/10 ADN normaux appariés en sont exempts

### Extension à 195 leucémies diverses

Aucune autre V600E retrouvée

Dont 38 leucémies à présentation similaire

## 2. Quelle technique utiliser ?

### Whole Genome Sequencing (WGS)

Séquençage direct d'ADNg fragmenté

- + **couverture homogène**
- + **données exhaustives** (translocations comprises)
- ... **trop exhaustives** (intergénique, introns ...)
- **cible étendue** (3.2 Gb, profondeur ou coûts élevés ...)

### RNA-sequencing

Capture des ARN (sondes poly-T)  
Filtration et normalisation éventuelles

- + **données complémentaires** (expression, jonctions d'exons)
- ~ **données biaisées** (ARN poly-A)
- ~ **cible raisonnable** (50 Mb, ~ WGS / 64)
- **couverture très hétérogène** (tributaire du niveau d'expression)

### Whole Exome Sequencing (WES)

Fragmentation de l'ADNg + capture (puce, billes ...)

- ~ **données biaisées** (exons répertoriés dans les banques)
- ~ **cible raisonnable** (50 Mb, ~ WGS / 64)
- **couverture hétérogène** (biais de %GC ...)

### Targeted resequencing

Fragmentation + capture  
Amplification multiplexe

- + **cible réduite** (<100 kb, possibilité de multiplexer)
- **couverture hétérogène** (biais de capture ou d'amplification)
- **données réduites** (sélection *a priori*, découvertes limitées)

### Et bien d'autres

#### **CHIP-seq**

Capture des régions qui interagissent avec une protéine donnée

#### **ATAC-seq**

Capture de l'ADN libre (euchromatine, empreinte des protéines)

...

### Un choix à faire selon ses objectifs

#### **Caractérisation d'un génome rare**

Le WGS reste le plus exhaustif et le moins biaisé

#### **Recherche de mutations codantes sans a priori**

Le WES offre un bon rapport quantité / prix

#### **Analyse conjointe des mutations et CNV**

WGS reste moins biaisé que WES en terme de CNVs  
Considérer aussi WES + aCGH, coûts similaires

#### **Analyse conjointe des mutations et expressions**

Vu le coût actuel des puces transcriptomiques,  
WES + puce peut s'avérer plus complet qu'un RNA-seq

#### **Analyse de l'épissage alternatif**

RNA-seq reste la seule solution, bien que limitée

#### **Confirmation de récurrence**

#### **Application clinique**

#### **Recherche de variant rare (intra/inter génome)**

Reséquençage ciblé pour minimiser les coûts

### Dans tous les cas

#### **Séquencer aussi l'ADN normal !!!**

Seul moyen fiable de distinguer somatique et constit.  
Double le nombre d'échantillons, les coûts

# 3. Quelles analyses en aval ?

## Détecter un variant somatique

Exemple : *Koboldt et al, Genome Res 2012 (VarScan2)*

Approche heuristique, basée sur le test exact de Fisher  
Reconnu comme l'un des plus sensibles, paramétrage délicat  
*Wang et al, Genome Medicine 2013*  
*Stead et al, Human Mutation 2013*

### 1. Pré-filtration des positions génomiques

- profondeur minimale (8x)
- MAF minimale (8%)
- minimum de reads non-référence (2x)
- minimum de qualité ( $q_{base} > 20, q_{map} > 1$ )

### 2. Attribution indépendante des génotypes

- <8% = wild-type
- >75 % = homozygote
- sinon = hétérozygote

### 3. Tests de Fisher

	Tumoral	RR	RV	VV
	RR	reference	somatic	somatic
Normal	RV	LOH	germline	LOH
	VV	?!!	?!!	germline

#### Normal = Tumoral

Comparaison du total de reads variants aux erreurs de séquençage attendues (0,1%)

	R	V
Normal + Tumoral	1423	6
0,1% d'erreurs	1428	1

$p = 0,12$

#### Normal ≠ Tumoral

Comparaison des nombres de reads référence et variants entre normal et tumoral

	R	V
Normal	31	3
Tumoral	18	14

$p = 0,0017$

## Mutations récurrentes

### Filtration fonctionnelle

Annotation automatisée (VEP, ANNOVAR ...)  
faux / non sens, algorithmes de prédiction (SIFT, PolyPhen ...)

### Recherche de récurrence entre patients

- A la même position génomique
- A la même position protéique (attention aux isoformes ...)
- Au sein d'un même domaine protéique
- Au sein d'un même gène

## Études d'enrichissement

### Recherche de récurrence au sein d'un pathway

Y a-t-il un enrichissement en gènes mutés (une fois ou plus) au sein de ce groupe de gènes (voie métabolique, annotation GO ...)?

#### - Tests de Fisher + correction des tests multiples

dans ce pathway	non	oui
gène muté	419	12
gène non muté	19420	43

$p = 1,7e-9$  }  $\times 10^3$  pathways

#### - Music PathScan

*Wendel et al, Bioinformatics 2011*

Probabilité par gène d'être muté en fonction de sa taille et du BMR  
Combinaison au sein du pathway, par patient puis dans la série

- ...

## Et bien d'autres

### Recherche de CNV

### Recherche de translocations

Ebauchés dans les cours précédents

# 4. A quelle profondeur séquencer ?

## Stead et al, Hum Mutat 2013

### Comparaison de variant callers

VarScan2, MuTect, BCFtools, GATK UG, LoFreq, SNVer, Atlas2

### Différents jeux de données

- Données synthétiques (Q37 + 1% d'erreurs)
- Données Illumina simulées (simNGS)
- Données réelles (dilution d'une lignée mutée par un ADN normal)

Attention, profondeur minimale  $\ll$  moyenne

Variant	1 BAM	normal + tumoral
50 %	30x	
25 %	60x	
10 %	> 100x	
5 %	> 250x	100x - 250x
2,5 %		500x
1 %		> 750x

} Génomique constit.  
 } Génomique somatique

## Shallow sequencing

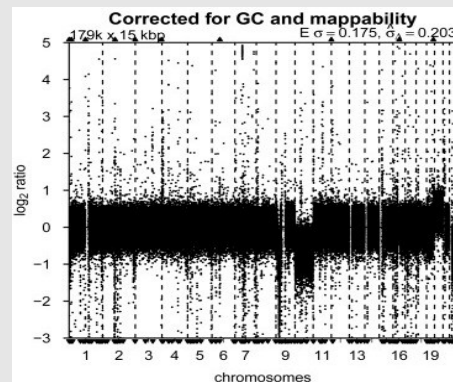
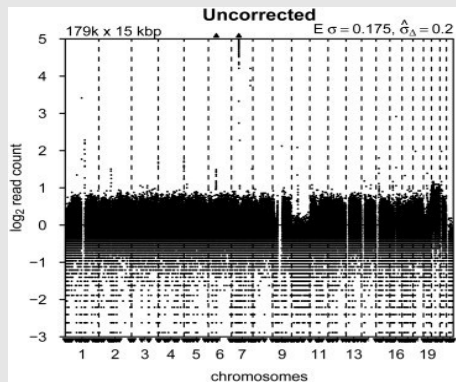
Exemple : Scheinin et al, Genome Res. 2014

### Séquençage

- Illumina HiSeq2000, 1x50 bp, génome entier
- Multiplexage ~20 échantillons / piste
- coût / patient inférieur à une puce CGH
- profondeur finale = **0,1x**

### Analyse

- Nombre de reads / bin (15kb ~ puce CGH 180k)
- Correction LOESS 2D (%GC, mappability)
- Exclusion des régions difficiles (répétitions ...)
- Segmentation CBS



## Deep sequencing

Exemple : Abdelhamid E et al, Am J Hematol. 2013

### Séquençage

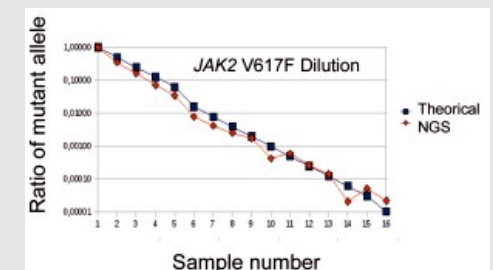
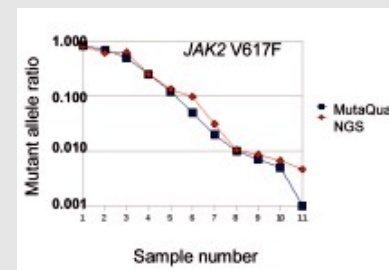
- Life PGM 1x100 bp, 1 amplicon de 81 bp
- Un échantillon par puce 314 (600 000 reads espérés)
- Ratio des reads mutés / totaux
- profondeur finale = **600 000 x**

### Comparaison à un test de biologie moléculaire

- Sensibilité jusqu'à 1e-3 (mutaQuant™)
- Corrélation fréquence d'allèles mutés > 0,95

### Mesure dans un échantillon dilué

- ADN muté dilué dans un ADN normal
- Corrélation jusqu'à une fréquence de 1e-5



# 5. Combien de patients séquencer ?

## Lawrence et al, Nature 2014

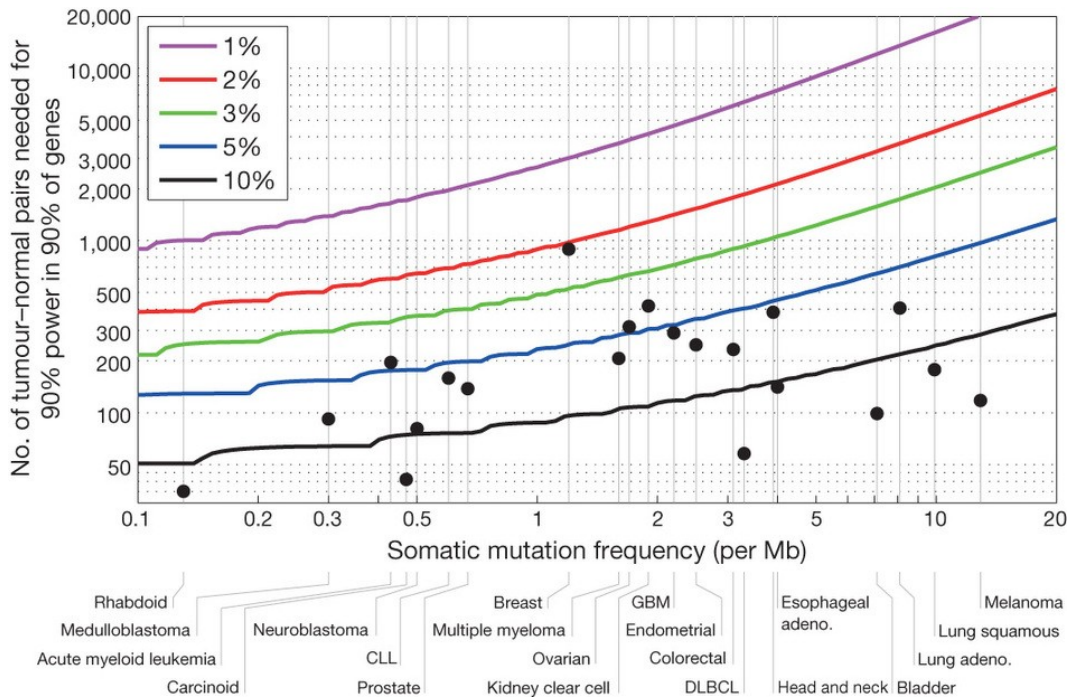
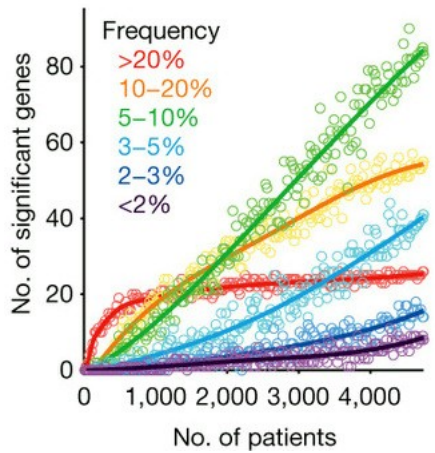
**Collecte de données d'exomes**  
 21 sous-types de cancers  
 4742 WES normal + tumoral  
 ~2000 issues de *The Cancer Genome Atlas*

**Etude conjointe**  
 Pipeline identique (Broad institute)  
 3e6 SNV, 77e3 indels, 30e3 block substitutions  
 Taux de mutation de 0,03 à 7000 mutations / Mb  
 Recherche de gènes fréquemment mutés

**Analyse de saturation**  
 Sous-échantillonnage progressif, suivi des découvertes  
 Nombre de gènes détectés comme mutés  
 Selon la proportion de patients mutés (≠ VAF)

**Analyse de puissance**  
 Objectif : 90% de puissance, 90% de gènes détectés  
 Autorise 10% de mutations manquées (faux négatifs)  
 Selon la proportion de patients mutés (+ BMR)

Tumour type	n	Median somatic mutation frequency (per Mb)	References
Acute myeloid leukaemia	196	0.4	TCGA
Bladder	99	7.1	TCGA
Breast	892	1.2	TCGA, Banerji et al. Nature 2012
Carcinoid	54	0.5	Francis et al. Nat Genet 2013
Chronic lymphocytic leukaemia	159	0.6	Wang et al. NEJM 2011, Landau et al. Cell 2013
Colorectal	233	3.1	TCGA, Bass et al. Nat Genet 2011
Diffuse large B-cell lymphoma	58	3.3	Lohr et al. PNAS 2012
Endometrial	248	2.5	TCGA
Oesophageal adenocarcinoma	141	4.0	Dulak et al. Nat Genet 2013
Glioblastoma multiforme	291	2.2	TCGA
Head and neck	384	3.9	TCGA, Stransky et al. Science 2011
Kidney clear cell	417	1.9	TCGA
Lung adenocarcinoma	405	8.1	TCGA, Imielinski et al. Cell 2012
Lung squamous cell carcinoma	178	9.9	TCGA
Medulloblastoma	92	0.3	Pugh et al. Nature 2012
Melanoma	118	12.9	Berger et al. Nature 2012, Hodis et al. Cell 2012
Multiple myeloma	207	1.6	Chapman et al. Nature 2011
Neuroblastoma	81	0.5	Pugh et al. Nat Genet 2013
Ovarian	316	1.7	TCGA
Prostate	138	0.7	Berger et al. 2011, Barbieri et al. Nat Genet 2012
Rhabdoid tumour	35	0.1	Lee et al. J Clin Invest 2012



## **Bruit de fond important**

Polymorphismes → séquencer l'ADN normal apparié est primordial  
Mutations somatiques aléatoires (BMR moyen de 0,1 à 20 mutations / Mb)

## **Recherche d'événements récurrents**

Position récurrente (génomique, protéique ...)  
Voie métabolique commune (études d'enrichissement)

## **Identifier ses besoins avant de séquencer**

### **Quelle technologie ?**

- **Illumina** reads courts (30 à 100 bp), qualité élevée
- **Life** taille et qualité des reads intermédiaires (100 à 400 bp)
- **PacBio** reads longs (> 1kb), qualité faible
- **nPore** reads encore plus longs (> 10kb), qualité très faible

### **Quelle technique ?**

WGS, WES, RNA-seq, target-seq ...

### **Quelle profondeur ?**

Deep-sequencing, shallow-sequencing ... VAF ciblée (cellularité comprise !)

### **Combien de patients ?**

Rareté de la mutation ciblée (fréquence au sein de la population)

→ **Pour quelle application ?**

## **Connaitre les limites de son design**

**Puissance** généralement limitée

- Ce qui est visible reste correct (peu de FP)
- Combien d'événements similaires a-t-on raté ? (FN)

## **Profiter des données publiques**

The Cancer Genome Atlas