

Targeted RNA-sequencing for the quantification of measurable residual disease in acute myeloid leukemia

Laura W. Dillon,¹ Sheida Hayati,^{1,2} Gregory W. Roloff,¹ Ilker Tunc,³ Mehdi Pirooznia,³ Antonina Mitrofanova² and Christopher S. Hourigan¹

¹Laboratory of Myeloid Malignancies, Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD; ²Department of Health Informatics, Rutgers School of Health Professions, Rutgers, The State University of New Jersey, Newark, NJ and ³Bioinformatics and Computational Biology Core Facility, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

©2019 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2018.203133

Received: July 26 2018.

Accepted: August 28, 2018.

Pre-published: August 31, 2018.

Correspondence: hourigan@nih.gov

SUPPLEMENTARY METHODS

Cell culture

Leukemia cell lines positive for fusion genes targeted by the AML MRD panel were cultured according to the supplier's guidelines. The K-562 cell line harbors the t(9;22)(q34;q11) genomic rearrangement and expresses the *BCR-ABL1* p210 fusion transcript¹. The Sup-B15 cell line (obtained from American Tissue Culture Collection, Cat#CRL-1929) harbors the t(9;22)(q34;q11) genomic rearrangement and expresses the *BCR-ABL1* p190 fusion transcript². The Kasumi-1 (Kas-1) cell line harbors the t(8;21)(q22;q22) genomic rearrangement and expresses the *RUNX1-RUNX1T1* fusion transcript³. The ME-1 cell line (kindly provided by Dr. Paul Liu, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD) harbors the inv(16)(p13q22) genomic rearrangement and expresses the *CBFB-MYH11* type A fusion transcript⁴. The NB4 cell line (kindly provided by Dr. Gabriel Ghiaur, Sidney Kimmel Cancer Center at Johns Hopkins, Johns Hopkins University School of Medicine, Baltimore, MD) harbors the t(15;17)(q22;q12) genomic rearrangement and expresses the *PML-RARA* fusion transcript⁵.

Clinical samples

Peripheral blood samples were collected from a healthy adult donor and from a 46-year-old female with monocytic AML who underwent myeloablative matched related donor allogeneic stem cell transplant (NHLBI protocol # 07-H-0113). Peripheral blood mononuclear cells (PBMCs) were isolated using Lymphocyte Separation Medium (Cat#

50494, MP Biomedicals, Solon, Ohio). Additional patient samples were collected on local institutional IRB-approved biobanking protocols by collaborators.

RNA isolation

RNA was isolated from cell lines and healthy donor PBMCs using the NucleoSpin RNA Plus Kit (Cat# 740984.50, Macherey-Nagel, Germany) and from patient PBMCs or bone marrow using the AllPrep DNA/RNA Mini Kit (Cat# 80204, Qiagen, Germany). RNA was quantitated using the NanoDrop Lite Spectrophotometer (Nano Drop Technologies, Wilmington, DE).

Primer design

Targeted amplicon primers were designed for the AML MRD target genes and fusion genes of interest (Supplementary Table 1) for use with the QIAseq Targeted RNA Panel system (Cat#333025, Qiagen, Germany). For each target, a BC and LA primer were designed to capture approximately 150-bp of the target sequence. The BC primer was designed as follows and targets the RNA sequence: 5'-RS2 sequence (AATGTACAGTATTGCGTTTG) – UMI (12 random nucleotides) – gene specific primer 1 (GSP1) -3'. The LA primer was designed as follows and targets the cDNA sequence: 5'- FS2 sequence (TTCTTAGCGTATTGGAGTCC) – gene specific primer 2 (GSP2)-3'. GSP1 and GSP2 primer sequences were designed using Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) with considerations to minimize off-target amplification, primer-dimer formation, overlap with known SNPs, and efficient mapping of the 150-bp sequencing read using the analysis pipeline outlined below.

Where possible, the GSP1 and GSP2 primer sequences matched or overlapped with regions targeted by the standardized TaqMan-based real-time quantitative PCR assays developed by the ‘Europe Against Cancer’ (EAC)/European LeukemiaNet (ELN) program^{6, 7}.

AML MRD targeted RNA-sequencing library preparation and sequencing

Targeted RNA-sequencing libraries were prepared using the QIAseq Targeted RNA Custom Panel kit (Cat#333025, Qiagen, Germany), with significant modifications. Briefly, 250ng of RNA was treated with ezDNase (Cat# 11766051, Invitrogen, Carlsbad, CA) for 2 minutes (min) at 37°C. UMI assignment and cDNA generation was performed using the SuperScript IV First Strand Synthesis System (Cat# 18091050, Invitrogen, Carlsbad, CA), where the DNase-treated RNA was pre-incubated with 100nM each of BC primer (Supplementary Table 1) at 65°C for 5 minutes and cDNA was synthesized by the SuperScript IV Reverse Transcriptase at 55°C for 10 min. The reaction was inactivated at 80°C for 20 min and residual RNA was removed from by incubating with RNase H at 37°C for 20 min. To ensure complete removal of excess BC primers, the cDNA synthesis reaction was subjected to two rounds of purification with QIAseq beads at 1.3X volume and eluted in 10µL of nuclease-free water. The purified cDNA was then incubated with 100nM each of LA primer (Supplementary Table 1), 600nM RS2 primer (5'-AATGTACAGTATTGCGTTTG-3'), and 6U HotStarTaq DNA Polymerase as follows: 95°C for 15 min, and 8 cycles of 95°C for 15 seconds (sec) and 60°C for 5 min. The DNA was purified with one round of QIAseq beads at 1.6X volume and eluted in 25µL of nuclease-free water. Finally, the purified DNA was amplified with universal adapter

primers for Illumina platforms (QIAseq Targeted RNA 96-Index I, Cat# 333117, Qiagen, Germany) and 12U HotStarTaq DNA Polymerase at the following conditions: 95°C for 15 min, and 26 cycles of 95°C for 15 sec and 60°C for 2 min. The resulting library was purified with one round of QIAseq beads at 1.1X volume and eluted in 25µL of nuclease-free water. The final library was quantitated using the Qubit dsDNA HS Assay kit (Cat# Q32851, Invitrogen, Carlsbad, CA).

Single-end 150bp sequencing was performed on the Illumina MiSeq (v3 reagents) or HiSeq 2500 (rapid run mode) using the QIAseq Read 1 Primer 1 custom primer (Qiagen, Germany) and following the manufacturer's protocol (Illumina, San Diego, CA).

Raw sequencing FASTQ files have been deposited in the NCBI Sequencing Read Archive (SRA) database under accession PRJNA421563.

Bioinformatics pipeline

Raw sequencing FASTQ files were processed using cutadapt (version 1.14)⁸ to remove the 5' and 3' adapter sequences and a custom Python script to remove the 12-mer UMI sequence from the beginning of each read and append the UMI sequence to the read name in the FASTQ file. Subsequently, the trimmed reads were aligned to the human reference genome (GRCh38) using the splice read aligner STAR (version 2.5.2b)⁹. The uniquely aligned and chimeric BAM files were used in the downstream analysis. SAMtools (version 1.3.1)¹⁰ was utilized to cluster the reads per amplicon per library. The

intended primer locations were used as target coordinates. Off-target reads were not considered.

UMI cutoffs were determined for each library based on *ABL1* counts to remove errors which may have arisen during library preparation/sequencing and to minimize background noise. First, the number of times each *ABL1* unique UMI was repeated (n) per library was counted. Then, unique UMIs with counts greater than $\omega = \mu(n) + 2\sigma(n)$, where μ is mean and σ is standard deviation, were removed. Next, the total number of times a unique UMI with (n) count was observed (m) was plotted on a graph, where (n) was plotted on x-axis and $\log_{10}(m)$ was plotted on y-axis. The inflection point (i) (i.e., flattening of the graph curve) of the graph was calculated with (n) in the x-axis and $\log_{10}(m)$ on the y-axis. We considered such an inflection point (i) as the UMI cutoff and only counted unique UMIs with (n) greater than (i) per amplicon. See Supplementary Figure 2 for an example.

The full bioinformatics pipeline is available for download at: www.github.com/NHLBI-BCB/QMRD.

Target panel coverage analysis using TCGA AML dataset

To determine the fraction of patients covered by the targeted RNA-seq panel, patient data from The Cancer Genome Atlas (TCGA) AML cohort¹¹ was analyzed. Only patients for which there was clinical data, mutation analysis, and RNA-seq were included in the analysis ($n=173$). Across the 173 patients, the proportion of patients with *NPM1* exon 12

insertion mutations, *PML-RARA*, *RUNX1-RUNX1T1*, *CBFB-MYH11*, and *BCR-ABL1* fusions were calculated. Utilizing the RNA-seq data, *WT1* or *PRAME* were considered overexpressed in a patient if the expression value was greater than the mean of the entire cohort. The percentage of patients overexpressing *WT1* or *PRAME* was calculated for the entire cohort and for patients who did not have *NPM1* insertion mutations or fusion.

Digital droplet PCR (ddPCR)

CBFB-MYH11 type A expression was determined for the ME-1 cell dilutions by ddPCR. The established EAC assay for *CBFB-MYH11* type A was adapted for use on the Raindance platform (RainDance Technologies, Billerica, MA, USA), as follows. First, 1 μ g of RNA was incubated with ezDNase (Cat# 11766051, Invitrogen) for 2 min at 37°C and cDNA was generated by random priming using the SuperScript IV First-Strand Synthesis System (cat# 18091050, Invitrogen). The final reaction was diluted with water to generate a 20ng/ μ L cDNA equivalent. Custom 40x primer/probe master mixes were synthesized by Integrated DNA Technologies (PrimeTime qPCR probe assays, IDT, Coralville, IA, USA). For detection of *CBFB-MYH11* type A, the primer/probe master mix consisted of a forward primer (5'-CATTAGCACAAACAGGCCTTGA-3'), reverse primer (5'-AGGGCCCGCTTGGACTT-3'), and Taqman probe (5'-6-FAM-TCGCGTGTC/ZEN/CTTCTCCGAGCCT-IABkFQ). For detection of the *ABL1* control gene, the primer/probe master mix consisted of a forward primer (5'-TGGAGATAACACTCTAACGCATAACTAAAGGT-3'), reverse primer (5'-GATGTAGTTGCTTGGGACCCA-3'), and Taqman probe (5'-5HEX-CCATTTTG/ZEN/GTTTGGGCTTCACACCATT-IABkFQ-3'). Amplification of 100ng

cDNA equivalent was performed in 25 μ L reactions consisting of 1x TaqMan Genotyping Master Mix (Thermo Fisher Scientific, Walkersville, MD, USA), 1x Droplet Stabilizer (RainDance Technologies, Lexington, MA), 1x *ABL1* primer/probe master mix, and 1x *CBFB-MYH11* type A primer/probe master mix. Droplets were generated on the RainDrop Source machine (Raindance Technologies) and amplified using the following optimized reaction conditions: 1 cycle of 95°C (2.5C/s ramp) for 10 min; 45 cycles of 95°C (2.5C/s ramp) for 15 sec, 58°C for 15 sec, 60°C for 45 sec; followed by 1 cycle of 98°C (2.5C/s ramp) for 10 min. Droplets were assessed using the RainDrop Sense machine and the data were analyzed with the RainDrop Analyst II v.1.1 software (RainDance Technologies).

ArcherDx Myeloid FusionPlex library preparation and analysis

Anchored multiplex PCR-based enrichment RNA-sequencing libraries were generated using the ArcherDx Myeloid FusionPlex assay for Illumina (cat# AB0037, ArcherDx, Boulder, CO), per manufacturer's instructions. In short, 250ng of RNA from the ME-1 and *NPM1* mutA-positive patient cell dilutions (down to 1:10,000) were subjected to first- and second strand cDNA synthesis. The resulting double-stranded cDNA was subjected to end repair, A-tailing, purification using SPRIselect reagent (Beckman Coulter, cat# B23317), and ligation with a universal ArcherDx molecular barcode adapter, which allows for unidirectional amplification using gene specific primers. Following the molecular barcode ligation, the libraries were subjected to two rounds of nested PCR for target enrichment. The resulting libraries were subjected to paired-end 150bp sequencing on the Illumina Miseq platform (Illumina), per manufacturer's instructions. Raw sequencing

FASTQ files have been deposited in the NCBI Sequencing Read Archive (SRA) database under accession PRJNA421563.

Raw sequencing FASTQ files were analyzed using the Archer Analysis software version 5.1.3, utilizing the default parameters. The downsampling feature was applied to provide consistency across the serial dilutions (1.6M and 2M for the ME-1 and *NPM1* mutA-positive patient serial dilutions, respectively). For the ME-1 dilution, *CBFB-MYH11* (%) is called by the software as the percentage of unique reads spanning the breakpoint and supporting the event divided by the total reads at that locus. *CBFB-MYH11/ABL1* was calculated by dividing the number of unique reads spanning the breakpoint and supporting the event divided by the average number of unique reads aligning to *ABL1* across all of the gene specific primers. For the *NPM1* mutA-positive patient serial dilution, *NPM1* mutA (%) is called by the software as the percentage of the number of alternate observations divided by the sequence coverage at that locus. *NPM1 mutA/ABL1* was calculated by dividing the number of alternate observations divided by the average number of unique reads aligning to *ABL1* across all of the gene specific primers.

Molecular analysis of *CBFB-MYH11* isoform

To confirm the *CBFB-MYH11* isoform present in the bone marrow of patient J (Table 1), 50ng of the final AML MRD targeted RNA-sequencing library generated from patient J was amplified in a mixture containing 0.02 U/ μ L Q5 High-Fidelity DNA Polymerase (cat# M0491, New England Biolabs, Ipswich, MA), 200 μ M 10mM dNTPs, 1X Q5 Reaction Buffer, 0.5 μ M forward primer (5'-CATTAGCACAAACAGGCCTTGA-3'), and 0.5 μ M

reverse primer (5'-AGGGCCCGCTTGGACTT-3'). PCR amplification was performed as follows: 98°C for 30 seconds; 35 cycles of 98°C for 10 seconds, 63°C for 20 seconds, and 72°C for 1 minute; 72°C for 2 minutes. The final PCR product was purified using the DNA Clean & Concentrator kit (Cat# D4013, Zymo Research, Irvine, CA) and visualized on the Agilent 4200 Tapestation (Agilent Technologies, Santa Clara, CA) using the Genomic DNA ScreenTape Assay (cat# 5067-5365, Agilent Technologies). Sanger sequencing was performed by Macrogen, USA (Rockville, MA) on the final purified PCR product using reverse primer (5'-AGGGCCCGCTTGGACTT-3').

SUPPLEMENTARY REFERENCES

1. Konopka JB, Watanabe SM, Witte ON. An alteration of the human c-abl protein in K562 leukemia cells unmasks associated tyrosine kinase activity. *Cell.* 1984;37(3):1035-1042.
2. Clark SS, McLaughlin J, Timmons M, et al. Expression of a distinctive BCR-ABL oncogene in Ph1-positive acute lymphocytic leukemia (ALL). *Science.* 1988;239(4841 Pt 1):775-777.
3. Asou H, Tashiro S, Hamamoto K, Otsuji A, Kita K, Kamada N. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood.* 1991;77(9):2031-2036.
4. Liu P, Tarle SA, Hajra A, et al. Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science.* 1993;261(5124):1041-1044.
5. Lanotte M, Martin-Thouvenin V, Najman S, Balerini P, Valensi F, Berger R. NB4, a maturation inducible cell line with t(15;17) marker isolated from a human acute promyelocytic leukemia (M3). *Blood.* 1991;77(5):1080-1086.
6. Gabert J, Beillard E, van der Velden VH, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program. *Leukemia.* 2003;17(12):2318-2357.
7. Cilloni D, Renneville A, Hermitte F, et al. Real-time quantitative polymerase chain reaction detection of minimal residual disease by standardized WT1 assay to enhance risk stratification in acute myeloid leukemia: a European LeukemiaNet study. *J Clin Oncol.* 2009;27(31):5195-5201.

8. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. 2011;17(1):
9. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012;
10. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-2079.
11. Cancer Genome Atlas Research N, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-2074.

Supplementary Table 1. List of primer sequences.

Target	BC Primer Sequence	LA Primer Sequence
<i>ABL1</i> *	AATGTACAGTATTGCGTTTGNNNNNNNNNCACTCAGACCCTGAGGCTAA	TTCTTAGCGTATTGGAGTCCTGCCTGCCCTGCATTTATCA
<i>WT1</i>	AATGTACAGTATTGCGTTTGNNNNNNNNNCGAGTACTGCTGCTCACCCA	TTCTTAGCGTATTGGAGTCCCCAGCCCGCTATTGCAATC
<i>PRAME</i>	AATGTACAGTATTGCGTTTGNNNNNNNNNCGTGATCCCACACTCATCAA	TTCTTAGCGTATTGGAGTCCCAGCGTCAGTCAGCTAAGTGT
<i>NPM1</i>	AATGTACAGTATTGCGTTTGNNNNNNNNNGCAGATATCAACTGTTACAGAAATG	TTCTTAGCGTATTGGAGTCCGAAGAATTGCTCCGGATGACT
<i>CBFB-MYH11</i>	AATGTACAGTATTGCGTTTGNNNNNNNNNTCTCATCTCCTCCATCTGGGT	TTCTTAGCGTATTGGAGTCCAAGACTGGATGGATGGCTG
<i>RUNX1-RUNX1T1</i>	AATGTACAGTATTGCGTTTGNNNNNNNNNTCAGCCTAGATTGCGTCTCA	TTCTTAGCGTATTGGAGTCCCAGTGTCTCACAAACCCACC
<i>PML-RARA bcr1/2</i> #	AATGTACAGTATTGCGTTTGNNNNNNNNNGCTGTAGATGCGGGGTAGAG	TTCTTAGCGTATTGGAGTCCACCTGGATGGACCGCCTAG
<i>PML-RARA bcr3</i> #	AATGTACAGTATTGCGTTTGNNNNNNNNNGCTGTAGATGCGGGGTAGAG	TTCTTAGCGTATTGGAGTCCCAGTGGCTCGACGAGTT
<i>BCR-ABL1 p190</i> *	AATGTACAGTATTGCGTTTGNNNNNNNNNCACTCAGACCCTGAGGCTAA	TTCTTAGCGTATTGGAGTCCCCTCGCAGAACCTCGAACAG
<i>BCR-ABL1 p210</i> *	AATGTACAGTATTGCGTTTGNNNNNNNNNCACTCAGACCCTGAGGCTAA	TTCTTAGCGTATTGGAGTCCCTGACCAACTCGTGTGAA

*Common BC primer sequences are used to amplify *ABL1*, *BCR-ABL1 p190*, and *BCR-ABL1 p210*

#Common BC primer sequences are used to amplify *PML-RARA bcr1/2* and *PML-RARA bcr3*

Supplementary Table 2. ArcherDx Myeloid FusionPlex assay mutation detection in serial dilution samples.

Cell Dilution	<i>NPM1</i> mutA (%) ^a	<i>CBFB-MYH11</i> (%) ^b
0	47.29	42.2
1:10 ¹	19.62	26
1:10 ²	Not detected	5.1
1:10 ³	Not detected	0.6
1:10 ⁴	Not detected	Not detected

^apercentage of the number of alternate observations divided by the sequence coverage at that locus

^bpercentage of unique reads spanning the breakpoint and supporting the event divided by the total reads at that locus

Supplementary Table 3. Comparison of mutation detection in diagnostic patient samples by AML MRD Targeted RNA-seq and qPCR.

Patient	Sample Type	Annotated Mutation	Mutation Detected by RNA-seq	Mutation Detected by qPCR
A	PB	<i>NPM1</i> mut	<i>NPM1</i> mutA	<i>NPM1</i> mutA
B	BM	<i>NPM1</i> mut	<i>NPM1</i> mutA	<i>NPM1</i> mutA
C	PB	<i>NPM1</i> mut	<i>NPM1</i> mutA	<i>NPM1</i> mutA
D	PB	<i>RUNX1-RUNX1T1</i>	<i>RUNX1-RUNX1T1</i>	<i>RUNX1-RUNX1T1</i>

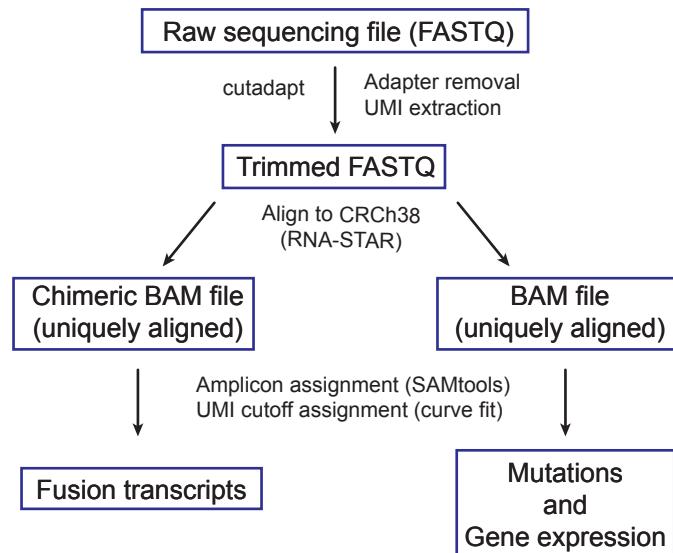
PB = peripheral blood, BM = bone marrow

Supplementary Table 4. Estimation of AML MRD RNA-seq assay cost by platform.

Illumina Platform	MiSeq	NextSeq
Size	150 cycles (v3)	150 cycles (high output)
Maximum reads per run	~25 million	~400 million
Maximum number of samples (fusion transcripts only)	25	400
Maximum number of samples (fusion transcripts and NPM1)	8	133
Approximate sequencing cost per run*	\$991	\$3,060

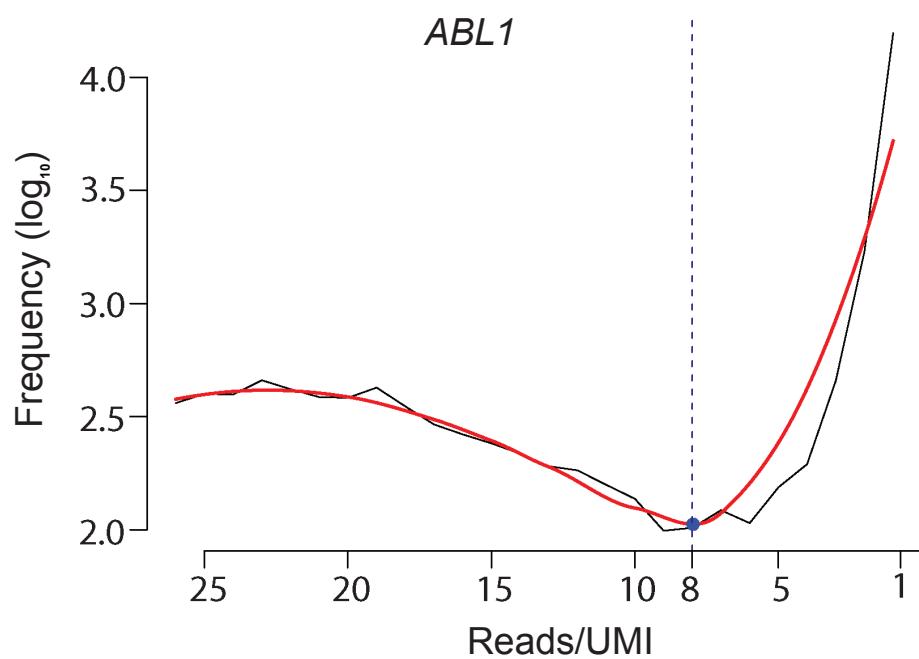
*Sequencing costs will vary depending on local pricing and platform availability and are estimated above for illustrative purposes only. Example prices listed are current US estimates (adapted from: <http://www.biotech.cornell.edu/brc/genomics/services/price-list>, accessed 17th August 2018). Costs associated with library preparation, labor, and computational resources are additional.

Supplementary Figure 1

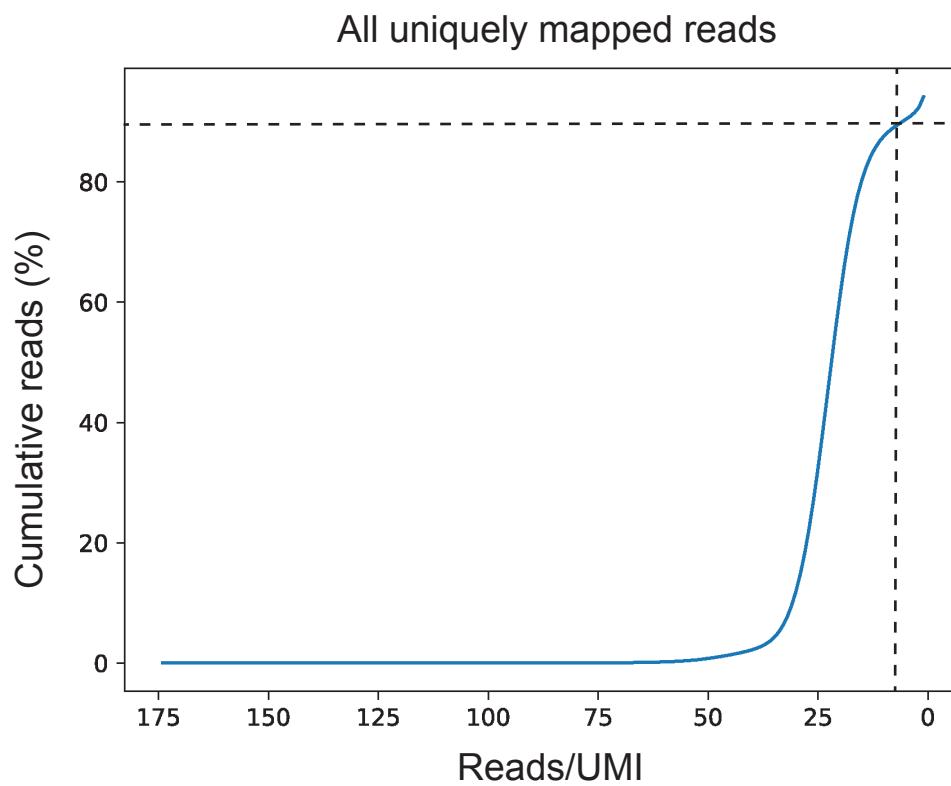


Supplementary Figure 2

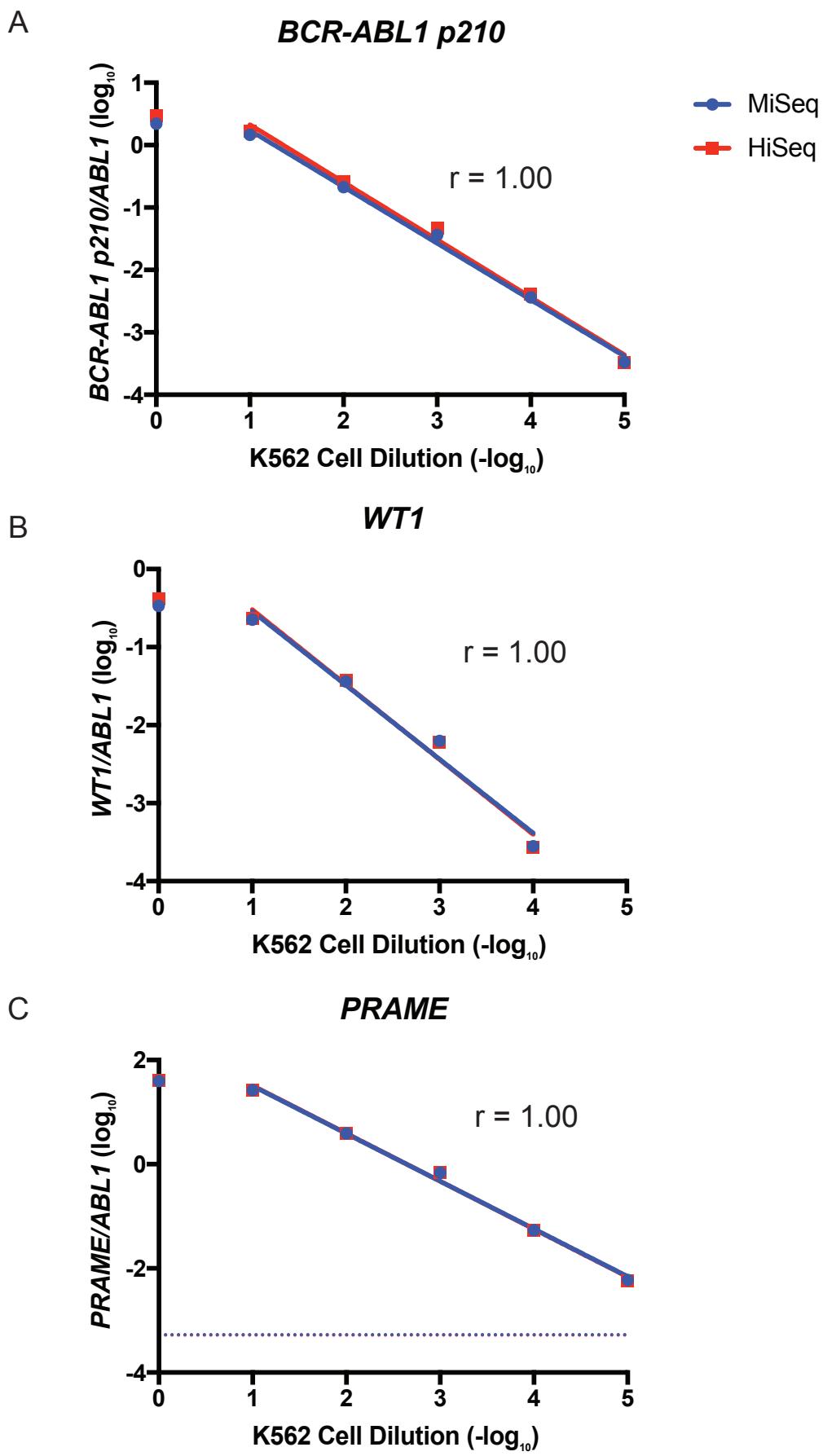
A



B

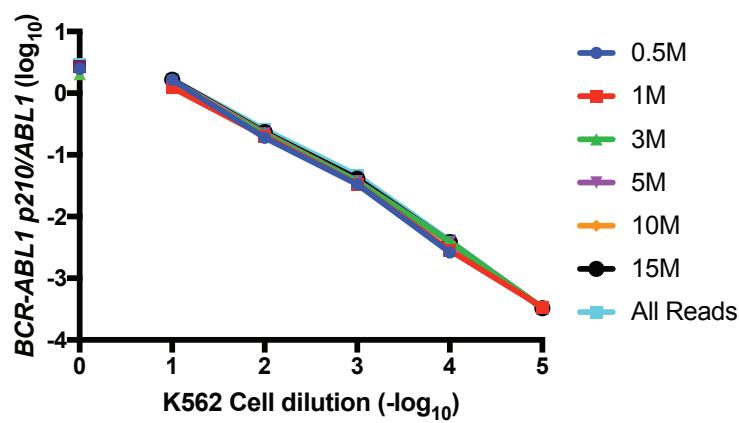


Supplementary Figure 3

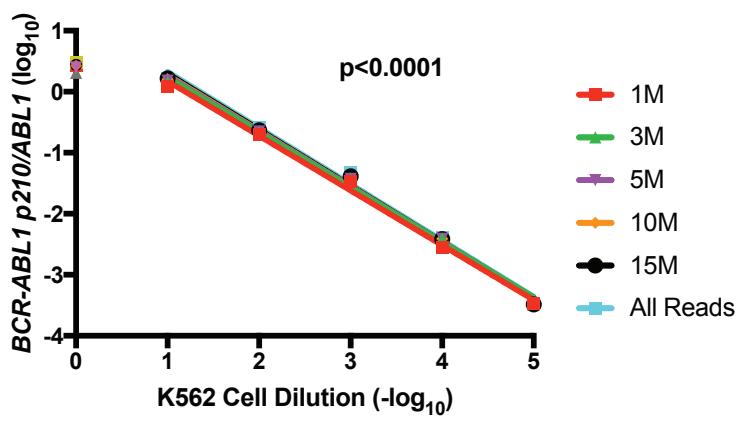


Supplementary Figure 4

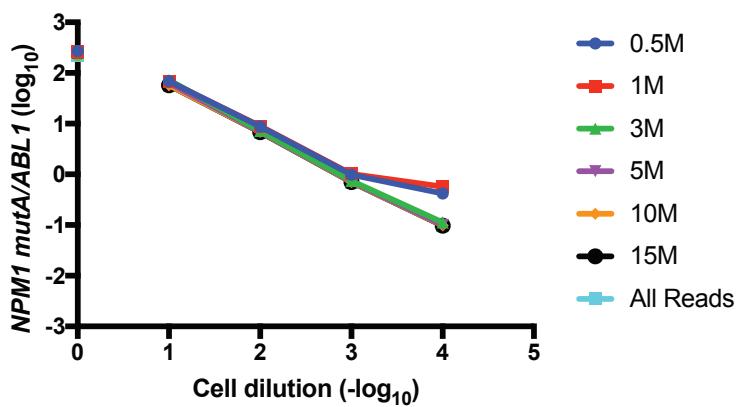
A



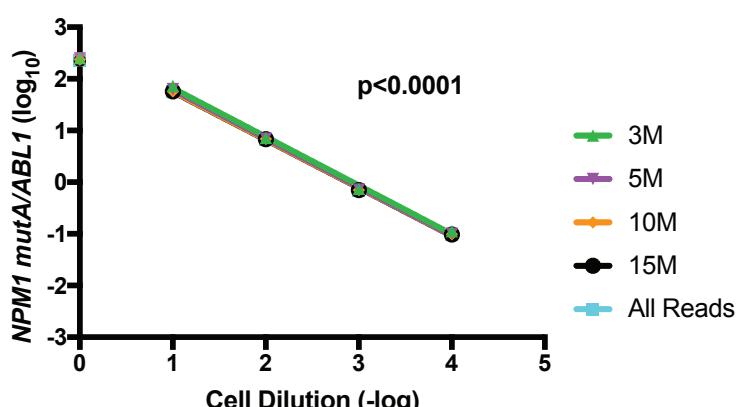
B



C



D



Supplementary Figure 5

