H62

LOD and Sensitivity of MyHEME: A Comprehensive Targeted Sequencing Strategy

Andrew R. Carson, Ph.D.[1], Bradley A. Patay, M.D.[1], Valerie McClain M.S.[1], Patrick Liu, M.S.[1], Suzanne M. Graham, Ph.D.[1], Tim Stenzel, M.D. , Ph.D.[1,2,3], and Jeffrey E. Miller, Ph.D.[1,2,3]

[1]Invivoscribe, San Diego, CA, [2] LabPMM LLC, San Diego, CA and [3]LabPMM GmbH, Martinsried, Germany

Abstract

Background: Next-generation sequencing (NGS) of hematopoietic and lymphoid neoplasm genomes promises to revolutionize oncology, with the ability to design and use targeted drugs, to predict outcome and response, and to classify patients' responses to treatment more thoroughly using more predictive combinations of mutations. It is of critical importance that the NGS platform chosen, the chemistry utilized and the well vetted bioinformatics are then applied consistently for future adoption in clinical decisions. To illustrate the increased capacity and resolution of NGS for the comprehensive characterization of patients with hematologic cancers, we sequenced both clinical patient samples and contrived cell lines using a novel specific targeted strategy involving DNA and RNA. Using contrived cell lines, we utilize maximum gene coverage, long read lengths, and higher sequencing depth to accurately detect variants, indels and breakpoints critical in both hematologic development and for tracking minimal residual disease (MRD). Furthermore, understanding mutations in the context of clonal architecture may prove crucial for personalized therapies. We demonstrate limit of detection that allows for complete characterization of the molecular changes, with high sensitivity and specificity, thus allowing clonal architecture discovery.

Methods: To examine these cancers, we targeted coding exons (571 genes) and potential genomic breakpoint regions within known somatic gene fusions (371 genes) comprising the MyHEME™ gene panel. We sequenced target loci on the Illumina® MiSeq® platform to an average depth of coverage of 1000x for cell lines and patient samples. Using a custom bioinformatics pipeline, we performed thorough mutation detection analyses to identify single nucleotide variants (SNVs), indels, inversions and translocations. In addition, we calculated allelic frequencies to investigate potential aneuploidy, LOH and clonality. Using contrived samples, we were able to determine limit of detection rates, sensitivity and specificity.

Results: Our analyses of targeted sequencing results from cell lines identified the published genomic variants within MyHEME targeted genes. Critically, our assay enabled detection of variants as low as 5%. In many cases, these variants were more fully characterized for their precise genomic breakpoints and inserted sequence content.

Conclusion: We demonstrate that by specifically targeting driver genes using the MyHEME gene panel, we can comprehensively characterize mutations for AML, ALL, Non-Hodgkin's, Multiple Myeloma cell lines and patients with these hematologic conditions. Our results show this assay can comprehensively characterize the cancer genome of patients, identifying not only primary clones, but secondary clones that are present in at least 5% of the patient's sample.

Materials and Methods

Target baits for MyHEME

- DNA baits: Targets the coding sequences of **571 genes**
- RNA baits: Targets the transcripts of **371 genes**

Analysis method

- 1µg of DNA or 0.1µg of RNA is used as input before hybridizing to the MyHEME baits.
- Captured targets are then sequenced on the Illumina® platform.
- Customized bioinformatics pipeline identifies and characterizes SNVs, indels, and SVs.

Samples used to evaluate quality metrics:

- NIST human reference sample **NA12878** (aka "Genome in a Bottle")
  - High confidence variants are used to evaluate true positives (TP) and false negatives (FN).
  - Sanger sequenced confident regions validated to contain no false positives or false negatives. Used to calculate false positives (FP) and true negatives (TN).
- Contrived samples containing dilutions of **6 cell lines at different allelic frequencies**.
  - Used to analyze LOD, reproducibility and linearity of variant detection.
- **6 cell lines with known gene fusions** used to evaluate the ability to detect fusions in RNAseq data.

DNA Target Gene Coverage and Sequencing Depth

We evaluated target coverage (Figure 1) and sequencing depth (Figure 2) across the coding sequences from 571 genes. These analyses incorporate data from 16 samples, including 8 runs of NA12878 and 2 runs of 4 different contrived samples from cell line dilutions.

Figure 1: MyHEME DNA Target Coverage

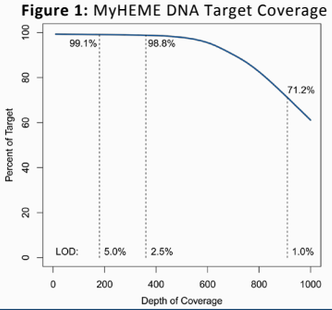
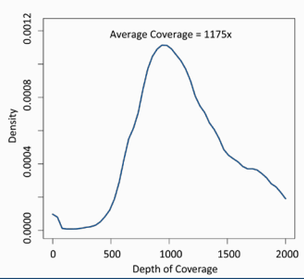


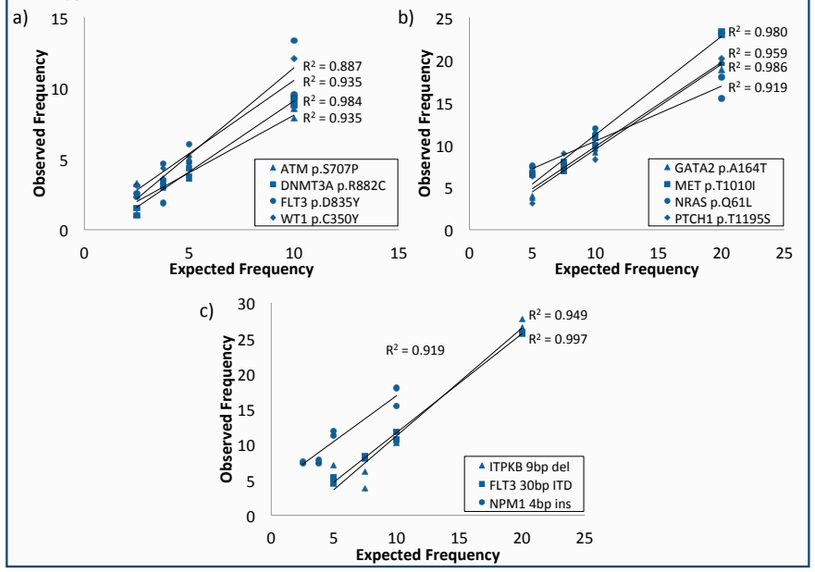
Figure 2: MyHEME DNA Sequencing Depth



MyHEME DNA Results: Limit of Detection and Linearity

To estimate the limit of detection (LOD) and linearity of DNA variant detection using MyHEME, we use contrived samples comprised of 6 AML cell lines. Five cell lines were diluted into a 6th cell line (background) at the following dilutions: **5%**; **7.5%**; **10%**; **20%**. Note, for heterozygous variants, the allelic frequency is half of the dilution, so we are testing LOD as low as 2.5%.

Figure 3: Linearity of a) 4 heterozygous SNVs; b) 4 homozygous SNVs and c) 1 heterozygous and 2 homozygous indels



MyHEME DNA Results: Sensitivity and Specificity

To evaluate MyHEME's DNA variant detection sensitivity and specificity, we sequenced the NIST human reference sample **NA12878**. The GIAB consortium sequenced this "Genome in a Bottle" multiple times on multiple platforms to generate an integrated "gold standard" dataset containing:

- A set of 3,641,994 high-confidence variants. Of these variants, there are:
  - **656** high-confidence coding variants (640 SNVs and 16 indels) within MyHEME targets
  - **2,171** high-confidence non-coding variants (1,948 SNVs and 223 indels) within MyHEME targets
  - High-confident variants were used as gold-standard **true positives** for sensitivity analyses
- High-confidence regions containing 2,565,300,578 bp with highly accurate genotype calls, include non-variant sites (homozygous reference calls). Of these bases:
  - **1,594,796** of these bases overlap with MyHEME coding targets and **2,202,265** bases overlap with MyHEME non-coding targets
  - Non-variant sites were used as gold-standard **true negatives** for specificity analyses

Table 1: Evaluation of Sensitivity and Specificity of high-confidence a) coding and b) non-coding variants in 8 NA12878 MyHEME analyses

	SNVs				Indels			
	Coding (n=5,120)		Non-Coding (n=128)		Coding (n=15,584)		Non-Coding (n=1,784)	
	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
Sensitivity	99.8%	99.8%	99.8%	99.8%	100%	100%	95.6%	95.6%
Specificity	94.9%	98.3%	95.7%	98.6%	87.1%	97.7%	83.1%	84.7%

Sensitivity is calculated as Detected True Positives / Gold-Standard True Positives (n in above table)
Specificity is calculated as Detected True Positives / All Detected Variants

- Using a SNV cutoff of 2.5% and an indel cutoff of 5.0%, we observe:
  - >95% sensitivity for both coding and non-coding SNVs and indels
  - 95% specificity for SNVs and >80% specificity for indels

MyHEME DNA & RNA Results: Translocations and Gene Fusions

MyHEME DNA baits contain targets to detect structural variants that occur within the breakpoint hotspot in *KMT2A* and in small introns adjacent to targeted exons.

- The 6 cell lines included in the contrived dilution samples contain the following detectable translocations:
  - t(9;22)(q34.12;q11.23*)(ABL1;BCR)* translocation
  - t(9;22)(q34.13;q11.1*)(NUP214;XKR3)* translocation
- These cell lines were sequenced 8 times at: 2.5%, 3.75%, 5% and 10% allelic frequencies

Table 2: DNA detection of translocations

Translocation	Genes	Detected	False Negatives
t(9;22)	*ABL1 – BCR*	8	0
t(9;22)	*NUP214 – XKR3*	8	0

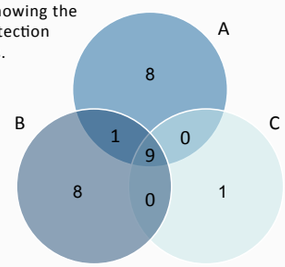
MyHEME RNA baits contain targets to detect gene fusions that occur within any of **371 genes**.

- We sequenced 6 different cell lines containing a known gene fusion:
  - t(1;19)(*TCF3;PBX*)
  - t(9;22)(*BCR;ABL1*) – b2a2 (e13/a2)
  - t(9;22)(*BCR;ABL1*) – b3a2 (e14/a2)
  - t(8;21)(*RUNX1;RUNX1T1*)
  - t(15;17)(*PML;RARA*) – "L-form"
  - inv(16)(*CBFB;MYH11*)
- We use 3 different RNA fusion finding programs to improve sensitivity and specificity for the detection of gene fusions
  - All fusions were detected with their expected fusion types/forms

Table 3: Evaluation of gene fusion detection and sensitivity using 3 gene fusion detection methods

Program	Total Fusions	Known Fusions	Sensitivity
A	18	6	100%
B	18	6	100%
C	12	6	100%
Combined	9	6	100%
2 of 3	10	6	100%

Figure 4: Venn diagram showing the overlap of gene fusion detection from 3 different programs.



Note: Of the 4 novel fusions using 2 of 3 programs, 3 are actually reciprocal gene fusions of one of the known fusions. The other is a *NUP214-XKR3* fusion observed with high confidence by all 3 programs, and confirmed by DNA translocation analysis.

Conclusions

Using sequence data obtained from 1) the NIST reference **NA12878**, 2) contrived samples containing **dilutions of 6 AML cell lines**, and 3) **6 cell lines with known gene fusions**, we established:

- **Variant Sensitivity > 95%**
  - Sensitivity was highest for SNVs (99.8%).
- **Variant Specificity of 95% for SNVs and >80% for indels**
  - Using an LOD of 5%, our coding specificity for both SNVs and indels is >97%
- **Limit of Detection of at least 5% allelic frequency for >99% of the coding bases of targeted genes**
  - In addition, as much as 98% of the coding bases of the targeted genes should have an LOD of at least 2.5% and potentially >70% of the coding bases should have an LOD of 1.0%
- **Significant linearity for detection of SNVs and indels, including pathogenic mutations such as *FLT3*/ITD**
- **We are able to detect structural variants using both DNA and RNA**
  - Can detect translocations with an LOD as low as 2.5%
  - Combining 3 gene fusion programs has a very high sensitivity with a low false positive detection rate.

We demonstrated that MyHEME is a highly sensitive, accurate and reproducible assay that can comprehensively characterize mutations within samples from a variety of hematological malignancies.

MiSeq® and Illumina® are registered trademarks of Illumina, Inc.