Published in final edited form as:

Cancer Res. 2016 April 15; 76(8): 2197–2205. doi:10.1158/0008-5472.CAN-15-1015.

Genomic profiling of pediatric acute myeloid leukemia reveals a changing mutational landscape from disease diagnosis to relapse

Jason E. Farrar[1,2], **Heather L. Schuback**[3], **Rhonda E. Ries**[3], **Daniel Wai**[4], **Oliver A. Hampton**[5], **Lisa R. Trevino**[5,6], **Todd A. Alonzo**[2,7], **Jaime M. Guidry Auvil**[8], **Tanja M. Davidsen**[9], **Patee Gesuwan**[9], **Leandro Hermida**[9], **Donna M. Muzny**[5], **Ninad Dewal**[5], **Navin Rustagi**[5], **Lora R. Lewis**[5], **Alan S. Gamis**[10], **David A. Wheeler**[5], **Malcolm A. Smith**[11], **Daniela S. Gerhard**[8], and **Soheil Meshinchi**[2,3]

[1]Arkansas Children's Hospital Research Institute and the University of Arkansas for Medical Sciences, Little Rock, AR

[2]Children's Oncology Group, Monrovia, CA

[3]Fred Hutchinson Cancer Research Center and the University of Washington School of Medicine, Seattle, WA

[4]Ron Matricaria Institute of Molecular Medicine at Phoenix Children's Hospital and the University of Arizona College of Medicine

[5]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX

[6]Doctors Hospital at Renaissance, Edinburg, TX

[7]University of Southern California, Los Angeles, CA

[8]National Cancer Institute, Office of Cancer Genomics, Bethesda, MD

[9]National Cancer Institute, Center for Bioinformatics and Information Technology, Rockville, MD

[10]Children's Mercy Hospitals & Clinics, Kansas City, MO

[11]National Cancer Institute, Cancer Therapy Evaluation Program, Rockville, MD

Abstract

The genomic and clinical information used to develop and implement therapeutic approaches for AML originated primarily from adult patients and has been generalized to patients with pediatric AML. However, age-specific molecular alterations are becoming more evident and may signify the need to age-stratify treatment regimens. The NCI/COG TARGET-AML initiative employed whole exome capture sequencing (WXS) to interrogate the genomic landscape of matched trios representing specimens collected upon diagnosis, remission, and relapse from 20 cases of de novo childhood AML. One hundred forty-five somatic variants at diagnosis (median 6 mutations per

Corresponding author: Soheil Meshinchi, MD, PhD, Clinical Research Division, Fred Hutchinson Cancer Research Center, Department of Pediatrics, University of Washington School of Medicine, 1100 Fairview Ave N, D5-380, PO Box 19024, Seattle, WA 98109-1024, ph: 206-667-4077, fax: 206-667-4310, ; Email: smeshinc@fredhutch.org.

Conflicts of Interest: Nothing to declare.

patient) and 149 variants at relapse (median 6.5 mutations) were identified and verified by orthogonal methodologies. Recurrent somatic variants (in {greater than or equal to}2 patients) were identified for 10 genes (FLT3, NRAS, PTPN11, WT1, TET2, DHX15, DHX30, KIT, ETV6, KRAS), with variable persistence at relapse. The variant allele fraction (VAF), used to measure the prevalence of somatic mutations, varied widely at diagnosis. Mutations that persisted from diagnosis to relapse had a significantly higher diagnostic VAF compared to those that resolved at relapse (median VAF 0.43 vs. 0.24, P<0.001). Further analysis revealed that 90% of the diagnostic variants with VAF >0.4 persisted to relapse compared to 28% with VAF <0.2 (P<0.001). This study demonstrates significant variability in the mutational profile and clonal evolution of pediatric AML from diagnosis to relapse. Furthermore, mutations with high VAF at diagnosis, representing variants shared across a leukemic clonal structure, may constrain the genomic landscape at relapse and help to define key pathways for therapeutic targeting.

Keywords

childhood AML; whole exome sequencing; genomic profiling; relapse

Introduction

Comprehensive cataloguing of somatic changes in acute myeloid leukemia (AML) in adults has revealed novel molecular pathways that contribute to the development and progression of the disease (1) though many lack prognostic significance. Historically, the management of pediatric AML has been largely based on data from adult AML studies, in part because of the higher number of adult cases available for clinical trials and the assumption that there was similar biology across age groups. However, molecular profiling of AML in pediatric and adult studies have started to characterize AML as a disease with distinct age dependent alterations (2). Many of the novel and clinically impactful alterations identified in adult AML by whole genome sequencing (1,3) are not present in childhood AML,(1,4) suggesting a distinct age associated biology and highlighting the need for comprehensive genomic profiling of this disease in children. The Therapeutically Applicable Research to Generate Effective Treatment (TARGET) AML initiative is a National Cancer Institute (NCI) and Children's Oncology Group (COG) collaborative project to establish the genomic alterations present in pediatric AML with the goal of identifying novel therapeutic targets. Here we present the results of whole exome capture sequencing (WXS) to define the somatic alterations in a cohort of patients without known high-risk features who experienced relapse after having initially achieved remission.

Materials and Methods

Study Patients

Patients treated on recent COG AML trials (5–7), who achieved an initial remission, had a subsequent relapse, and a diagnostic blast percentage >50% were selected for this study. These clinical trials randomized type and timing of induction therapy (CCG-2961) and randomized additions to backbone therapy (AAML03P1 and AAML0531) but did not include molecularly-targeted treatment. All karyotypes were centrally reviewed, and patients

with high-risk cytogenetic features (monosomy 7, monosomy 5 or del5q) were excluded. Selected molecular features (e.g. *KIT, RAS, NPM, WT1, CEBPA, IDH1* mutations and *FLT3/ITD*) were clinically available (4,8,9).

Library construction

The complete library construction, exome capture, and DNA sequencing protocols can be accessed at ftp://caftpd.nci.nih.gov/pub/dcc_target/AML/Discovery/WXS/METADATA/.

Mapping Reads

Illumina HiSeq bcl files were processed using BCLConvertor v1.7.1. All reads from the prepared libraries that passed the Illumina Chastity filter were formatted into fastq files. The fastq files were aligned to human reference genome build37 (NCBI) using BWA (bwa-0.5.9-R16) with default parameters with the following exceptions: seed sequence [40], bpseed mismatch [2], total mismatches allowed [3]. BAM files generated from alignment were preprocessed using GATK (v1.3-8-gb0e6afe)(10) to recalibrate and locally realign reads.

Data Deposition

The raw data for all 60 specimens are available at dbGaP, study phs000465.

Mutation Detection

Full methods are available at ftp://caftpd.nci.nih.gov/pub/dcc_target/AML/Discovery/WXS/. The somatic variant allele fraction (VAF), defined as the number of variant reads/total reads at sites of variants not detected in remission specimens, was used to estimate the prevalence of leukemic mutations.

ITD Detection

We used ITD Assembler (11), a combined *de novo* assembly/algorithmic approach that takes the entire set of unmapped and significantly soft-clipped reads, and employs a De Bruijn graph assembly algorithm to select read sets that form cycles, indicative of repetitive sequence structures, in order to find reads that span duplications. Read sets that formed De Bruijn graph cycles are independently assembled using the Overlap Layout Consensus (OLC) methodology of the Phrap algorithm thereby alleviating the collapse of repeat sequences from De Bruijn graph assembly approaches. Resulting OLC assembled contigs are locally aligned to the reference sequence and the mapped location data from aligned soft-clipped reads and aligned-unaligned read pairs from that contig, are utilized to annotate the position of detected internal tandem duplications (ITDs). *FLT3/ITDs* were also verified by fragment length analysis utilizing Life Technologies' GeneMapper software (Life Technologies, Grand Island, NY). Variant allele fractions are reported as the size of the ITD peak divided by the sum of the wild type and ITD peaks.

Copy Number Aberrations (CNA) Detected by LOHcate

CNA analysis was performed using LOHcate (12), a method that identifies CNA events in whole exome tumor sequence data via detecting enrichment in variant or reference allele quantities per site across polymorphic exonic sites. The per-site quantities are plotted two-

dimensionally between matched normal and tumor. Significant sites are then clustered in this Euclidean 2D space using an optimized version of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm (13), after which clusters are classified to denote the appropriate CNA events: somatic gain, loss of heterozygosity (LOH), or copy neutral LOH (cnLOH). The sites within these classified CNA clusters are mapped back to the exome, which is subsequently segmented into CNA regions. Regions and sub-regions of low to high recurrence are identified by comparing CNA regions across samples. Segments with 200 or more markers were visualized by Partek software (Partek Inc., St. Louis, MO) and utilized in the analysis.

Validation on 454 and Ion Torrent Platforms

Whole exome capture sequencing (WXS) mutation calls were validated using alternative instrumentation and chemistries to avoid systematic errors inherent to the processes described above. Variant sites were selected for amplification by PCR followed by analysis on 454 sequencing instruments as previously described (14). Amplicons were also prepared for sequencing using the Life Technologies Ion Xpress and Ion OneTouch protocols and reagents (Life Technologies, Inc., Grand Island, NY). Briefly, amplicons were clonally amplified on Ion Sphere Particles (ISPs) through emulsion PCR and then enriched for template-positive ISPs. For Ion Torrent runs, approximately 35 million template-positive ISPs per run were deposited onto the Ion 318C chips (Life Technologies, Cat. No. 4466617) by a series of centrifugation steps that incorporated alternating the chip directionality. Sequencing was performed with the Ion Personal Genome Machine Sequencing Kit (Life Technologies, Cat. No. 4474004) using the 440 flow (200 bp) run format. The instrumentation used for validation is indicated in the MAF files. Verification sequencing results were analyzed by employing a two-step mapping process. First, fastq sequence files were aligned to the human genome reference using BLAT (v33); the top-scoring alignment was reported from the cognate amplicon hits only if it was greater than 90% of the next-best hit. Second, the passing BLAT hits were pair-wise aligned to their respective amplicon sequence using cross-match (v1.080812). A putative mutation was considered verified in this secondary assay for sites with a minimum of 50 reads in the secondary screening.

Statistical Methods

Study data were frozen as of December 31, 2012. Differences in medians were compared using the Wilcoxon Rank Sum test for two medians and Kruskal-Wallis test for three medians. Fisher’s exact test was used to compare two proportions whereas analysis of variance (ANOVA) was used to compare three proportions.

Results

Patients

Demographic, clinical and laboratory characteristics of the 20 study patients are listed in Supplemental Table 1 and available at http://target.nci.nih.gov/dataMatrix/TARGET_DataMatrix.html. Age at diagnosis ranged from 1 to 17 years (median of 10.1 years). AMLs had variable morphology, karyotype and molecular alterations (Supplemental Figure 1A and B). Patients with high-risk cytogenetics (monosomy 7, 5q-, complex

karyotype) were excluded. Those with core binding factor translocations (t(8;21) or inv(16), CBF AML) constituted 35% of the study population; of the remaining patients, two had normal karyotype, one an MLL rearrangement and the remainder had a cytogenetic abnormality without established clinical significance.

Whole Exome Sequencing

Whole exome capture sequencing obtained from trios of diagnostic, remission, and relapse specimens from 20 children with AML were interrogated to define the somatic alterations at diagnosis and relapse. The coding region of approximately 30,000 genes from RefSeq, CCDS and miRBase databases were sequenced to a minimum depth of 20X in >94% of target bases. The somatic mutation profile at diagnosis was compared to that at relapse to identify events that persist, develop or disappear after treatment. All computationally derived variants underwent verification by Roche 454 or Ion Torrent sequencing at a median depth of 535X (range 10–9716).

A broad range of non-synonymous somatic mutations, including missense, nonsense, insertion/deletion (indel, both in-frame and frameshift), as well as splice site variants, were identified. Overall, 145 variants in 120 genes at diagnosis and 149 variants in 136 genes at relapse were successfully verified through secondary methodologies. A list of verified, somatic variants are provided in Supplemental Table 2. One variant at diagnosis, 3 at relapse, and 3 present at both diagnostic and relapse timepoints were identified in 3 patients in regions of copy alteration or copy-neutral LOH; these were tallied for description but excluded from subsequent analyses of VAF changes with disease progression. The number of verified, somatic diagnostic mutations ranged from 1 to 18 per patient with a median of 6 mutations per patient. Similar findings were seen in relapse, with a median of 6.5 mutations per patient (range 0–18). The combined total number of verified mutations at diagnosis or relapse varied from 1 to 36 with a median of 12.5 (Figure 1A). The majority of verified mutations represented missense changes (120/145 in the diagnostic and 132/148 in relapse samples) (Figure 1B). Of the missense mutations, PolyPhen (15) prediction analysis estimated 72% were possibly and/or probably functionally damaging lesions. In addition, we evaluated the distribution of single-nucleotide base pair changes observed in somatic variants. The most common base-pair change was a cytosine-to-thymine (C→T) transition seen in 73% of all variants at diagnosis and 60% of variants at relapse (Figure 1C).

Younger patients (< 2 years) had substantially fewer mutations at diagnosis (median of 3.5 mutations per patient) compared to older patients (2–17 years of age; median of 8 mutations, P=0.006). Evaluation of the number of mutations identified at diagnosis in specific cytogenetic or molecular cohorts demonstrated variability, although these differences were not statistically significant. A median of 10 mutations per patient was seen in CBF AML compared to 5 in patients with *MLL*/Other karyotype abnormalities or *FLT3/ITD* (Figure 2A and B).

Of the combined 294 non-synonymous mutations in 179 genes identified at diagnosis or relapse, variants in ten genes were seen in two or more patients, including *FLT3* in eight patients, *NRAS* in seven, *PTPN11*, *WT1*, and *TET2* in three, and *DHX15, DHX30, KIT, KRAS,* and *ETV6,* in two patients (Figure 3). Gene expression data did not demonstrate

significant differences in expression in comparison of mutated and non-mutated specimens (Supplemental Figure 2).

With the exception of *FLT3/ITD*, all previously ascertained mutations in these samples (*KIT*, *RAS*, *NPM*, *WT1*, and *CEBPA*) were independently identified in the WXS data. The variation in length and site of insertion of ITDs in *FLT3* has made detection of such mutations in short-read sequence data a computationally challenging process. Sequence data underwent additional computational analysis by two algorithms (ITD Assembler and Pindel) specifically optimized for detection of internal tandem duplications (ITDs). Both algorithms identified *FLT3/ITD* mutations in all but one case with clinically annotated FLT3/ITD, which was known to have low allele fraction *FLT3/ITD*. Although presence of FLT3/ITD could be successfully confirmed in 8 of 9 specimens with clinically annotated mutation, the concordance of allelic burden measured by this sequencing-based approach was poor in comparison to established clinical testing.

Recurrent Mutations and Common Pathways

Although recurrent mutations of individual genes were not prevalent, the mutations clustered into groups of 11 gene families with established relevance to leukemogenesis (Figure 4). Mutations in nine tyrosine kinase genes, including *FLT3* (N=8), *KIT* (N=2), *JAK2* (N=1), *EPHA5* (N=1), *POR1* (N=1), *OBSCN* (N=1), *MAP4K1* (N=1), *TRPM7* (N=1), and *PRKD1* (N=1), were identified from 12 patients in the cohort. Similarly, from diagnostic samples, mutations in 13 RAS/MAPK/MEK family genes were identified from 14 patients, mutations in 14 transcription factor genes in ten patients, and mutations in 11 epigenetic modifier genes in nine patients (Figure 4A). While tyrosine kinase (TK) genes or RAS/MAPK/MEK pathway alterations were present in 90% of cases, most patients (15/18) with mutations in one of these gene families had mutations in one of the additional functional groupings. All patients had a mutation in at least one of these functional groups. Overall, we identified somatic mutations in 78 genes from 11 functional groups with the majority implicating key leukemia-driving signaling pathways, including TK, RAS/MAPK/MEK pathways, and secondly, transcription factors and epigenetic modifiers (Figure 4B).

Clonal Evolution from Diagnosis to Relapse

Sequence Variants—Evaluation of the genomic alterations from diagnosis to relapse revealed significant heterogeneity, including numerous mutations with resolution of diagnostic variants, emergence of novel mutations at relapse as well as marked changes in VAFs from diagnosis to relapse. Mutations at diagnosis and relapse, along with their corresponding VAFs, are depicted in Supplemental Figure 3.

Of the 141 verified diagnostic somatic, non-synonymous mutations, only 83 (58%) persisted and were identified at relapse. Conversely, of the 143 mutations detected at relapse, 60 (42%) appeared to be newly evolved variants that were not detected at diagnosis. Of all somatic mutations detected at diagnosis or at relapse, 58 (29%) were present only in diagnostic specimens, 60 (30%) were present only at the time of relapse and 83 (41%) variants were present in both specimens (Figure 5A) highlighting a significant molecular evolution with loss and gain of genomic events from diagnosis to relapse.

Correlation of Diagnostic Variant Allele Fraction and Relapse—The VAFs for all verified mutations in diagnostic samples varied significantly from 0.037–0.86 (median VAF 0.35). In order to define whether diagnostic VAF correlated with the genomic landscape at relapse, we compared the diagnostic VAF in variants that persisted to those that were lost at relapse. The median diagnostic VAF for mutations that persisted to relapse was 0.43 (range 0.055–0.72) vs. 0.24 (range 0.048–0.86, P<0.0001) for the cleared variants (Figure 5B), suggesting that those with high diagnostic VAF are more likely to persist to relapse. We further evaluated different diagnostic VAF threshold with persistence to relapse. Sixty mutations had a diagnostic VAF of ≥0.4, of which 54 (90%) were also detected in the relapse specimens. In contrast, of the diagnostic mutations with VAF of <0.2 (N=38), only 11 (28%) were detected at relapse (P<0.001) while those with intermediate VAF had an intermediate association with persistence at relapse (Figure 5C). These trends are evident at the individual as well as aggregate scale. Of 14 patients in whom at least one high VAF (>0.4) variant was identifiable at presentation, 13 (93%) had a persistent alteration identified at relapse. Interestingly, of the 6 patients who lacked at least one dominant variant, 4 had core binding factor translocations.

Fifty-nine variants in 50 genes present at diagnosis were not detected in the corresponding relapse specimens. These genes included recurrent alterations in genes with established relevance in driving myeloid malignancy such as *FLT3* (N=3, 2 *FLT3/ITD*s and 1 *FLT3* TKD mutations), 5 *NRAS* mutations (N=4), and 3 *PTPN11* mutations (N=2). Of the 8 total *FLT3* alterations identified in diagnostic or relapse specimens (7 ITDs and 1 point mutation), 3 were only detected at diagnosis, 4 persisted from diagnosis to relapse and 1 was detected at relapse only. Similarly, of the 11 *NRAS* variants identified in 7 patients at diagnosis, 5 were not present at relapse. It is notable that of the 19 variants at diagnosis involving *FLT3*, *NRAS,* and *PTPN11*, all variants that resolved from diagnosis to relapse had initial VAFs <0.2, suggesting that the identification of dominant mutations at diagnosis, rather than simple categorical presence of a mutation, may be more relevant in the selection of targeted therapies.

Copy Number Alterations (CNAs)—We used enrichment for variant or reference alleles across polymorphic exonic regions to identify segmental chromosomal alterations. In addition to confirming major karyotypic CNAs, we identified novel regions of copy number gain or loss, and copy neutral loss of heterozygosity (cnLOH) in diagnostic and relapse specimens (Figure 6A). In total, 155 segments varying in length from 0.75 Mb to 168.81 Mb (median 7.20 Mb) were detected at diagnosis (N=72) or relapse (N=83), of which 23 segments were present in both diagnostic and relapse specimens (46/155 segments) (Figure 6B). Accordingly, 49 distinct segments were identified at diagnosis only and 60 segments were identified at relapse only. Thus, of the total CNA segments present at diagnosis or relapse, only 30% were identified at both time points, demonstrating somatic CNAs, like mutations, evolve significantly from diagnosis to relapse (Figure 6C). In addition to areas of LOH associated with deletions, segmental cnLOH regions were detected in 15 of 20 patients (75%), highlighting the prevalence of this previously undetected genomic alteration in AML. We further examined CNA regions of less than 5Mb in length that would be below the resolution threshold of conventional karyotyping. We identified ten recurrent (> 2 patients)

CNA segments (<5Mb) with two regions on 11p and 16p in five patients. Supplemental Table 3 contains a list of all segments.

Clonal Architecture at Diagnosis and Relapse—Given the significant genomic heterogeneity of AML at diagnosis and variable clonal evolution involving sequence and copy number alterations at relapse, we employed clonal cluster analysis to map changes in clonal architecture from diagnosis to relapse. Variant allele fractions for mutations present at diagnosis were used to study the genomic complexity at diagnosis and track the genomic evolution to relapse. Each patient's mutations were clustered using the VAF scores (DBSCAN) and graphed according to time point. In addition, CNV data was used to further inform the somatic VAF (deletions or CN-LOH impact calculated VAF). Using this information, we defined possible scenarios for clonal evolution and tracked each mutation from diagnosis to relapse to map clones that persisted, those that resolved from diagnosis to relapse as well as novel mutations that emerged at relapse. Noting that these data are open to variable interpretation, Supplemental Figure 4 shows various scenarios for clonal evolution involving known mutations, such as *FLT3* and *NRAS,* as well as novel mutations. These data depict the genomic complexity of childhood AML and delineate the association of clonal dominance at diagnosis with likelihood of persistence of mutations from diagnosis to relapse.

Discussion

The ongoing NCI/COG cooperative TARGET AML initiative aims to interrogate the childhood AML genome, transcriptome and epigenome in order to define biologic biomarkers that can be used for more precise risk stratification and therapeutic intervention. The current study included WXS of 20 matched trios of diagnostic, remission and relapse specimens to better understand molecular evolution between presentation and relapse. We demonstrate the heterogeneous nature of the somatically altered AML genome but also the paucity of recurrently mutated genes in childhood AML, as has previously been reported for adult *de novo* AML. However, although highly recurrent individual mutations were not identified, mutations in specific gene families and key cellular pathways are common. Protein-altering mutations in 78 genes belonging to only 11 known pathways were identified among the cases sequenced. Mutations in tyrosine kinase and RAS/MAPK/MEK were identified in 90% of patient samples, and were the predominantly altered pathways. In addition, mutations in transcription factors and epigenetic modifiers were also common and appeared to overlap with the activated TK and RAS/MAPK variants. Further, all phosphatase mutations occurred in the background of TK or RAS/MAPK pathway mutations, suggesting that alterations of phosphatases may cooperate with TK/RAS/MAPK mutations during AML leukemogenesis.

These data also demonstrate that the presence or lack of specific somatic variants may not operate as a binary entity. Rather, the clonal prevalence of a mutation, a property reflective of both order of acquisition during leukemogenesis as well as the proliferative drive induced by the accumulated constellation of mutations, rather than its mere presence may be a significant contributor to the clinical picture. We have previously demonstrated that leukemogenic *FLT3* variants that develop in the early progenitor population have

significantly different clinical implications than the same mutations that evolve in the more mature, differentiated cells (16). In these studies Pollard et al., demonstrated that presence of *FLT3/ITD* in CD34+/CD33− progenitors was highly associated with relapse; in contrast, those where *FLT3/ITD* was restricted to the CD34+/CD33+ fraction, a more favorable outcome was observed. We have also demonstrated significant association of allelic variation of *FLT3/ITD* and clinical outcome, where those patients with AML having a high mutant to normal allelic ratio *FLT3/ITD* had adverse outcomes, while those with AML having a lower *FLT3/ITD* allelic ratio showed outcomes similar to patients without *FLT3* mutations (17). This observation has been validated in other trials (18,19) and has been incorporated in the last two COG AML studies, where patients with high allelic *FLT3/ITD* ratio are allocated to the high risk arm and directed to receive hematopoietic stem cell transplantation (20). Similar observations have also been reported for other mutations in AML, including *KIT* and *CBL*, where allelic burden at diagnosis correlated with outcome (21).

The findings presented here suggest these mechanisms may operate more broadly in AML as well as offer a potential explanation for differences in outcome based on the clone or subclone of origin. The observation that the majority of dominant variantspersist from diagnosis to relapse, and the corollary, that many subclonal variants do not, may reflect both the context of acquisition during leukemic ontogeny in addition to the potential to drive proliferation. Surprisingly, even mutations of genes with established roles as drivers in myeloid leukemogenesis such as *NRAS* and *PTPN11*, were lost at relapse when initially present as subclonal, low VAF variants. This is in contrast to recently reported data in JMML, where essentially all somatic pathogenic mutations at diagnosis were present at relapse (22). This suggests hat the presence of even driver somatic variants may not be sufficient to drive myeloid leukemogenesis outside of limited contexts. The clinical significance of a variant may be more closely linked with whether it occurs during ancestral leukemic development (dominant variants) or in a more distant sub-clone (minor variants) (16,23).

Current pediatric AML treatment stratification schema incorporate data on mutation status of several frequently mutated genes for the classification of risk, although the number of known mutations significantly exceeds those incorporated for risk stratification, since many apparently lack prognostic significance. The clinical incorporation of allelic burden testing of specific mutations provides an opportunity for prioritization of mutations for therapeutic targeting and risk-based therapy allocation. In such a scenario, those variants with higher diagnostic VAF would be preferentially targeted, as dominant variants would be more clinically relevant. In addition to the loss of variants present at diagnosis, evolution of novel variants at relapse may provide insight into the selective pressures that predispose to their appearance and contribution to relapse. The contribution of the new, emerging mutations at relapse must be carefully assessed, as events that lead to the evolution of primary disease and those that lead to emergence of resistance and relapse may be different.

The genomic landscape at relapse appears to be fluid and in part shaped by the selective pressure exerted by the applied therapy. Emerging data in a specific subset of patients have clarified that different therapeutic agents can lead to evolution of distinct mutations at relapse. As an example, patients with *FLT3/ITD* who receive conventional chemotherapy vs.

targeted therapy with tyrosine kinase inhibitors (TKIs) have a distinctly different profile at relapse. For instance, treatment with TKIs selectively leads to secondary mutations in the activation loop domain of the gene, where those who relapse after exposure to TKI show evolution of secondary *FLT3* mutations that are associated with TKI resistance (24,25).

The WXS data presented in this study highlight the heterogeneity of AML at diagnosis, as well as the changing genomic make up during therapy and at relapse. This remains a critical issue for understanding the biological drivers of these complex myeloid malignancies in children. The mutation profile, guided by VAF, has important implications in treatment assignment as we are moving away from a morphological classification system towards a more precise, genomic guided therapy allocation. A more complete understanding of the contribution these myriad of somatic mutations to the leukemic process is critical for more precise prioritization of the genomic events for intervention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial Support: This work was supported by COG Chairs U10 CA180886-01 and U10CA98543, NCI R01 CA 114563 (SM); NCI TARGET U10 CA 98543 (SM/RJA); Arkansas Biosciences Institute (JEF); the Ronald Matricaria Institute of Molecular Medicine, the Lund Foundation, the Najafi Foundation, The Children's Cancer Network, St. Baldrick's Foundation Consortium Award (RJA); American Society of Clinical Oncology Young Investigator Award, Alex's Lemonade Stand Foundation Young Investigator Award, St. Baldrick's Foundation Fellowship Award (HLS)

This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

The authors would like to gratefully acknowledge the important contributions of the late Dr. Robert Arceci to the AML TARGET initiative.

References

1. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. The New England journal of medicine. 2010; 363(25):2424–33. [PubMed: 21067377]
2. Schuback HL, Arceci RJ, Meshinchi S. Somatic characterization of pediatric acute myeloid leukemia using next-generation sequencing. Seminars in hematology. 2013; 50(4):325–32. [PubMed: 24246700]
3. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. The New England journal of medicine. 2009; 361(11):1058–66. [PubMed: 19657110]
4. Ho PA, Kutny MA, Alonzo TA, Gerbing RB, Joaquin J, Raimondi SC, et al. Leukemic mutations in the methylation-associated genes DNMT3A and IDH2 are rare events in pediatric AML: a report from the Children's Oncology Group. Pediatric blood & cancer. 2011; 57(2):204–9. [PubMed: 21504050]
5. Lange BJ, Smith FO, Feusner J, Barnard DR, Dinndorf P, Feig S, et al. Outcomes in CCG-2961, a children's oncology group phase 3 trial for untreated pediatric acute myeloid leukemia: a report from the children's oncology group. Blood. 2008; 111(3):1044–53. [PubMed: 18000167]

6. Cooper TM, Franklin J, Gerbing RB, Alonzo TA, Hurwitz C, Raimondi SC, et al. AAML03P1, a pilot study of the safety of gemtuzumab ozogamicin in combination with chemotherapy for newly diagnosed childhood acute myeloid leukemia: a report from the Children's Oncology Group. Cancer. 2012; 118(3):761–9. [PubMed: 21766293]
7. Sung L, Aplenc R, Alonzo TA, Gerbing RB, Gamis AS, Group AP. Predictors and short-term outcomes of hyperleukocytosis in children with acute myeloid leukemia: a report from the Children's Oncology Group. Haematologica. 2012; 97(11):1770–3. [PubMed: 22801969]
8. Pollard JA, Alonzo TA, Gerbing RB, Ho PA, Zeng R, Ravindranath Y, et al. Prevalence and prognostic significance of KIT mutations in pediatric patients with core binding factor AML enrolled on serial pediatric cooperative trials for de novo AML. Blood. 2010; 115(12):2372–9. [PubMed: 20056794]
9. Ho PA, Alonzo TA, Gerbing RB, Kuhn J, Pollard JA, Hirsch B, et al. The prognostic effect of high diagnostic WT1 gene expression in pediatric AML depends on WT1 SNP rs16754 status: report from the Children's Oncology Group. Pediatric blood & cancer. 2014; 61(1):81–8. [PubMed: 23956224]
10. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010; 20(9):1297–303. [PubMed: 20644199]
11. Rustagi N, Hampton O. ITDetect: A tool for Internal Tandem Duplication Detection from Short Read Massively Parallel Sequencing. In Preparation.
12. Dewal N. LOHcate: Robust detection and analysis of aneuploidy in whole exome sequence data from cancer genomes. In Preparation.
13. Ester, M.; Kriegel, H-PSJ.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Han, ESJ.; Fayyad, U., editors. Second International Conference on Knowledge Discovery and Dana Mining (KDD-96). AAAI Press; Menlo, Park, California: 1996. p. 226
14. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487(7407):330–7. [PubMed: 22810696]
15. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Human molecular genetics. 2001; 10(6):591–7. [PubMed: 11230178]
16. Pollard JA, Alonzo TA, Gerbing RB, Woods WG, Lange BJ, Sweetser DA, et al. FLT3 internal tandem duplication in CD34+/CD33− precursors predicts poor outcome in acute myeloid leukemia. Blood. 2006; 108(8):2764–9. [PubMed: 16809615]
17. Meshinchi S, Alonzo TA, Stirewalt DL, Zwaan M, Zimmerman M, Reinhardt D, et al. Clinical implications of FLT3 mutations in pediatric AML. Blood. 2006; 108(12):3654–61. [PubMed: 16912228]
18. Schnittger S, Bacher U, Kern W, Alpermann T, Haferlach C, Haferlach T. Prognostic impact of FLT3-ITD load in NPM1 mutated acute myeloid leukemia. Leukemia. 2011; 25(8):1297–304. [PubMed: 21537333]
19. Gale RE, Green C, Allen C, Mead AJ, Burnett AK, Hills RK, et al. The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia. Blood. 2008; 111(5):2776–84. [PubMed: 17957027]
20. Gamis AS, Alonzo TA, Meshinchi S, Sung L, Gerbing RB, Raimondi SC, et al. Gemtuzumab ozogamicin in children and adolescents with de novo acute myeloid leukemia improves event-free survival by reducing relapse risk: results from the randomized phase III Children's Oncology Group trial AAML0531. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2014; 32(27):3021–32. [PubMed: 25092781]
21. Allen C, Hills RK, Lamb K, Evans C, Tinsley S, Sellar R, et al. The importance of relative mutant level for evaluating impact on outcome of KIT, FLT3 and CBL mutations in core-binding factor acute myeloid leukemia. Leukemia. 2013; 27(9):1891–901. [PubMed: 23783394]
22. Stieglitz E, Taylor-Weiner AN, Chang TY, Gelston LC, Wang YD, Mazor T, et al. The genomic landscape of juvenile myelomonocytic leukemia. Nature genetics. 2015; 47(11):1326–33. [PubMed: 26457647]

23. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature. 2014; 506(7488):328–33. [PubMed: 24522528]

24. Smith CC, Wang Q, Chin CS, Salerno S, Damon LE, Levis MJ, et al. Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. Nature. 2012; 485(7397):260–3. [PubMed: 22504184]

25. Meyer JA, Wang J, Hogan LE, Yang JJ, Dandekar S, Patel JP, et al. Relapse-specific mutations in NT5C2 in childhood acute lymphoblastic leukemia. Nature genetics. 2013; 45(3):290–4. [PubMed: 23377183]

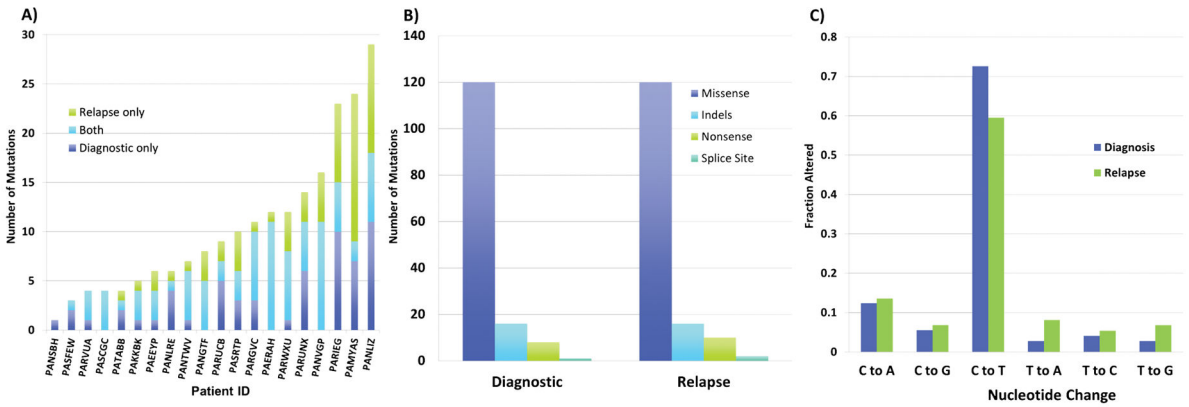

Figure 1.

Mutation distribution and types of variants. A) Number of verified, non-synonymous, somatic mutations per patient in all samples. B) Types of verified, somatic, non-synonymous mutations found in diagnostic and relapse samples. C) Nucleotide changes observed among the missense mutations.

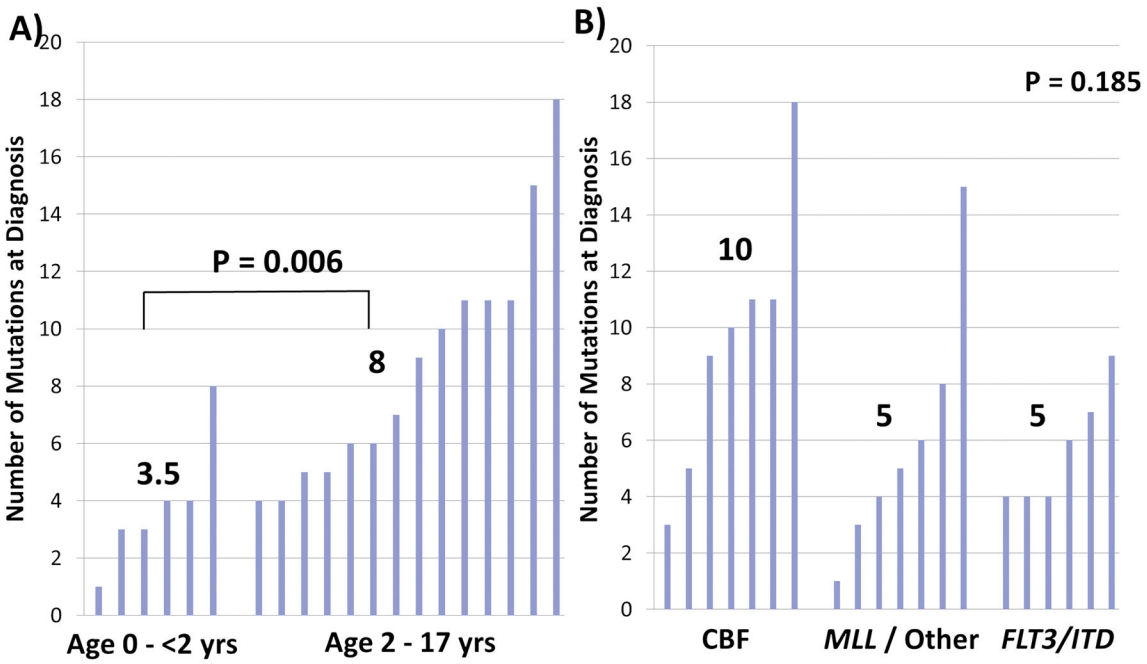

Figure 2.

Number of diagnostic, verified, somatic non-synonymous mutations by group. A) Infant AML, age 0 – <2 years vs. 2 – 17 years and B) risk-group, core binding factor abnormality (CBF) vs. MLL rearrangement and other karyotypic abnormalities vs. *FLT3/ITD* positive (B). Numbers above each cluster signify the median for that risk group. P value determined by Wilcoxon Rank Sum test for (A) and Kruskal-Wallis test for (B).

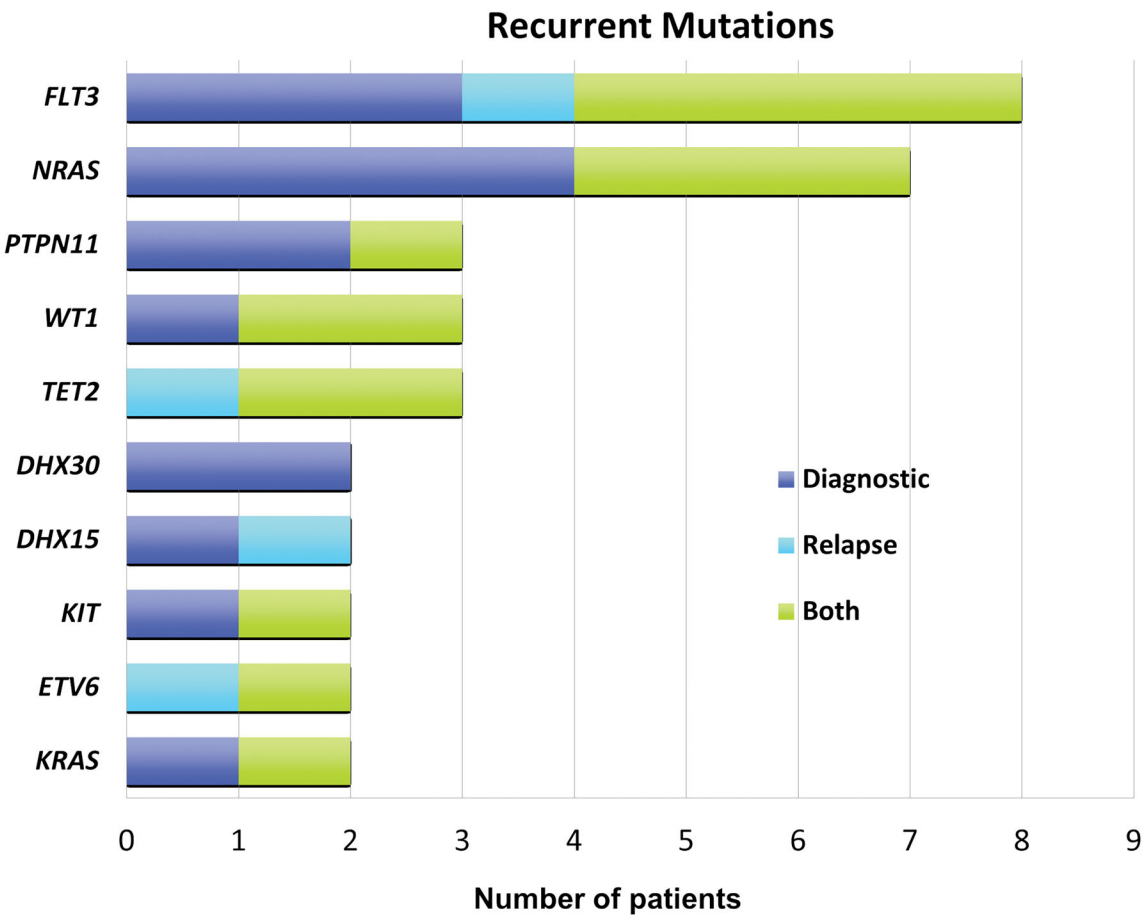

Figure 3.
Number of patients with recurrent mutations.

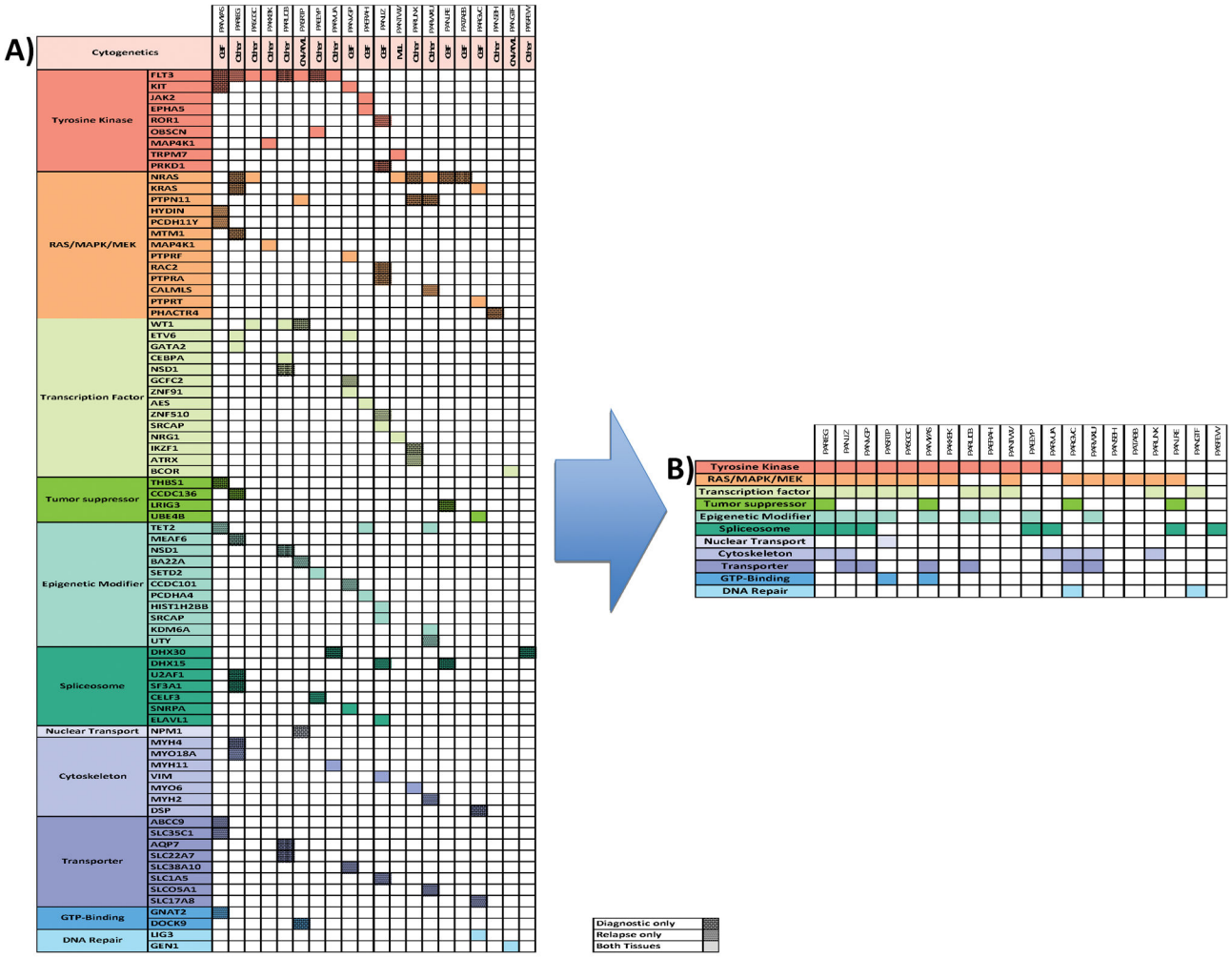

Figure 4.

Pathway analysis of identified mutations. A) The cytogenetic category is indicated below each patient identifier: normal karyotype (CN-AML), core binding factor AML (CBF) indicating presence of either t(8;21) or inv16, MLL rearrangement, or other karyotype abnormality. Tissue type is indicated by degree of shading per legend above (diagnostic only: dark shading, relapse only: medium shading, both tissues: no shading). Both refers to mutations identified at diagnosis that persist to relapse. B) is a collapsed version of (A) demonstrating the prevalence of mutations in each pathway. Figure 5. Table displaying number of amplifications, deletions or cnLOH found at diagnosis and relapse (A). Venn diagram depicting copy number segments present in diagnostic only, relapse only, or both tissues (B). Venn diagram depicting mutations present at diagnosis only, relapse only, or in both tissues (C).

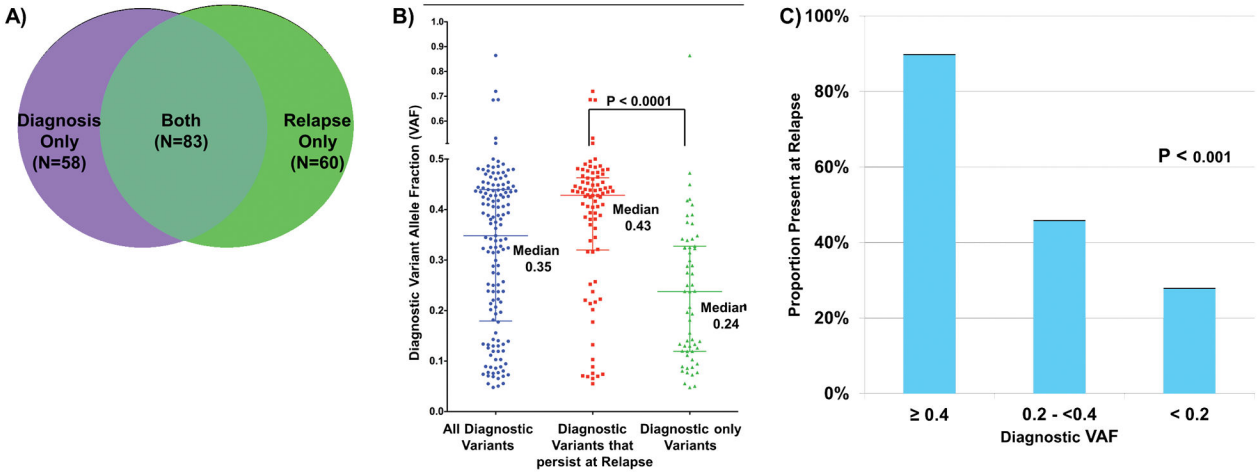

Figure 5.

Number of mutations and VAF in tissues. A) Venn diagram depicting mutations present at diagnosis only, relapse only, or in both tissues. B) Distribution of allele fraction in all diagnostic variants, diagnostic variants that persist at relapse and diagnostic variants only. P value determined by Kruskal-Wallis test. C) Proportion of variants present at relapse with diagnostic allele fraction greater than 0.40, between 0.20 and 0.40 and less than 0.20. P value determined by ANOVA.

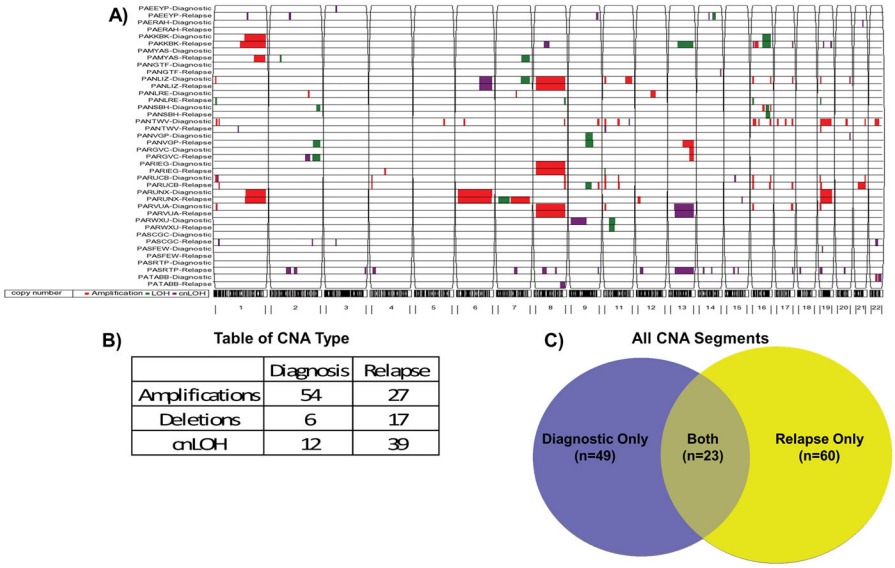

Figure 6.

Copy number segments determined by WXS. A) CNA segments were predicted by LOHcate, filtered to include only regions >200 markers, and visualized by Partek. Segments include somatic amplifications shown in red, loss of heterozygosity (LOH) shown in green, and copy neutral loss of heterozygosity (cnLOH) shown in purple. Paired patient samples (Diagnostic and Relapse) are on the y-axis while chromosomes are indicated on the x-axis. B) Table displaying number of amplifications, deletions or cnLOH found at diagnosis and relapse. C) Venn diagram depicting copy number segments present in diagnostic only, relapse only, or both tissues.