invivoscribe

H03

Analysis and Characterization of Hematologic Cancers using a Comprehensive NGS Panel Comprised of DNA and RNA Baits Targeting 704 Genes

Andrew R. Carson[1], Bradley A. Patay[1], Suzanne M. Graham[1], Valerie McClain[1], Zhiyi Xie[1], Timothy Stenzel[1,2,3], and Jeffrey E. Miller[1,2,3]
[1]Invivoscribe, Inc., [2]LabPMM, LLC, San Diego, CA, and [3]LabPMM, GmbH, Martinsried, Germany

Abstract

Introduction: As next-generation sequencing (NGS) methodologies improve, so does the ability to characterize hematopoietic and lymphoid neoplasm genomes. This promises to revolutionize oncology, allowing more accurate and precise classification of patients and potentially leading to novel targeted and combination therapies with improved outcomes.

Methods: The MyHEME™ targeted-sequencing panel is comprised of two independent bait sets that target a combined 704 genes known or predicted to contribute to hematologic cancers (DNA baits for 571 genes and RNA baits for 361 genes; 228 genes are found in common between the two bait sets). Libraries were constructed using 1 µg of DNA or 0.1 µg of RNA and sequenced on an Illumina® platform. Sequenced reads are analyzed using proprietary MyInformatics™ software to identify single-nucleotide variants (SNVs), indels, and structural variants (SVs). Both the MyHEME panel and MyInformatics software were created under ISO 13485-design control. To characterize the performance metrics of the MyHEME panel, we used the NIST human reference sample NA12878 along with combinations of hematologic cancer derived cell lines with known pathogenic variants at various allelic frequencies.

Results: Analytical validation of the MyHEME panel established an average read depth of 1,175x (with a median read depth of 1,088x) for the DNA targets and an average transcripts per million (TPM) of 2,256 (with a median TPM of 743) for the RNA targets. For the DNA targets, we establish sensitivity >95% (99.8% for SNVs at a 2.5% limit of detection (LoD); 100% for coding indels at a 5.0% LoD) and specificity >95% (95.5% for SNVs at a 2.5% LoD; 97.7% for coding indels at a 5.0% LoD). We also show the ability to cross-confirm results between the 228 genes common to both the DNA and RNA targets. Importantly, novel gene fusions, which are generally difficult to detect and validate, were cross-confirmed when observed in both the DNA and RNA targets. For example, we identified a novel t(9;22) translocation causing a *NUP214-XKR3* gene fusion using both the DNA and RNA targets. Additionally, while RNA data provides the fused exons of the transcripts, DNA data gives the precise genomic breakpoint coordinates.

Conclusion: MyHEME LDT assay service is an extensive panel for sensitively and specifically identifying SNV, indel, and SV mutations in 704 target genes. This panel can comprehensively characterize mutations in multiple, diverse hematologic cancer samples, including AML, ALL, Non-Hodgkin Lymphoma and Multiple Myeloma. By utilizing a high depth of coverage, MyHEME can accurately detect clones present down to 5% of a patient's sample. In addition, by targeting both DNA and RNA, MyHEME contains a built in validation method to cross-confirm novel variants of interest.

Materials & Methods

704 genes are targeted by MyHEME baits

- Genes were chosen due to published associations with hematologic cancers
- DNA baits: Targets the coding sequences of **571 genes**
- RNA baits: Targets the transcripts of **371 genes**
- **228 genes** are targeted by both DNA and RNA baits, allowing cross-validation of observed variants

Analysis method

- 1 µg of DNA or 0.1 µg 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 confirm true positives (TP) and identify false negatives (FN)
  - Sanger-sequenced regions are used to identify false positives (FP) and confirm 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

MyHEME DNA Results: Coverage and Sequencing Depth

We evaluated DNA 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 contrived samples from cell-line dilutions.

Figure 1: MyHEME DNA Target Coverage

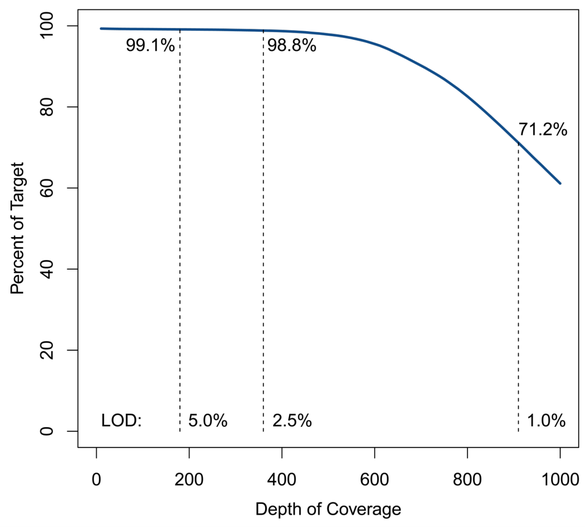
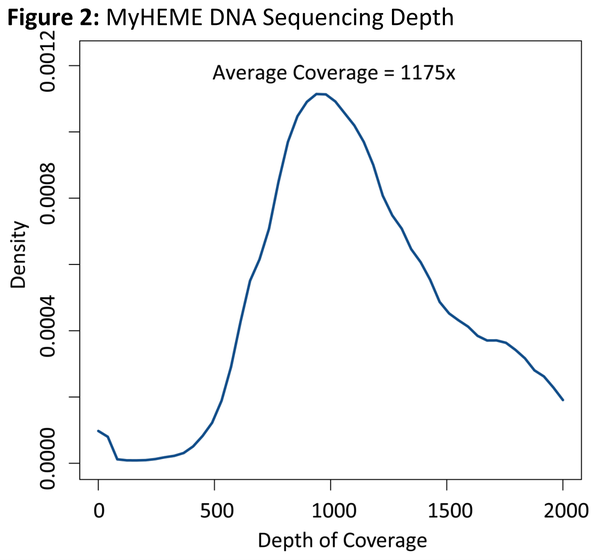


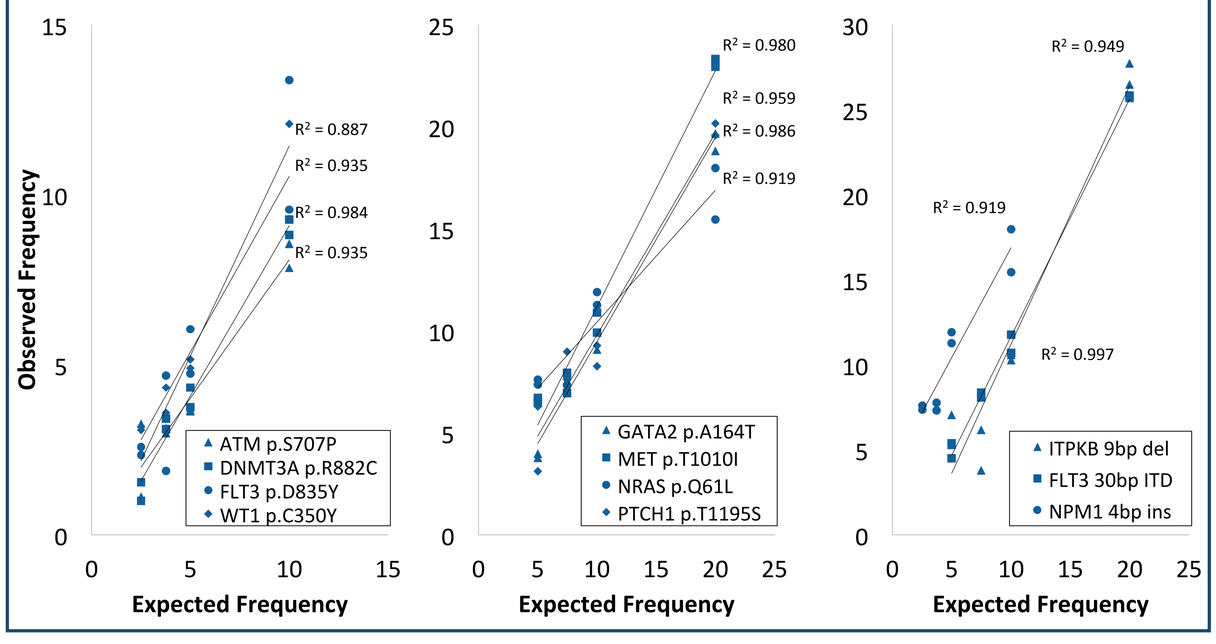
Figure 2: MyHEME DNA Sequencing Depth



MyHEME DNA Results: Limit of Detection and Linearity

To estimate the LoD and linearity of DNA variant detection using MyHEME, contrived samples comprised of 6 AML cell lines were used. Five cell lines were diluted into a 6th cell line (background) at the following dilutions: 5%, 7.5%, 10%, and 20%. Note, for heterozygous variants, the allelic frequency is half of the dilution, so LoD was tested at levels 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 MyHEMEs DNA variant detection sensitivity and specificity, the NIST human reference sample **NA12878** was sequenced. 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:
  - **1,594,796** bp and **2,202,265** bp overlap MyHEME coding and non-coding targets (respectively)
  - Non-variant sites were used as gold-standard **true negatives** for specificity analyses

Table 1: Sensitivity and Specificity of a) coding and b) non-coding variants in 8 NA12878 MyHEME runs at VAF cutoffs of 2.5% and 5.0%

	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%, the following was observed:
  - >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 in the contrived dilution samples contain the following detectable translocations:
  - t(9;22)(*BCR;ABL1*) – b3a2 (e14a2)
  - t(9;22)(*NUP214;XKR3*)
- 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

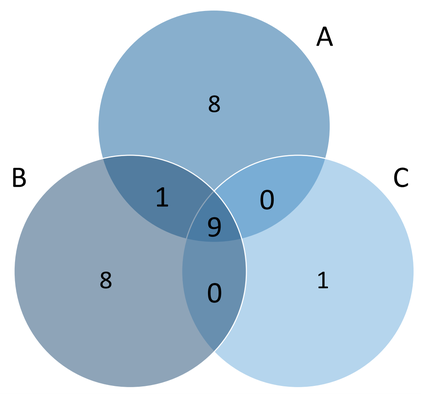
MyHEME RNA baits contain targets to detect gene fusions that occur within any of **371 genes**.

- We sequenced 6 different cell lines containing a known gene fusion:
  - t(1;19)(*TCF3;PBX*)
  - t(9;22)(*BCR;ABL1*) – b3a2 (e14a2)
  - t(15;17)(*PML;RARA*) – "L-form"
  - t(9;22)(*BCR;ABL1*) – b2a2 (e13a2)
  - t(8;21)(*RUNX1;RUNX1T1*)
  - inv(16)(*CBFB;MYH11*)
- We use 3 RNA fusion finding programs 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 reciprocal gene fusions of one of the known fusions. The other is a t(9;22)(*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**, the following was 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%
- **LoD 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**
- **Structural variants using both DNA and RNA can be detected**
  - Translocations with an LoD as low as 2.5% can be detected
  - Combining 3 gene fusion programs results in a very high sensitivity with a low false-positive detection rate

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

AMP Global, April 3–5 2017, Berlin, Germany

invivoscribe