# Next-Generation Sequencing applied to tailor targeted therapies in lymphoma : the RELYSE project

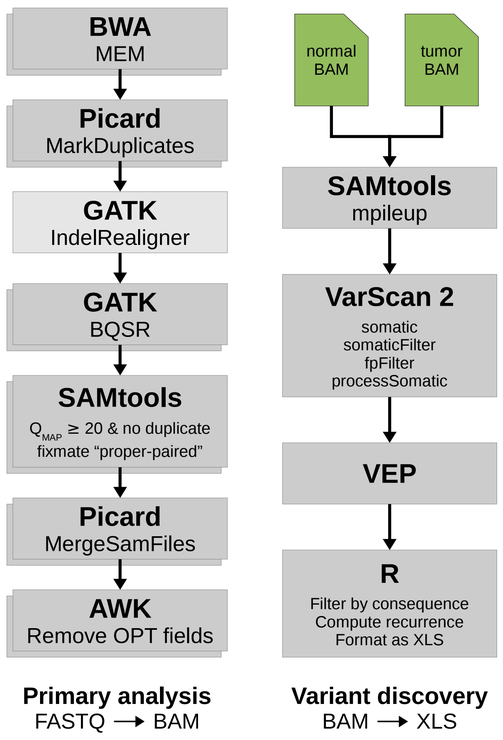
Sylvain Mareschal [1,4], Pierre-Julien Viailly [1,4,5], Philippe Bertrand [1], Fabienne Desmots-Loyer [2], Elodie Bohers [1], Catherine Maingonnat [1], Karen Leroy [3], Thierry Fest [2], Fabrice Jardin [1,4]

[1] INSERM 918, Centre Henri Becquerel, IRIB, Rouen, France
[2] INSERM 917, Centre Hospitalo-Universitaire de Rennes, France
[3] INSERM 955, AP-HP, Hôpital Henri Mondor, Créteil, France
[4] Université de Rouen, Normandie Université, France
[5] LITIS EA 4108, IRIB, Saint-Etienne du Rouvray, France

*Non-Hodgkin Lymphomas (NHL) are lymphoid cell malignancies accounting for about 4% of all cancers, with an incidence rate of 12 cases per 100,000 and per year in Europe. While recently developed immuno-chemotherapies like Rituximab have significantly enhanced their prognosis, a large part of these patients still relapses or is refractory to treatments. Recent advances in the field of high-throughput sequencing have provided concrete solutions to support these patients, as whole-exome sequencing of large NHL cohorts has highlighted several recurrent mutations, and benchtop sequencers are now available to quickly screen these mutations at diagnosis to tailor personalized therapies. The RELYSE project is an embodiment of this strategy, initiated by the french LYSA group in 2013.*

# 1. Target selection

## Whole Exome Sequencing




To complete published datasets with variants from atypical lymphomas, WES was performed in **4 "leg-type"** DLBCLs and **14 DLBCLs refractory** to treatments (relapse < 1 year).

Sequencing was performed in matched normal and tumoral tissues using the **SureSelect™** exon capture kit and **Illumina** sequencers (GA IIx and HiSeq 2500 respectively).

Average coverages of 53|82 x were met, mapping 96|98 % of the 63|91 $10^6$ reads to the reference genome, from which 75|71 % were on target.

660|2,810 variants were retained after filtration, 5|50 loci and 27|265 genes were found mutated in more than one sample.

## Confrontation with published datasets

| | Morin et al Blood 2013 | Zhang et al PNAS 2012 | RELYSE Refractory | RELYSE Leg-type | Bohers et al GCC 2014 | Lohr et al PNAS 2011 | Rossi et al SCB 2013 | |
|---|---|---|---|---|---|---|---|---|
| **Pathology** | DLBCL | DLBCL | DLBCL | leg-type | DLBCL | DLBCL | | |
| **Series** | 53 WES | 94 WES | 14 WES | 4 WES | 161 Sanger | 49 WES | | Review |
| **Scope** | RELYSE target regions | | | | Targets | Full genes | | |
| ***B2M*** | 9% | 4% | | | NA | 10% | | |
| ***BCL2*** | 13% | 7% | 7% | | NA | 22% | | |
| ***BRAF*** | | | | | NA | | | HCL(100%) |
| ***CARD11*** | 11% | 12% | 14% | | 11% | 20% | | DLBCL(10%) |
| ***CD58*** | 4% | | 7% | | NA | 10% | | |
| ***CD79A*** | | | | | 3% | | | DLBCL(20%) |
| ***CD79B*** | 8% | 6% | | | 6% | 16% | | DLBCL(20%) |
| ***CDKN2A **** | | | | | | | | |
| ***CDKN2B **** | | | | | | | | |
| ***CIITA*** | | | 14% | | NA | 10% | | |
| ***CREBBP*** | 9% | 6% | | 25% | NA | 16% | | |
| ***EP300*** | 8% | | | | NA | 6% | | |
| ***EZH2*** | 21% | 4% | | | 13% | 14% | | |
| ***FOXO1 ***** | 2% | | | | NA | | | |
| ***GNA13*** | 11% | 7% | 29% | | NA | 20% | | |
| ***ID3*** | 2% | | | | NA | | | BL(70%), SMZL(34%) |
| ***IRF4*** | 2% | | 21% | | NA | 2% | | |
| ***ITPKB*** | 2% | | 21% | | NA | 10% | | |
| ***MEF2B*** | 8% | 3% | 14% | | NA | 18% | | |
| ***MFHAS1*** | 2% | | 14% | | NA | 2% | | |
| ***MLL2*** | 23% | | 21% | | NA | 29% | | |
| ***MYC*** | 8% | 3% | | | NA | 8% | | |
| ***MYD88*** | 13% | 15% | 21% | 75% | 19% | 12% | | WM(90%), DLBCL(30%) |
| ***NOTCH1*** | | | | | NA | | | MCL(12%), CLL(11%), DLBCL(8%) |
| ***NOTCH2*** | 2% | 4% | 7% | | NA | 4% | | SMZL(20%), DLBCL(8%) |
| ***PIM1*** | 13% | 10% | 29% | 50% | NA | 31% | | |
| ***PRDM1*** | 6% | 2% | 7% | | NA | | | |
| ***SOCS1*** | 2% | 2% | 21% | | NA | 6% | | |
| ***STAT6*** | 4% | | 14% | | NA | 4% | | |
| ***TCF3*** | 2% | | | | NA | 2% | | BL(70%) |
| ***TNFAIP3*** | | 7% | 14% | | NA | 2% | | MCL(44%), WM(40%), DLBCL(30%) |
| ***TNFRSF14*** | 11% | | | | NA | 22% | | |
| ***TP53*** | 17% | 14% | 21% | | NA | 24% | | |
| ***XPO1*** | 2% | | 14% | | NA | 4% | | |

* *CDKN2A/B* were found deleted in 40% of 203 DLBCLs (Lenz et al, PNAS 2008)

** *FOXO1* was found mutated in 8% of 301 DLBCLs (Trinh et al, Blood 2013)

**BL**
Burkitt Lymphoma

**CLL**
Chronic Lymphocytic Leukemia

**DLBCL**
Diffuse Large B-Cell Lymphoma

**HCL**
Hairy Cell Leukemia

**MCL**
Mantle Cell Lymphoma

**SMZL**
Splenic Marginal Zone Lymphoma

**WM**
Waldenström Macroglobulinemia

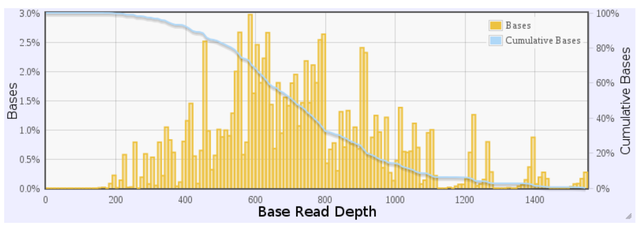
The integration of WES results from internal and published series, in close collaboration with clinical hematologists, led to the selection of **272 loci** (exons, functional domains or mutation hotspots) from **34 genes** of interest in lymphomas, covering **87.7 kb** of the human genome. The resulting Ampliseq™ design includes 872 amplicons, divided into 3 pools.

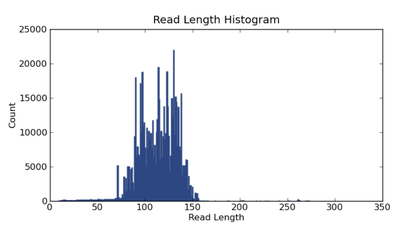
# 2. PGM™ implementation

## AmpliSeq™ VS Haloplex™

### Ampliseq™, Cancer Hotspot Panel V2

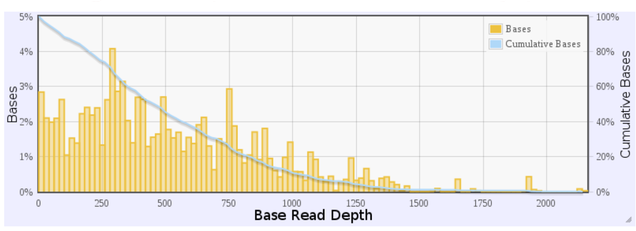
10 ng of template DNA, two day protocol, unique mapping pattern

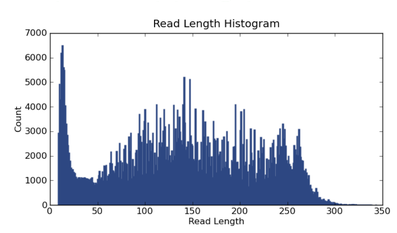



### Haloplex™, Cancer Panel

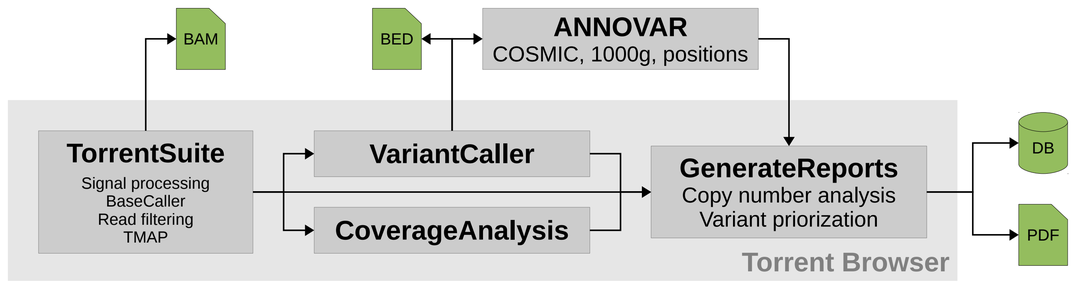
200 ng of template DNA, three day protocol, multiple mapping patterns




To compare the two targeting chemistries available for the PGM implementation of the RELYSE panel, 4 biopsies of blastic NK cell lymphomas were assessed on two "314" chips using the Ampliseq™ and Haloplex™ cancer panels. The analysis, restricted to the 15.5 kb of overlap in 36 genes between the two designs, showed a more homogeneous depth distribution in Ampliseq™ results, with less uncovered positions and unusable reads. This coupled with lighter protocol and DNA template requirements made us favor the **Ampliseq™** chemistry for this project.

## Analysis pipeline




Most parts of the **Torrent Suite ™** pipeline provided by Life Technologies were retained for data analysis, as it mainly consists of open source tools modified to benefit from IonTorrent™ chemistry specifics.

A plugin for the Torrent Browser ™ interface was developed to complete this analysis: after calling, variants are **annotated** by ANNOVAR, prioritized according to their proteic impact and annotation in databases (dbSNP, COSMIC), then formatted into a per-sample **PDF report**. Finally, depth comparison with a pool of normal sample results provides a probability for each gene of the panel to be **amplified** or **deleted** in the sample.

# 3. From bioinformatics to bedside

## Result validations

| Variant score | Sanger – | Sanger + |
|---|---|---|
| < 9.5 | 30 | 1 |
| gray zone | 14 | 7 |
| > 22 | 3 | 224 |

522 of the 4,375 variants detected in a first batch of 57 patients were validated by **Sanger** sequencing, focusing on recurring COSMIC variants. 11 recurring variants never validated in multiple assays were discarded and flagged as "**technical artefacts**" for future filtering. VariantCaller™ scores of the remaining variants were used to define thresholds separating **false positives** (Q<9.5) from **true positives** (Q>22), and a **gray zone** in which Sanger validation is required for confirmation (7.2 / 76.7 variants per analyzed sample in average).

**Illumina MiSeq™** and **PyroMark™** validation assays are currently under development, to optimize the thresholds for low frequency variants currently under Sanger detection range.

## Prospective application

As of today, the RELYSE panel has the potential to **recommend an experimental targeted therapy** in lymphomas with alterations in *BCL2*, *CREBBP*, *EP300*, *EZH2*, *CD79A/B*, *CARD11*, *MYD88* or *TNFAIP3* (Intlekofer & Younes, NRCO 2014). To this end, 300 patients are planned to be screened at relapse, beginning early 2015.

For each of the 57 patients in the first batch, **on average 7.0 of 76.7 variants were found informative** (COSMIC annotated or predicted effect, true positives) and 47.3 silent (no effect, true positives).

**Monthly meetings** between clinicians, pathologists and bioinformaticians are being set up in our institution as part of the CAMELEON project, to discuss sequencing results from *de novo* DLBCLs and orient treatments according to the molecular profile of the tumor and available clinical trials.

13th European Conference on Computational Biology (ECCB)
2014 September 7-10, Strasbourg

Many thanks to Sydney Dubois for her consciencious copy-editing and color picking !