

Evaluation of acute myeloid leukemia using genomic proximity mapping-based next generation cytogenomics

Cecilia C.S. Yeung,^{1,2*} Stephen M. Eacker,^{3*} Olga Sala-Torra,¹ Mary Wood,³ Lan Beppu,¹ David W. Woolston,¹ Ivan Liachko,³ Maika Malig,³ Derek Stirewalt,^{1,4} Alexander Muratov,³ Min Fang^{1,2#} and Jerald Radich^{1,4#}

¹Translational Science and Therapeutics Division, Fred Hutchinson Cancer Center;

²Department of Laboratory Medicine and Pathology, University of Washington Medical Center; ³Phase Genomics and ⁴Department of Medicine, University of Washington Medical Center, Seattle, WA, USA

*CCSY and SME contributed equally as first authors.

#MF and JR contributed equally as senior authors.

Correspondence: C.C.S. Yeung
cyeung@fredhutch.org

Received: June 17, 2025.

Accepted: January 7, 2026.

Early view: January 15, 2026.

<https://doi.org/10.3324/haematol.2025.288461>

©2026 Ferrata Storti Foundation

Published under a CC BY-NC license



Abstract

Cytogenetic analysis encompasses a suite of standard-of-care diagnostic testing methods that is applied routinely in cases of acute myeloid leukemia (AML) to assess chromosomal changes that are clinically relevant for risk classification and treatment decisions. In this study, we assess the use of Genomic Proximity Mapping® (GPM) for cytogenomic analysis of AML diagnostic specimens for detection of cytogenetic risk variants included in the European Leukemia Network (ELN) risk stratification guidelines. Archival patients' samples (N=48) from the Fred Hutchinson Cancer Center Leukemia Bank with historical clinical cytogenetic data were processed for GPM and analyzed with the CytoTerra cloud-based analysis platform. GPM showed 100% concordance for all specific variants that have associated impacts on risk stratification as defined by ELN 2022 criteria and 78% concordance when considering all variants reported by the Cytogenetics Laboratory at Fred Hutchinson Cancer Center. Notably, the percentage of blasts (ranging from 5–96%) did not have a clear effect on the ability to detect these variants. In two cases, GPM identified a recurrent inv(9)(p13.3p13.1). These findings demonstrate GPM's effectiveness for the evaluation of known AML-associated risk variants and a source for biomarker discovery.

Introduction

Cytogenetic diagnostic testing is considered standard-of-care and is routinely applied in acute myeloid leukemia (AML) and other hematologic malignancies.¹ Decades of cytogenetic testing (including karyotyping, fluorescence *in situ* hybridization [FISH], and chromosome genomic array testing [CGAT]) have identified recurrent translocations, inversions, deletions, duplications, and copy-neutral loss of heterozygosity (cnLOH) in AML which drive tumorigenesis through oncogenic fusion genes, disrupted tumor suppressor genes, or amplified oncogenes. These biomarkers are clinically useful for assessing prognosis and guiding treatment.² Each cytogenetic method has specific strengths, but also limitations including low resolution (karyotyping), the need for living, dividing cells (karyotyping), limited scope of detection (FISH), and inability to detect balanced rearrangements (CGAT). These partially overlapping strengths and limitations necessitate costly multi-modal cytogenetic and molecular testing for AML that takes days to weeks to

complete. Nonetheless, current risk classification schemes as well as guidelines used in AML treatment and clinical trial design require cytogenetic findings.

We have developed Genomic Proximity Mapping® (GPM), a next-generation sequencing (NGS)-based assay that uses the proximity of interacting DNA sequences in intact cells to determine the linear structure of chromosomes.^{3,4} The method involves crosslinking DNA within intact nuclei (Figure 1A), extracting then fragmenting the crosslinked DNA, and generating chimeric DNA molecules by ligating DNA molecules that were in close three-dimensional proximity.⁵ Sequences that are closer along a chromosome are more likely to be proximal and interact with each other than sequences that are distant (Figure 1B). Interaction frequencies identify structural variants by detecting changes in expected interaction patterns. When visualized as a heatmap, patterns of pairwise interactions can be used to interpret the structure of structural variants including translocations, inversions, insertions, deletions and duplications (Figure 1C, D).^{6–8} For instance, a balanced translocation is visualized as

an excess of pairwise interactions between two chromosomes in a distinct pattern, as illustrated in Figure 1D. The signal decays moving away from the breakpoint following a power law function⁵ with sharp boundaries in the signal dictated by the sequences involved in the interaction. The inter-chromosomal signal is reciprocal on the heatmap in cases of balanced rearrangements, creating a ‘bowtie’ pattern. In cases of unbalanced rearrangements, only half of this ‘bowtie’ pattern is present, but the pattern on the heatmap retains sharp boundaries and the highest signal intensity is observed at the breakpoint. Similar patterns are observed for inversions but as intra-chromosomal signals. These patterns of long-range sequence interactions form the basis of detection of structural variants with GPM, illustrating a distinct advantage in detecting structural variant breakpoints with high confidence even with lower overall reads as compared to standard whole-genome sequencing. The clinical applicability of GPM analysis to resolve complex chromosome abnormalities in constitutional genetic cases has been published recently.^{9,10} Importantly, GPM can be performed on freshly isolated, frozen, or formalin-fixed, paraffin-embedded patients’ samples making it a broadly applicable cytogenetic analysis platform.^{11,12} In this study, the CytoTerra® Cytogenetics Platform was assessed for the use of GPM as a method for cytogenomic analysis of AML. Using cryopreserved diagnostic specimens, the CytoTerra platform detected all known risk variants included in the European Leukemia Network (ELN) risk stratification guidelines² previously identified by standard-of-care cytogenetics. GPM also identified additional clinically significant variants absent from cytogenetic reports at the time of diagnosis. Furthermore, a recurrent inversion not previously observed in AML was identified in the study population of 48 samples. These observations support the effectiveness of GPM to enhance cytogenomic characterization of AML cases.

Methods

Study population and cytogenetic evaluation

This study conforms with The Code of Ethics of the World Medical Association (Declaration of Helsinki) and was approved by the Fred Hutchinson (FH) Institutional Review Board. Written consent was obtained from all patients. Inclusion criteria included a confirmed diagnosis of AML and consent for whole-genome sequencing. Exclusion criteria included insufficient cryopreserved samples in the FH Leukemia Biobank for orthogonal testing. Samples included various AML-associated genetic abnormalities identified by standard-of-care testing. Data from institutional medical records were verified manually.

CytoTerra library construction and sequencing

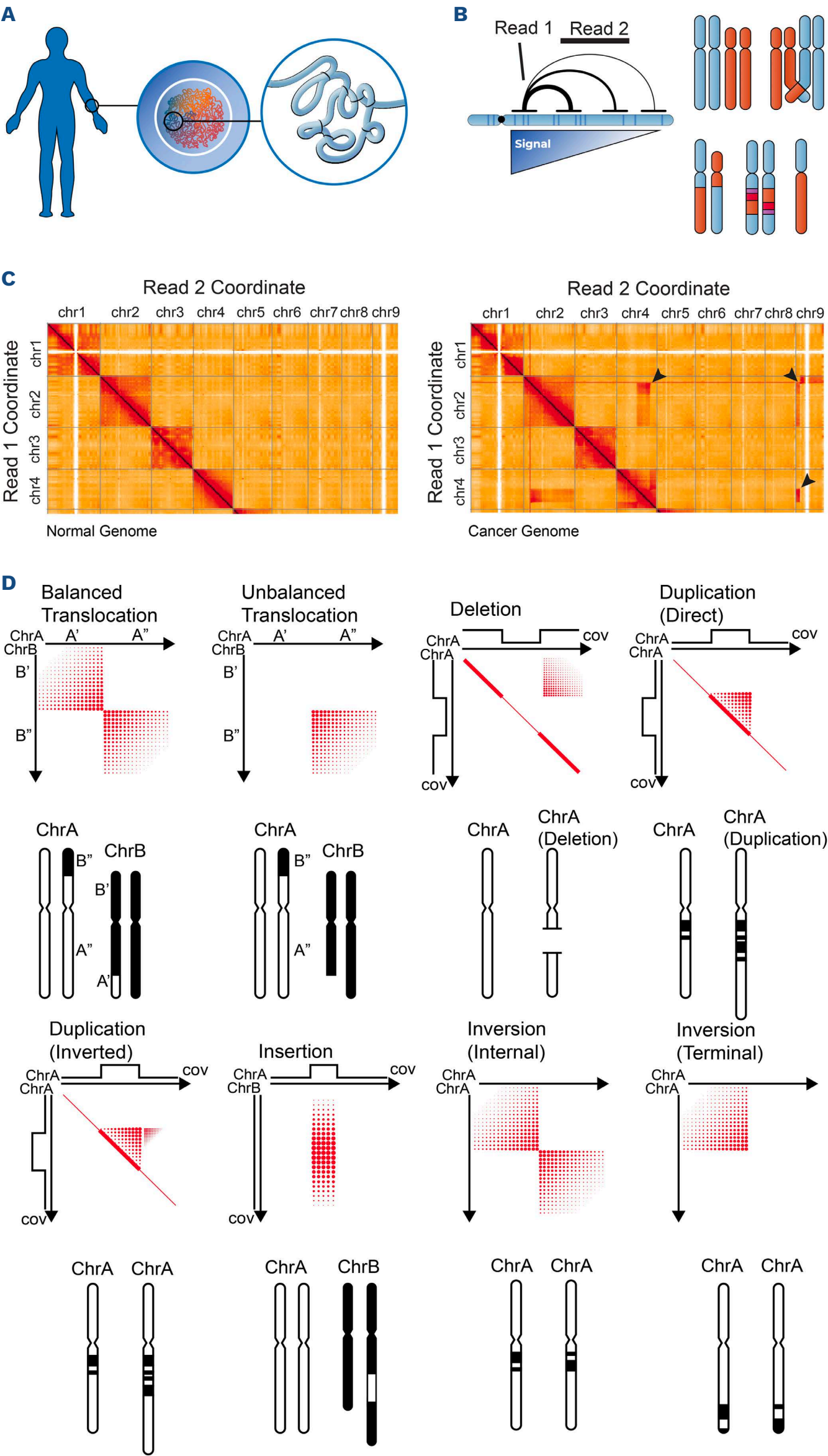
Thawed cryopreserved samples were counted by a hemo-

cytometer then stored in PGShield™ at 4°C. Approximately 200,000–500,000 cells were used for library preparation with the Phase Genomics CytoTerra® kit (v1.1) (Phase Genomics, Inc., Seattle, WA, USA) following the manufacturer’s protocol. In brief, 200,000 – 1 million (M) cells were crosslinked in 1% formaldehyde for 15 minutes. Following quenching, cells were lysed and crosslinked chromatin was immobilized on magnetic beads. DNA was digested using restriction enzymes and the ends were filled in with biotinylated nucleotides. Bead-bound chromatin was subjected to proximity ligation and crosslinks reversed. Proximity ligated junctions were enriched using streptavidin magnetic beads. Streptavidin bead-bound fragments were used to generate an Illumina® sequencing library sequenced on a NovaSeq™ 6000 (Illumina, Inc., San Diego, CA, USA) in the paired-end 150 bp format to an average depth of 150 M read pairs. Further details about the Illumina NGS methods are provided in the *Online Supplementary Methods*. Library performance was evaluated using Phase Genomics’ open-source quality control tool hic_qc.¹³ Libraries passing primary quality control metrics were advanced to analysis and include same-strand read pairs (>20%), high-quality reads (uniquely mapping sequences, >50%), and duplicate read pairs (<35%).

CytoTerra data analysis

Raw FASTQ files were analyzed with the CytoTerra cloud-based analysis platform. First, reads were aligned to the reference genome GRCh38 using bwa mem v0.7.18;¹⁴ duplicate reads were marked with samblaster v0.1.26; and after gathering quality control statistics, duplicate reads were removed using matlock. In-house scripts filtered out low-quality alignments and calculated coverage. The Pairix v0.3.7¹⁵ and Cooler v0.10.2¹⁶ packages generated the contact frequency matrix. Small variants were predicted from filtered alignments using a convolutional neural network, and an in-house script extracted allele frequency data. The proprietary CytoTerra analytic suite is composed of: (i) a convolutional neural network trained to detect variants in heatmap images generated from the contact frequency matrix. The neural network generates a probability confidence value for each call; (ii) a support vector machine caller that predicts aneuploidy from coverage and allele frequency data and produces a probability confidence value for each variant; (iii) a bioinformatics tool specialized in breakpoint detection from GPM alignments which uses a log-odds method to produce a confidence value; and (iv) a trio of bioinformatics tools specialized in detecting copy number and cnLOH based on alignment data and allele frequency from small variants with each reporting an Expect value to judge confidence.

After variant predictions were made, putative false positives were flagged using a rules-based approach reliant on call confidence statistics and a convolutional neural network trained in-house to classify calls based on heatmap image



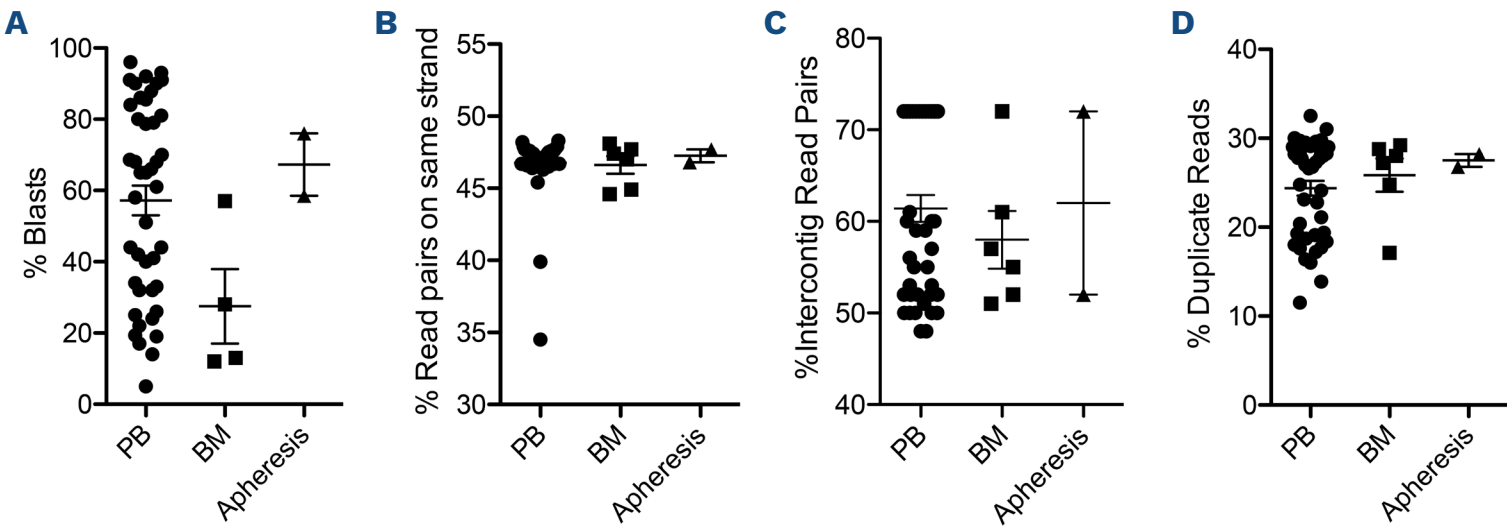


Figure 2. Blast counts and library parameters for samples used in this study. (A) Blast percentage estimates for peripheral blood, bone marrow aspirates, or apheresis-derived samples used for Genomic Proximity Mapping® library construction. (B–D) Quality control metrics for libraries generated by sample type. (B) ‘Reads on same strand’ measures the percentage of read pairs from library inserts that map to the same strand of the human reference genome and are therefore the product of a proximity ligation event. Because reads can be proximity ligated to either the same or different strand configuration, the same strand percentage multiplied by two gives an estimate of the fraction of library fragments derived from proximity ligation events. (C) ‘Inter-contig mapping read pairs’ measure the fraction of read pairs likely to be derived by spurious ligation. (D) Duplicate reads are the fraction of reads that are the result of either polymerase chain reaction or optical duplication. PB: peripheral blood; BM: bone marrow.

data. The CytoTerra Curator platform was used for variant analysis, annotation, and visualization. Calls were reviewed by a blinded operator. Variants >100,000 bp were reported except cnLOH; while CytoTerra’s resolution for cnLOH was 1 Mb, a 10 Mb threshold was employed for likely somatic cnLOH events.

Results

Clinical acute myeloid leukemia sample performance in proximity ligation sequencing

Cryopreserved AML samples were identified from the FH AML sample repository. The study population was composed of 48 diagnostic samples derived from peripheral blood, bone marrow aspirates, and apheresis samples. Details of the patients with therapy information are provided in Table 1. Blast percentage estimates varied between 5% and 96% (Figure 2A). GPM libraries generated from 200,000–500,000 cells were successful for all samples. Post-sequencing quality control analysis of the resulting libraries showed acceptable performance for all libraries (Figure 2B–D, *Online Supplementary Table S1*). Two normal bone marrow samples and a diverse set of normal lymphoblastic cell lines with no structural variants that met reporting criteria¹⁷ were used as controls.

Detection of interchromosomal and intrachromosomal rearrangements

Based on previous cytogenetic risk assessments, the blinded CytoTerra assessment identified all ELN-identified translocations in the study population. These include the t(8;21) (q22;q22.1) translocation that generates the *RUNX1::RUNXT1* fusion (Figure 3A). Beyond the ELN-classified translocations

Table 1. Clinical demographics.

Parameters	Values
Total patients, N	48
Sex, N (%)	
Male	28 (58)
Female	20 (42)
Age, years, mean (range)	54 (21-84)
GPM specimen source, N (%)	
Peripheral blood	42 (88)
Bone marrow	4 (8)
Apheresis	2 (4)
Chemo/TX regimens, N (%)	
Incomplete data	3 (6)
1	31 (65)
2-4	12 (25)
>4	2 (4)
Transplant, N (%)	
Yes	16 (34)
No	32 (66)

GPM: Genomic Proximity Mapping®; TX: treatment.

identified, several additional rearrangements of known and unknown significance were observed in the CytoTerra analysis. These variants included an *NUP98::KDM5A* fusion created by a t(11;12)(p15.4;p13.33) translocation, a variant associated with poor prognosis and chemoresistance (Figure 3D).¹⁸ This rearrangement was not previously detected in the cytogenetics report for this patient. Though not previously reported in the literature, an identified unbalanced t(6;7)(p23;q36.3) translocation demonstrated a small region of non-reciprocal exchange resulting in a deletion of the *JARID2* gene, a known tumor suppressor in myeloid neo-

plasms (Figure 3E).¹⁹ The remaining variants identified were of no known significance (*Online Supplementary Table S2*). We observed a single complex karyotype in this study population that included six interchromosomal breakpoints, in addition to a -7 and dup(21)(q22.12q22.12) (Figure 4C) among other less complex karyotypes (Figure 4A, B).

Among this dataset, inv(16)(p13.1q22) is the most common, observed in 4-5% of AML patients who undergo cytogenetic work-ups.^{20,21} CytoTerra identified all four inv(16) rearrangements observed by cytogenetics in the study population (Figure 5A). The next most common class of inversions observed are related rearrangements involving the *MECOM* (*EVI1*) locus.²¹ In this study, both cytogenetics and CytoTerra identified a single instance of an inv(3)(q21.3q26.1) (Figure 5B). However, only CytoTerra identified an inv(3)(p24.3q26.2), an unusual but previously documented rearrangement.²² Both

of these rearrangements are classified as adverse by the ELN 2022 classification.² We also observed an instance of an inv(12)(p13.32p13.2), a variant that has not been previously documented (Figure 5D).

Among the 48 cases in this set, we observed a recurrent inversion not previously documented in AML in two cases. The inv(9)(p13.3p13.1) was seen in two unrelated cases and occurs between two paralogous genes *ANKRD18A* and *ANKRD18B* (Figure 5C). These genes lie within regions of segmental duplication in the pericentromeric region of chromosome 9 at approximately 5 Mbp from each other.

Detection of insertions

In this study population, we identified two insertions not reported by conventional cytogenetic analysis (*Online Supplementary Tables S3 and S4*). In one case, 8 Mbp of chromo-

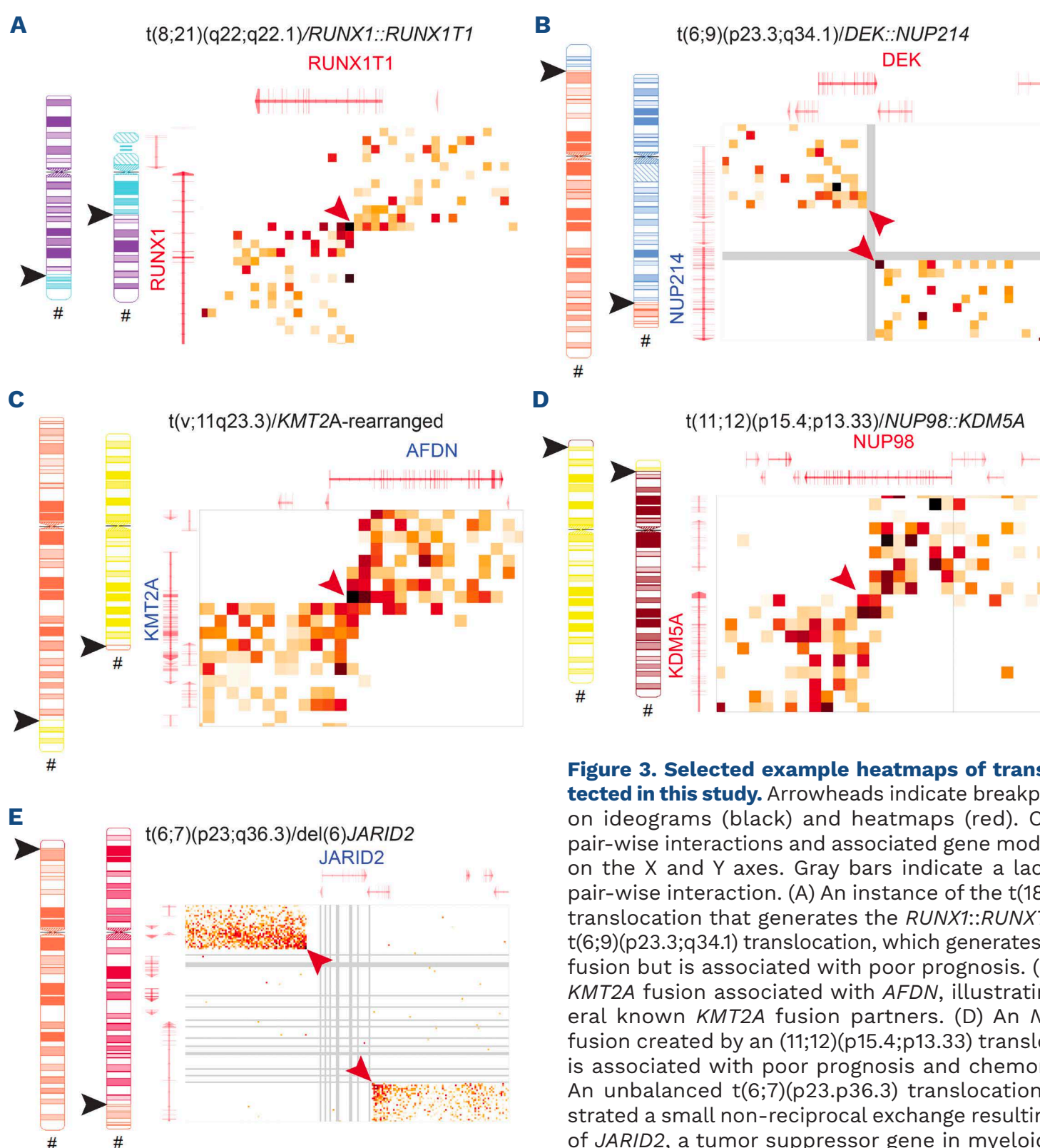


Figure 3. Selected example heatmaps of translocations detected in this study. Arrowheads indicate breakpoints observed on ideograms (black) and heatmaps (red). Coordinates of pair-wise interactions and associated gene models are labeled on the X and Y axes. Gray bars indicate a lack of detected pair-wise interaction. (A) An instance of the t(18;21)(q22;q22.1) translocation that generates the *RUNX1::RUNX1T1* fusion. (B) A t(6;9)(p23.3;q34.1) translocation, which generates a *DEK::NUP214* fusion but is associated with poor prognosis. (C) An example *KMT2A* fusion associated with *AFDN*, illustrating one of several known *KMT2A* fusion partners. (D) An *NUP98::KDM5A* fusion created by an (11;12)(p15.4;p13.33) translocation, which is associated with poor prognosis and chemoresistance. (E) An unbalanced t(6;7)(p23;p36.3) translocation that demonstrated a small non-reciprocal exchange resulting in a deletion of *JARID2*, a tumor suppressor gene in myeloid neoplasms.

some 5 was inserted into chromosome 13 150 kbp upstream of *FLT3* and disrupting the *PAN3* gene (Figure 6A). In another, 120 kbp of chromosome 12 was inserted into another site on chromosome 12 (Figure 6B). This small insertion is copy-neutral but does disrupt *DDX11*, which when mutated is associated with negative outcomes in AML.²³

Detection of copy number aberrations

In the current study population, only two observations of -7 were made by CytoTerra. Numerous aneuploidies and smaller

deletions and duplications of unknown significance were detected by CytoTerra (Online Supplementary Tables S4 and S5). Among these were two instances of deletions of the *TET2* gene not previously detected in cytogenetic reporting. In the study population, a single instance of cnLOH was detected on chromosome 13 by both CGAT and CytoTerra (Online Supplementary Table S6).

Concordance between cytogenetics and CytoTerra

Blinded reviews of variant calls generated by the CytoTerra

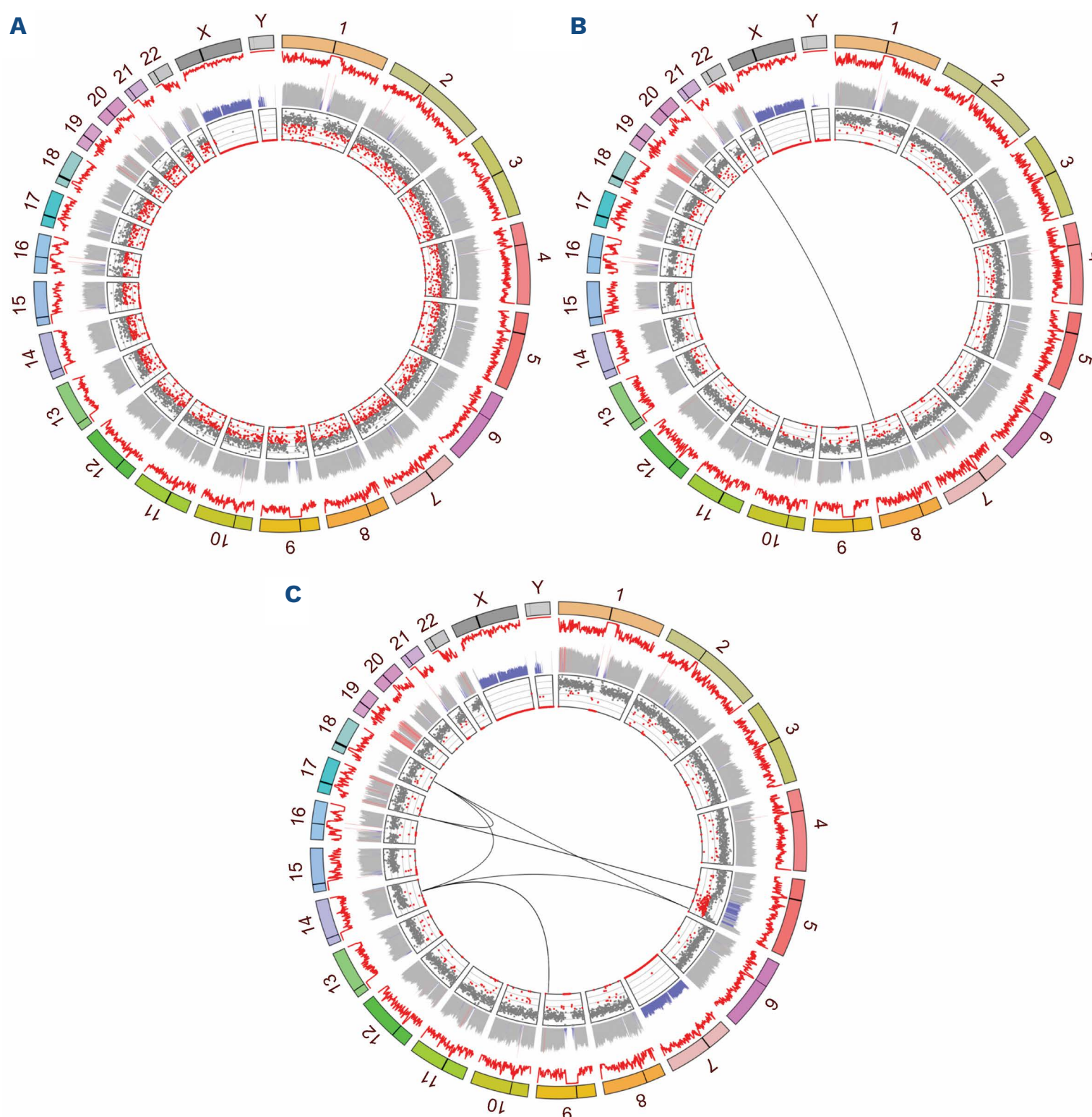


Figure 4. Circos plots illustrating the range of complexity observed in this study. (A) A sample from a patient with a normal karyotype (46, XY). (B) A patient presenting with 45,X,-Y,t(8;21)(q22;q22.1). (C) A complex series of genomic rearrangements uncovered using Genomic Proximity Mapping®. The inferred International System for Cytogenomic Nomenclature for this case: 45,XY,del(5)(q22q35),der(5)t(5;17)(q14.3;p13.3),-7,der(9)t(9;14)(q33.3;q23.1),der(14)del(14)(q22.2q23.1)t(9;14),der(17)t(17;18)(p13.3;q21.1),der(18)(18pter->18q21.1::14q23.1::5q35.1->5qter). Outer ring: chromosomes represented as colored boxes, the black bar illustrates the location of the centromere. Middle rings: red line illustrates raw coverage with inferred copy number illustrated as bars below. Gray = copy 2, blue = copy 1, red > copy 2. Inner ring: minor allele frequency (MAF). Gray dots indicate the expected MAF for copy 2 while red dots indicate a deviation from expected frequency.

platform were compared with the clinical cytogenetics records. CytoTerra showed 100% concordance for all specified variants that have associated impacts on risk stratification as defined by ELN 2022 criteria (Table 2). Notably, blast percentage did not have a clear effect on the ability to detect these variants, with the percentage of blasts ranging between 5% and 96%. When considering variants that meet the ELN categorization of “cytogenetic and/or molecular abnormalities not classified as favorable or adverse”, CytoTerra demonstrated a 77.8% concordance rate with cytogenetics (Table 2). A majority (6/10) of discordant calls were aneuploidies (two +8, two +22, and one case of 4N, tetraploidy). Nineteen copy number and structural variant calls were discordant between GPM and

clinical cytogenetics in 12 unique samples. These underwent orthogonal testing with whole-genome sequencing to corroborate their GPM or cytogenetic presentation. Fourteen GPM calls were confirmed over their cytogenetic presentation, four cytogenetic calls were corroborated over their GPM presentation, and there was one case with full-genome tetraploidy which neither GPM nor NGS can detect (*Online Supplementary Table S7*).

Discussion

This study shows the ability of whole-genome sequencing with GPM to detect cytogenetic aberrations including trans-

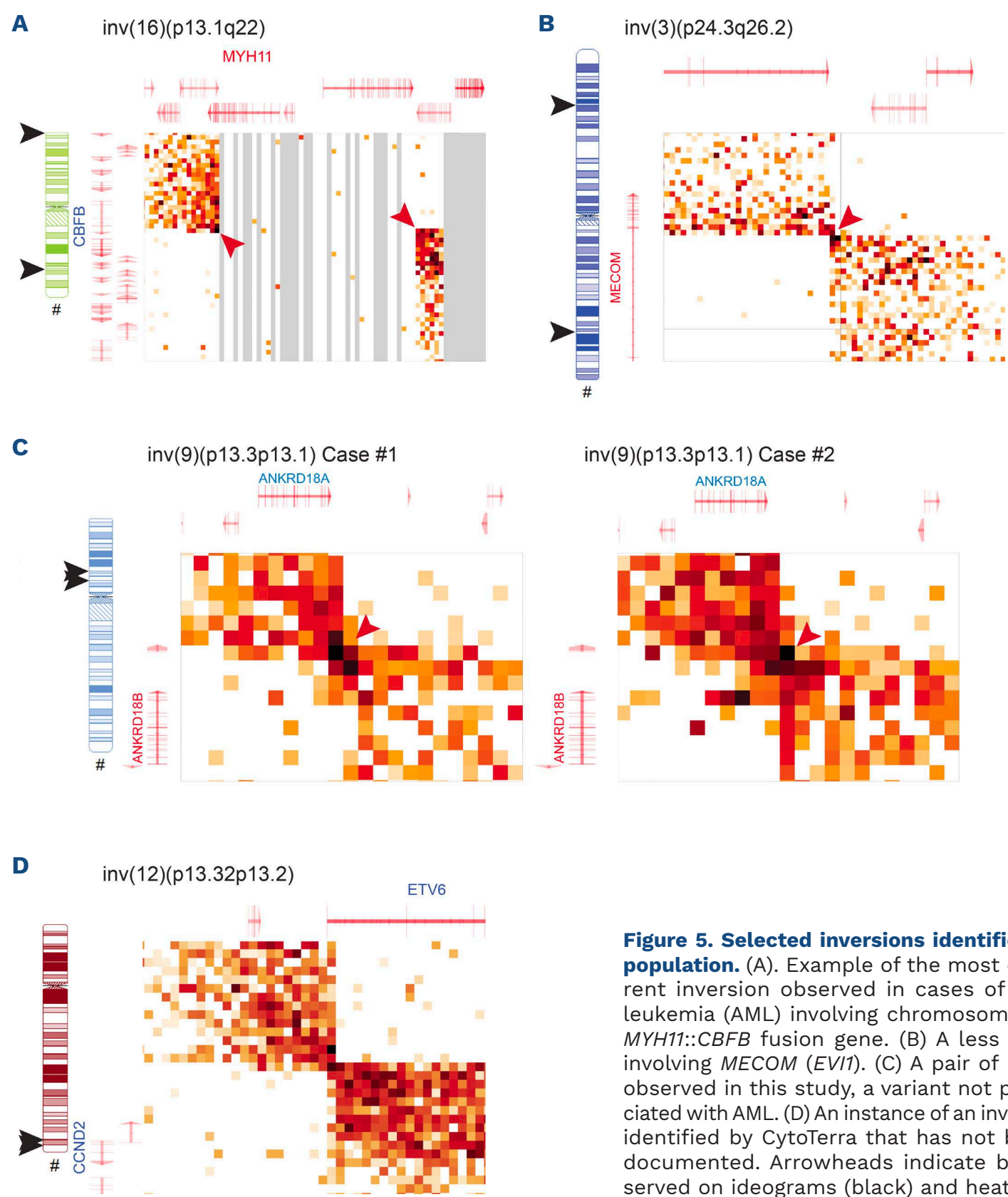


Figure 5. Selected inversions identified in the study population. (A). Example of the most common recurrent inversion observed in cases of acute myeloid leukemia (AML) involving chromosome 16, creating a *MYH11::CBFβ* fusion gene. (B) A less common inv(3) involving *MECOM* (*EVI1*). (C) A pair of recurrent inv(9) observed in this study, a variant not previously associated with AML. (D) An instance of an inv(12)(p13.32p13.2), identified by CytoTerra that has not been previously documented. Arrowheads indicate breakpoints observed on ideograms (black) and heatmaps (red).

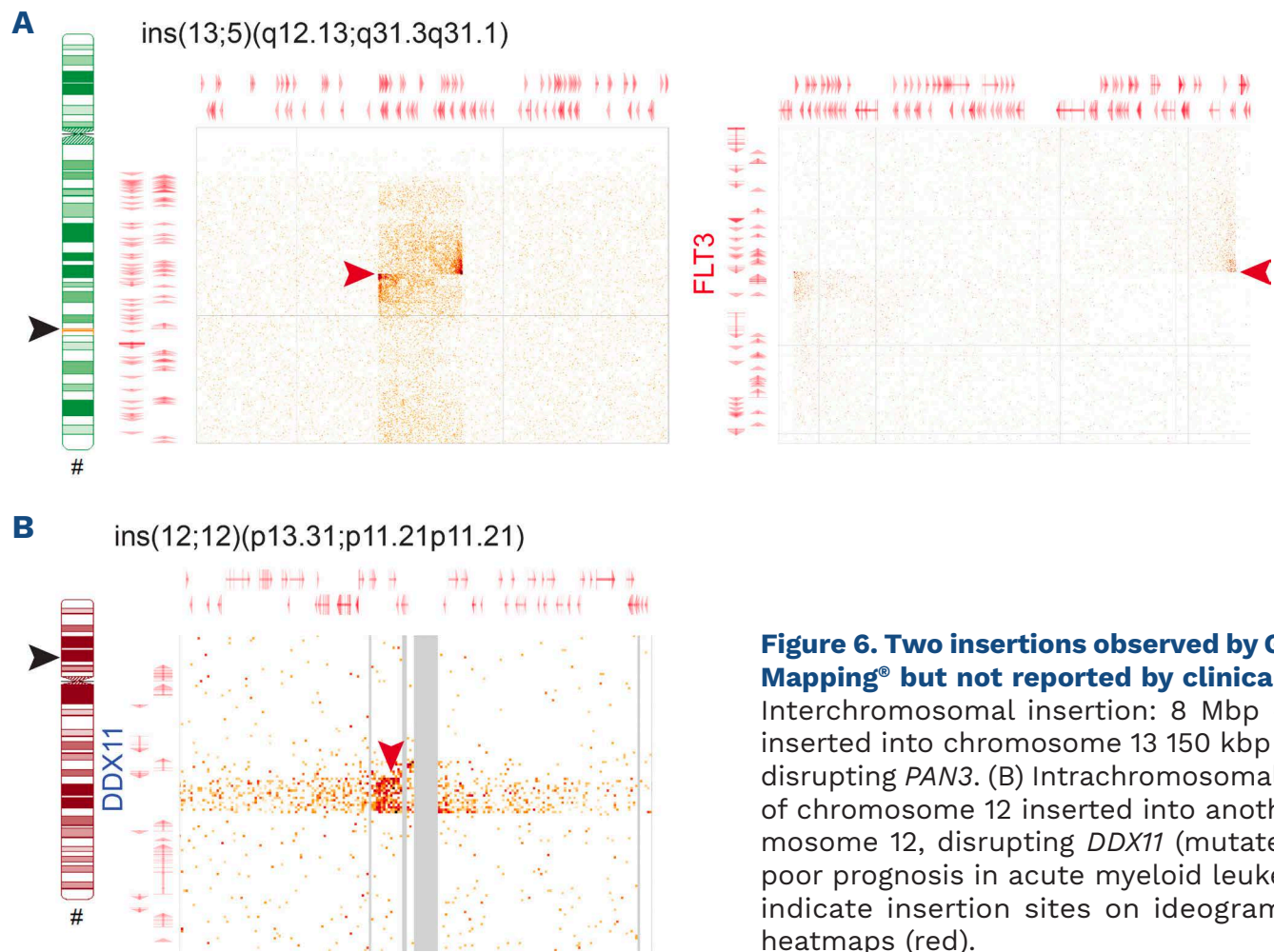


Figure 6. Two insertions observed by Genomic Proximity Mapping® but not reported by clinical cytogenetics. (A) Interchromosomal insertion: 8 Mbp of chromosome 5 inserted into chromosome 13 150 kbp upstream of *FLT3*, disrupting *PAN3*. (B) Intrachromosomal insertion: 120 kbp of chromosome 12 inserted into another region in chromosome 12, disrupting *DDX11* (mutated *DDX11* portends poor prognosis in acute myeloid leukemia). Arrowheads indicate insertion sites on ideograms (black) and on heatmaps (red).

locations, inversions, copy number alterations, insertions, and deletions, as well as cnLOH. Over an initial set of AML patients representing the full range of ELN prognostic risk categories, we challenged the GPM assay to recapitulate the prognostic data produced by archival clinical grade molecular and cytogenetic assays. Cytogenetics encompasses at least three major technologies that are used for diagnostic and research purposes: karyotyping, FISH, and CGAT.¹ Each technology has specific strengths and limitations, necessitating the use of all three technologies to achieve a comprehensive assessment of genomic alterations. While karyotyping and CGAT both offer genome-wide assessment, karyotyping has limited resolution and CGAT is unable to identify balanced rearrangements. These limitations are likely why variants such as *NUP98::KDM5A* – a balanced rearrangement involving a submicroscopic (~3 Mbp) sequence – eluded detection in the initial cytogenetic workup. While FISH can readily identify rearrangements, such as t(11;12), that lead to the *NUP98::KDM5A* fusion,²⁴ it has practical limitations for the number and variety of FISH probes that can be applied to any one diagnostic sample. GPM offers an alternative that captures balanced and unbalanced alterations in a genome-wide manner, including those involving small (<100,000 bp) amounts of sequence. GPM is one of several technologies being applied for the detection of chromosomal abnormalities for clinical research. Whole-genome sequencing using either short reads (Illumina, Element, Ultima, etc.) or long reads (PacBio, Oxford Nanopore) is being increasingly applied in clinical settings.

Table 2. Summary of concordance between Genomic Proximity Mapping® and standard-of-care cytogenetics.

Variant class	Concordant	Discordant	Added
Specified ELN risk variants, N			
Translocation	3	0	0
Inversion	6	0	0
Copy number variant	0	0	0
Aneuploidy	3	0	0
Overall, N (%)	12 (100)	0 (0)	0
Cytogenetic and/or molecular abnormalities not classified as favorable or adverse			
Translocation	7	1	7
Inversion	7	0	2
Insertion	0	0	2
Copy number variant	5	3	29
Copy-neutral LOH	1	0	0
Aneuploidy	15	6	0
Overall, N (%)	35 (77.8)	10 (22.2)	38

ELN: European Leukemia Network; LOH: loss of heterozygosity. Chromoseq™,²⁵ a short-read whole genome sequencing assay, is currently offered through Washington University, St. Louis Department of Pathology. Although it is based on

whole-genome sequencing, the assay only reports a defined list of recurrent structural variants, likely due to the limitations of short-read sequencing to identify breakpoints produced by inversions and translocations. Sequencing-based detection of rearrangements requires that a set of reads map to the junction between the sequences participating in the rearrangement, thus demanding very high coverage data. If, as is often the case, a rearrangement is mediated by a repetitive element,²⁶⁻²⁸ it is usually impossible for short-read sequencing to span the rearrangement junction, making these aberrations undetectable. Current estimates place the sensitivity of standard short-read sequencing for large aberration detection between 10%²⁹ and 70%³⁰ with an exceptionally high false-positive rate of up to 89%.³⁰⁻³³ GPM overcomes this limitation of short-read sequencing by capturing ultra-long-range sequence information at relatively low sequencing depth. For instance, Chromoseq™ relies on ~1 billion read pairs of Illumina® sequencing whereas GPM recommends only 150 M read pairs of data. Long-read sequencing technologies have made enormous progress by increasing accuracy, throughput, and reducing costs in recent years.^{34,35} Long reads can help overcome the ambiguity in breakpoint identification due to the extended sequence context, and an explosion in variant-calling algorithms supporting this activity has accompanied improvements in the technology. Despite these efforts, long-read sequencing deployment to the clinic is limited by the technical challenges of large amounts of high-molecular weight DNA needed for library construction and the cost associated with sequencing, especially in the case of somatic disease for which detection of clonal variants is a requirement.³⁶

Optical genome mapping (OGM) is another high-molecular weight DNA-based technology that can be used to identify structural variants.³⁷ Rather than sequencing DNA, OGM uses sequence-specific DNA labeling to generate long DNA fragments which are electrophoresed through a capillary channel. The distance between labeled sequences on the large fragments is used to map the order and orientation of genomic sequences, analogous to restriction fragment length polymorphism mapping. OGM has been applied in a systematic evaluation of AML genomes and it was demonstrated that, like GPM, OGM detected known variants of significance with a high degree of sensitivity.³⁸ OGM struggled to identify cases of trisomy 8 (6/9 identified),³⁸ similar to what was observed with GPM in this study (4/6 identified) (*Online Supplementary Table S5*). This may be in part attributed to the sensitivity of these methods to detect whole chromosome aneuploidy or bias in outgrowth of abnormal myeloid clones during culture for karyotype analysis. Identification of cryptic rearrangements, similar to those identified in this study, has been a highlight of more recent studies supporting the utility of OGM for resolving structural variants in AML patients' samples.^{12,39,40} A non-trivial challenge to overcome in these OGM-based

studies is the isolation of high-molecular weight DNA from limited amounts of patients' samples. By contrast, GPM operates on approximately one-tenth the input of that necessary for OGM and is compatible with a wider variety of sample types including formalin-fixed paraffin-embedded tissue.^{10,11} GPM also has the distinct advantage of running on ubiquitous short-read sequencing platforms that have established a foothold in diagnostic laboratories, thereby increasing its accessibility.

The ELN 2022 risk classification guidelines identify a set of recurrent translocations seen in AML. Though less commonly observed than translocations, a number of recurrent inversions are known contributors to the AML phenotype. The t(8;21) translocation which results in the *RUNX1::RUNX1* fusion is associated with favorable outcomes and may be an important biomarker for treatment with CDK4/6 inhibitors.⁴¹ The inv(16)(p13.1q22) inversion is the most common intra-chromosomal rearrangement and generates a *CBFB::MYH11* fusion transcript, a rearrangement that portends a more favorable prognosis. The t(6;9)(p23.3;q34.1) translocation similarly generates a *DEK::NUP214* fusion transcript but is associated with poor prognosis (Figure 3B).⁴² Members of the nucleoporin gene family, including *NUP98* and *NUP214*, are known to drive AML through a variety of different partner genes that are detectable using the GPM approach. Like the nucleoporin genes, *KMT2A* is known to associate with a variety of fusion partners,⁴³ including the *AFDN* gene observed in this study population (Figure 3C). While standard-of-care cytogenetics identified both *KMT2A* and *NUP214* fusions, only CytoTerra identified the *NUP98* fusion, a variant associated with poor outcomes.¹⁸ In this case, CytoTerra offers benefit in identifying variants of importance over existing standard methods.

Detecting insertions by traditional cytogenetics is dependent on the size and genomic content of the inserted DNA segment. Some insertions are routinely detected over the course of AML diagnosis (e.g., *FLT3*-ITD), but most of the variants are anonymous and are of unknown significance. In this study, GPM detected two insertions that were not detected by cytogenetics: one 8 Mbp and one 120 kbp in length. The resolution of GPM enables more thorough description of the insertions identified in this study, both of which are associated with genes (*FLT3*, *DDX11*) of known clinical importance in AML. Interestingly, a recurrent inversion was observed in two cases involving inv(9)(p13.3p13.1) between two paralogous genes *ANKRD18A* and *ANKRD18B* spanning a 5 Mb region, lying in a pericentromeric region and involving a sub-microscopic interval of the genome. This inversion has not been previously documented in AML genomes. Rearrangements such as this can be challenging to detect by cytogenetics and standard short-read sequencing methods. However, this variant has been observed previously in a number of cases of acute lymphoblastic leukemia,⁴⁴ detected only through a targeted resequencing effort of this pericentromeric region. These cytogenetically

cryptic classes of rearrangement represent a relevant class of variant for which CytoTerra shows promise of making a clinical impact.

Unsurprisingly, most additional variants called by CytoTerra involved copy number changes below the level of cytogenetic resolution (<5 Mb). They may also reflect changes deemed unreportable from a clinical perspective as many of these variants are of unknown significance, including potentially constitutional variants. However, as noted above, CytoTerra identified additional variants of known significance not specified in the ELN risk criteria, including *NUP98::KDM5A*. This, and six other translocations, and two additional inversions were uncovered by CytoTerra in this study. Previous cytogenetic analysis likely failed to observe these rearrangements because of the genomic location and/or size of the genomic interval involved. For example, *NUP98* lies at the 11p terminus and has multiple oncogenic partners, making it particularly challenging to detect.

Some limitations of GPM must be mentioned. Relatively few copy number aberrations are considered informative for risk stratification by the ELN 2022 guidelines. Copy number aberrations can also be included under the catch-all category of ‘cytogenetic and/or molecular abnormalities not classified as favorable or adverse’, which imparts intermediate risk. Other specific abnormalities are cited in the ELN guidelines including -5, del(5q), -7, -17 (or -17p), all of which are associated with adverse risk. One of GPM’s limitations is its relatively lower sensitivity in detecting copy number aberrations. This limitation may be overcome by increased depth of sequencing, allowing for higher confidence detection of subtle changes of minor allele frequency. This highlights the challenge of detecting mosaic changes in whole chromosome copy number in all sequencing-based coverage data. Tetraploidy also represents a challenge because, in the case of whole genome duplication, the allele frequency remains in balance and is undetectable by sequencing, array, or OGM. One patient demonstrated a cnLOH of 13q, a variant frequently observed in AML with *FLT3*-ITD mutations.^{45,46} This singular cnLOH variant was confirmed by CGAT but more extensive studies will be necessary to determine GPM’s performance detecting this class of structural variant. The limit of detection has yet to be systematically determined but anecdotal detection of variants in a 5% myeloblast sample were observed in this study. In a constitutional genetics study, a variant previously estimated to be found in 7% of the patient’s

peripheral blood was successfully identified, consistent with a sub-10% abundance limit of detection.¹⁰

In conclusion, this study demonstrates GPM’s capability to comprehensively interrogate the entire genome including detection of cryptic chromosomal aberrations at a higher resolution than conventional karyotyping and CGAT. The identification of a novel recurrent AML variant in this 48-sample study demonstrates the potential of GPM as a tool for biomarker discovery. The improved detection of ELN risk variants with GPM warrants a comprehensive study to evaluate CytoTerra for improved accuracy in the risk stratification of patients.

Disclosures

IL, MM and SME are employees of Phase Genomics, Inc., a company developing the GPM technology. CCSY consulted for TwinStrand Bioscience; this relationship has ended.

Contributions

CCSY, SME, IL, MF, JR and OS-T conceived the study. OS-T, LB and DS curated samples. LB and MM prepared and tested samples. CCSY, SME, OST, MW, DWW, IL, MM, DS, AM, MF and JR interpreted or analyzed data. CCSY, SME, OS-T, MW, LB, DWW, IL, MM, DS, AM, MF and JR prepared the manuscript. MF and JR revised the manuscript for important intellectual content. CCSY and SME provided supervision.

Acknowledgments

The authors thank the Genomics Core at Fred Hutchinson Cancer Center, especially Andy Marty and Alex Zevin, PhD, for the preparation and analysis of the whole-genome sequencing Illumina libraries.

Funding

This work was supported by an SBIR phase II grant from NCI/NIH R44CA278140 to SME and Phase Genomics. This work is also partially supported by UG1 CA233338-02 and CA175008-06 to JR at Fred Hutchinson Cancer Center.

Data-sharing statement

This is an active project with ongoing enrollment of subjects for the study, so the data will be made publicly available at a later date when the study is ended. However, data presented in this study will be made available upon request to the corresponding author. Please contact cyeung@fredhutch.org.

References

1. Wan TSK. Cancer cytogenetics: an introduction. In: Wan TSK, Cancer Cytogenetics. Springer New York; 2017. p. 1-10.
2. Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood*. 2022;140(12):1345-1377.
3. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol*. 2013;31(12):1119-1125.
4. Kaplan N, Dekker J. High-throughput genome scaffolding from in vivo DNA interaction frequency. *Nat Biotechnol*.

- 2013;31(12):1143-1147.
5. Lieberman-Aiden E, Van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-293.
 6. Harewood L, Kishore K, Eldridge MD, et al. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biol*. 2017;18(1):125.
 7. Chakraborty A, Ay F. Identification of copy number variations and translocations in cancer cells from Hi-C data. *Bioinformatics*. 2018;34(2):338-345.
 8. Dixon JR, Xu J, Dileep V, et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet*. 2018;50(10):1388-1398.
 9. Fang H, Eacker SM, Wu Y, et al. Genetic and functional characterization of inherited complex chromosomal rearrangements in a family with multisystem anomalies. *Genet Med Open*. 2025;3:103423.
 10. Fang H, Eacker SM, Wu Y, et al. Evaluation of genomic proximity mapping (GPM) for detecting genomic and chromosomal structural variants in constitutional disorders. medRxiv. 2025 Apr 25. doi: 10.1101/2025.04.23.25326303 [preprint, not peer-reviewed].
 11. Dennis MJ, Pavlick DC, Kacew A, et al. Low PD-L1 expression, MAP2K2 alterations, and enriched HPV gene signatures characterize brain metastases in head and neck squamous cell carcinoma. *J Transl Med*. 2024;22(1):960.
 12. Gonzales PR. Integration of newer genomic technologies into clinical cytogenetics laboratories. *Genes (Basel)*. 2025;16(6):688.
 13. Phase Genomics. HiC_QC. GitHub. https://github.com/phasegenomics/hic_qc Accessed November 1, 2024.
 14. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
 15. Lee S, Vitzthum C, Alver BH, Park PJ. Pairs and Pairix: a file format and a tool for efficient storage and retrieval for Hi-C read pairs. *Bioinformatics*. 2022;38(6):1729-1731.
 16. Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020;36(1):311-316.
 17. Ebert P, Audano PA, Zhu Q, et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science*. 2021;372(6537):eabf7117.
 18. Mohanty S. NUP98 rearrangements in AML: molecular mechanisms and clinical implications. *Onco*. 2023;3(3):147-164.
 19. Celik H, Koh WK, Kramer AC, et al. JARID2 functions as a tumor suppressor in myeloid neoplasms by repressing self-renewal in hematopoietic progenitor cells. *Cancer Cell*. 2018;34(5):741-756.e8.
 20. Grimwade D, Walker H, Oliver F, et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. *Blood*. 1998;92(7):2322-2333.
 21. Bendari M, Khoukha N, Cherkaoui S, et al. Current cytogenetic abnormalities in acute myeloid leukemia. In: Çelik TA, Dey S, Chromosomal Abnormalities. London: IntechOpen; 2020. p. 65-78.
 22. Haferlach C, Bacher U, Grossmann V, et al. Three novel cytogenetically cryptic EVI1 rearrangements associated with increased EVI1 expression and poor prognosis identified in 27 acute myeloid leukemia cases. *Genes Chromosomes Cancer*. 2012;51(12):1079-1085.
 23. Zhang T, Auer P, Dong J, et al. Whole-genome sequencing identifies novel predictors for hematopoietic cell transplant outcomes for patients with myelodysplastic syndrome: a CIBMTR study. *J Hematol Oncol*. 2023;16(1):37.
 24. Heald JS, López AM, Pato ML, et al. Identification of novel NUP98 fusion partners and comutations in acute myeloid leukemia: an adult cohort study. *Blood Adv*. 2024;8(11):2691-2694.
 25. Duncavage EJ, Schroeder MC, O'Laughlin M, et al. Genome sequencing as an alternative to cytogenetic analysis in myeloid cancers. *N Engl J Med*. 2021;384(10):924-935.
 26. Chen J-M, Cooper DN, Férec C, Kehrer-Sawatzki H, Patrinos GP. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol*. 2010;20(4):222-233.
 27. Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011;470(7332):59-65.
 28. Kidd JM, Graves T, Newman TL, et al. A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell*. 2010;143(5):837-847.
 29. Huddleston J, Chaisson MJP, Steinberg KM, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res*. 2017;27(5):677-685.
 30. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75-81.
 31. English AC, Salerno WJ, Reid JG. PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics*. 2014;15(1):180.
 32. Tattini L, D'Aurizio R, Magi A. Detection of genomic structural variants from next-generation sequencing data. *Front Bioeng Biotechnol*. 2015;3:92.
 33. Teo SM, Pawitan Y, Ku CS, Chia KS, Salim A. Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics*. 2012;28(21):2711-2718.
 34. Marx V. Method of the year: long-read sequencing. *Nat Methods*. 2023;20(1):6-11.
 35. Oehler JB, Wright H, Stark Z, Mallett AJ, Schmitz U. The application of long-read sequencing in clinical settings. *Hum Genomics*. 2023;17(1):73.
 36. Ahsan MU, Liu Q, Perdomo JE, Fang L, Wang K. A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. *Nat Methods*. 2023;20(8):1143-1158.
 37. Mak ACY, Lai YYY, Lam ET, et al. Genome-wide structural variation detection by genome mapping on nanochannel arrays. *Genetics*. 2016;202(1):351-362.
 38. Levy B, Baughn LB, Akkari Y, et al. Optical genome mapping in acute myeloid leukemia: a multicenter evaluation. *Blood Adv*. 2023;7(7):1297-1307.
 39. Loghavi S, Wei Q, Ravandi F, et al. Optical genome mapping improves the accuracy of classification, risk stratification, and personalized treatment strategies for patients with acute myeloid leukemia. *Am J Hematol*. 2024;99(10):1959-1968.
 40. Ballesta-Alcaraz L, Bernal M, Vilchez JR, et al. Application of optical genome mapping for the diagnosis and risk stratification of myeloid and lymphoid malignancies. *Int J Mol Sci*. 2025;26(12):5763.
 41. Swart LE, Heidenreich O. The RUNX1/RUNX1T1 network: translating insights into therapeutic options. *Exp Hematol*. 2020;94:1-10.
 42. Slovak ML, Gundacker H, Bloomfield CD, et al. A retrospective study of 69 patients with t(6;9)(p23;q34) AML emphasizes the need for a prospective, multicenter initiative for rare 'poor prognosis' myeloid malignancies. *Leukemia*. 2006;20(7):1295-1297.

43. Bill M, Mrózek K, Kohlschmidt J, et al. Mutational landscape and clinical outcome of patients with de novo acute myeloid leukemia and rearrangements involving 11q23/ KMT2A. *Proc Natl Acad Sci USA*. 2020;117(42):26340-26346.
44. Sarhadi VK, Lahti L, Scheinin I, et al. Targeted resequencing of 9p in acute lymphoblastic leukemia yields concordant results with array CGH and reveals novel genomic alterations. *Genomics*. 2013;102(3):182-188.
45. Stirewalt DL, Pogossova-Agadjanyan EL, Tsuchiya K, Joaquin J, Meshinchi S. Copy-neutral loss of heterozygosity is prevalent and a late event in the pathogenesis of FLT3/ITD AML. *Blood Cancer J*. 2014;4(5):e208.
46. Gronseth CM, McElhone SE, Storer BE, et al. Prognostic significance of acquired copy-neutral loss of heterozygosity in acute myeloid leukemia. *Cancer*. 2015;121(17):2900-2908.