Published in final edited form as:

Cancer Genet. 2015 ; 208(1-2): 1–18. doi:10.1016/j.cancergen.2014.11.003.

Integration of cytogenomic data for furthering the characterization of pediatric B-ALL: a multi-institution, multi-platform microarray study

LB Baughn^a, JA Biegel^b, ST South^c, T Smolarek^d, S Volkert^a, A Carroll^e, NA Heerema^f, KR Rabin^g, PA Zweidler-McKay^h, M Lohⁱ, and B Hirsch^{a,j}

^aDepartment of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota, USA

^bDepartment of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

^cARUP Laboratories, Department of Pathology, University of Utah, Salt Lake City, Utah, USA

^dDivision of Human Genetics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

^eDepartment of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA

^fDepartment of Pathology, The Ohio State University Wexner Medical Center, Columbus, Ohio, USA

^gDivision of Pediatric Hematology/Oncology, Texas Children's Cancer Center, Baylor College of Medicine, Houston, Texas, USA

^hDepartment of Pediatrics, University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

ⁱDepartment of Pediatrics, Benioff Children's Hospital, University of California San Francisco, California, USA

Abstract

It is well documented that among subgroups of B-ALL, the genetic profile of the leukemic blasts has significant impact on prognosis and stratification for therapy. Recent studies have documented the power of microarrays to screen genome-wide for copy number aberrations (CNAs) and regions of copy number neutral loss of heterozygosity (CNLOH) that are not detectable by G-banding or FISH. These studies have involved application of a single array platform for the respective cases. The present investigation demonstrates the feasibility and usefulness of integrating array results from multiple laboratories (ARUP, Children's Hospital of Philadelphia, Cincinnati Children's Hospital Medical Center, and University of Minnesota Medical Center) that utilize different array

^jCorresponding author: Betsy Hirsch, PhD, Dept. of Laboratory Medicine and Pathology Director of Cytogenetics, University of Minnesota Medical School, Mayo Mail Code 609, 420 Delaware Street, Minneapolis, MN. 55455 (hirsch003@umn.edu) 612-273-4952.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

platforms (Affymetrix, Agilent, or Illumina) in their respective clinical settings. Sixty five patients enrolled on the Children's Oncology Group (COG) study AALL08B1 were identified for study, as cytogenetic and fluorescence-in-situ hybridization studies had also been performed on these patients, with central review of those results available for comparison. Microarray data were first analyzed by the individual laboratories with their respective software systems; raw data files were then centrally validated using NEXUS software. The results demonstrated the added value of integrating multi-platform data with cytogenetic and FISH data and highlight novel findings identified by array including the co-occurrence of low and high risk abnormalities not previously reported to coexist within a clone, novel regions of chromosomal amplification, clones characterized by numerous whole chromosome LOH that do not meet criteria for doubling of a near-haploid, and characterization of array profiles associated with *IKZF1* deletion. Each of these findings raises questions that are clinically relevant to risk stratification.

Keywords

B-ALL; microarray; ETV6-RUNX1; hypodiploid; IKZF1; iAMP21

Introduction

Acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy [1]. While overall survival of pediatric B-ALL exceeds 90% with current treatment protocols [2], there is marked variability between subgroups of ALL, with the genetic profile of the leukemic blasts having a significant impact on prognosis. In the United States the majority of pediatric patients with ALL are entered on clinical trials coordinated through the Children's Oncology Group (COG). Patients are initially classified based on the following metrics: age, initial white blood cell count, extramedullary disease status, steroid pretreatment, cytogenetic abnormalities and measurement of minimal residual disease one week into treatment and at the end of induction therapy. Cytogenetic findings have been used to classify patients based on risk of relapse. Currently there are two subgroups associated with a highly favorable prognosis: the cryptic translocation t(12;21)(p13;q22) generating an *ETV6-RUNX1* fusion and the high hyperdiploid karyotype (51-65 chromosomes) that includes trisomies of chromosomes 4 and 10 [3]. In contrast, there are four subgroups associated with a poor prognosis: the t(9;22)(q34;q11.2) generating a *BCR-ABL1* gene fusion, any translocation involving the *KMT2A* (previously known as *MLL*) gene at 11q23, a hypodiploid karyotype with chromosome numbers <43 and amplification of *RUNX1* on chromosome 21 (iAMP21) [2]. These cytogenetic findings are used to help stratify patients for optimal therapy. Local laboratories must perform fluorescence in situ hybridization (FISH) to detect *ETV6-RUNX1*, *BCR-ABL1*, *KMT2A (MLL)* rearrangements, and trisomies 4 and 10 at the time of leukemia diagnosis as a requirement for COG study AALL08B1, a classification study of pediatric ALL that is a prerequisite for patient enrollment on therapeutic COG ALL trials. The karyotypes and FISH results are centrally reviewed by the COG cytogenetics committee before the patient is stratified for therapy after initial induction chemotherapy. As molecular genetic and molecular cytogenetic methods have advanced, additional recurring submicroscopic genomic alterations in pediatric B-ALL have been identified. High-resolution genomic arrays, gene expression arrays and whole exome/whole genome

sequencing technologies are being applied to characterize the genomic landscape of each leukemia. Aberrations in pathways controlling B cell development, cell cycle progression and tyrosine kinase signaling have been identified including frequent copy number aberrations (CNAs) in *CDKN2A/B*, *PAX5*, *ETV6* and *IKZF1* [4-10] and activating mutations involving the *JAK/STAT* and *TP53* pathways [11, 12]. These data are generating novel markers for improved risk stratification and identification of therapeutic targets. For example, a subgroup of B-ALL patients demonstrate a gene expression profile referred to as 'Ph-like' that results from CNA and/or mutations of genes in these pathways [11, 13]. Lesions such as those resulting in *CRLF2* overexpression, and rearrangements of *PDGFRB* and *ABL1* have been further associated with sensitivity of the blasts to tyrosine kinase inhibitors [11].

The large discovery studies that have generated the copy number data have primarily been research investigations with centrally performed testing on samples collected from single or multiple institutions or from tumor banks such as that maintained by COG. For each study, a series of patient specimens is evaluated with the same platform and the same statistical algorithms. However, as genomic microarray testing is now being performed in many laboratories for patients referred for constitutional cytogenetic studies, many local laboratories are adding microarray analyses, in addition to G-banding and FISH analyses for studies of patients with malignancies. The consistency and reliability of microarray data generated and analyzed at multiple sites have not yet been determined for COG or other clinical trials groups. Here, we demonstrate that despite the different array platforms being used, the data generated are highly concordant, and feasible for integration with other cytogenetic data to display a more accurate and comprehensive genomic representation of each patient's leukemic clone(s). The different measures were complementary, with G-banding and FISH facilitating interpretation of the array findings with respect to structural rearrangements and expected ratios for the CNA, and the array data providing accurate breakpoints for the unbalanced structural rearrangements, and the regions of copy number neutral loss of heterozygosity (CNLOH) and small CNA that were not detectable by G-banding. As expected, secondary CNA and CNLOH findings were identified in this study that associated with the different primary cytogenetic abnormalities used for risk stratification in AALL08B1. In addition, the array analyses revealed novel findings that raise questions relevant to clinical care and cytogenetic risk stratification.

Thus, genomic microarray analysis at the time of diagnosis performed in local laboratories can yield information not only valuable for diagnosis and monitoring of the specific patient, but also for furthering translational efforts in the field.

Materials and Methods

Clinical samples

Genomic microarray data were obtained on the B-ALL samples as part of routine clinical testing by local laboratories on B-ALL patients enrolled on the COG Risk Classification Study (AALL08B1) during the time period of 2011-2012. All institutions enrolling patients onto the study have had the AALL08B1 protocol approved by their local or central institutional review board, and all enrolled patients were consented to have genomic data

analyzed. Approval for this pilot microarray study was provided by AALL08B1 Study Chairs and endorsed by the Cytogenetics Committee of Children's Oncology Group. Participating laboratories included: ARUP Laboratories (ARUP), Cincinnati Children's Hospital Medical Center (CCHMC), The Children's Hospital of Philadelphia (CHOP) and the University of Minnesota Medical Center (UMMC).

Sixty-five pediatric B-ALL patients enrolled in AALL08B1 who had microarray analyses performed as part of their diagnostic work-up were included in this study. None of the patients had received therapy before the studies were performed. Based on the FISH and G-banded karyotypes, 20 were *ETV6-RUNX1* fusion positive [cryptic t(12;21)], 20 were hyperdiploid with trisomies of 4 and 10, five were hyperdiploid without the double 4 and 10 trisomies, three had a doubling of a near haploid complement, four were positive for *iAMP21*, eight were normal by G-banding and FISH and five had miscellaneous abnormalities not recognized as one of the well documented prognostically significant recurring abnormalities in pediatric B-ALL (Table 1). As there were very limited numbers of patients with MLL rearrangements and *BCR-ABL1* gene fusions enrolled on AALL08B1 and referred to the four laboratories for microarray testing during this study period, those cytogenetic subgroups were not included.

Microarray Processing

DNA was extracted from either fresh bone marrow aspirate or from residual cytogenetic fixed pellets at each participating laboratory and processed for array analysis utilizing one of three platforms (Affymetrix, Agilent or Illumina) following the methods standardized in each of these four CLIA certified laboratories. Briefly, at CCHMC, DNA was extracted using MagnaPure Compact automated extraction kits from Roche, (Indianapolis, IN) following the manufacturer's protocols, and patient DNA was analyzed using the 850 K Illumina Infinium® microarray according to Illumina protocols (Illumina Inc., San Diego, CA). At CHOP, DNA was extracted using a Gentra Puregene kit from Qiagen (Valencia, CA), and patient DNA was analyzed using the 850 K Illumina Infinium® microarray assay (Illumina Inc.) by the Center for Applied Genomics as previously described [10]. At UMMC, DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) following manufacturer's instructions, restriction digested and labeled with fluorochrome Cyanine 5 using random primers and exo-Klenow fragment DNA polymerase. DNA from a single sex-matched control was labeled concurrently with fluorochrome Cyanine 3. The patient and control DNA were combined and microarray analysis was performed with a SurePrint G3 Cancer CGH and SNP 4x180K microarray kit with 110,712 CGH and 59,647 SNP oligos with a 25 Kb overall median probe spacing. The ratio of patient to control DNA for each oligonucleotide was calculated using Feature Extraction software 11.0 (Agilent Technologies, Santa Clara, CA). At ARUP, DNA was extracted using a Gentra Puregene kit from Qiagen, and patient DNA was analyzed using the CytoScan HD® microarray assay (Affymetrix, Santa Clara, CA) which contains 743,304 SNP-based oligonucleotides and 1,953,246 non-polymorphic oligonucleotides and following manufacturer's recommended instructions.

Microarray Data Analysis

Initial analysis—Data were initially analyzed by local laboratories according to their standard operating procedures. Each laboratory utilized the array manufacturer's provided software: Affymetrix (Chromosome Analysis Suite (ChAS) version 1.2), Agilent (Cytogenomics version 2.5) and Illumina (GenomeStudio V2009.2 and BeadStudio). For analysis of Agilent arrays, statistical algorithms included ADM1 and ADM2 with threshold values set at an absolute value of 0.10 on a log2 scale and an LOH threshold of 6.0. For Illumina assays performed at CCHMC, B-allele frequency and log2R ratio were analyzed with Illumina GenomeStudio V2009.2 analysis software. DNA copy number changes were prioritized using output from cnvPartition Plug-in v2.3.4 software. The software identifies LOH based on the presence of homozygosity in the B-allele frequency but no change in the log2R ratio, to exclude regions that are hemizygous due to deletion. ROH that were interrupted by homozygous deletions or genotyping errors were manually adjusted. For Illumina assays performed at CHOP, B-allele frequency and log2R ratio were analyzed by GenomeStudio software as described in [10]. Exceptions included heterozygous or homozygous deletions in known cancer associated genes.

Cancer analyses require careful consideration of ratios and threshold to be used in the determination of copy number calls. Based on preliminary data and discussion, the following criteria were applied for the determination of copy number gains and losses. The minimum number of probes above threshold (for duplication) or below threshold (for deletion) varied for each platform. For Agilent, 3 probes for deletions and duplications, for Affymetrix, 25 probes for deletion and 50 probes for duplication and for Illumina, 15 probes for deletions and duplications. FISH for the primary abnormalities was helpful for generating expectations of the magnitude of the threshold. For example, if FISH from a single case showed 50% of interphase cells to have an *ETV6-RUNX1* fusion, then an associated region of single copy number loss for this patient would be expected to show a maximum ratio of -0.42 (Supplementary Table 1). Evaluation of the allele pattern can provide independent confirmation of a copy number call, detect copy neutral stretches of homozygosity, help distinguish mosaic and likely somatic versus germline alterations and in some cases determine ploidy alterations [14]. Chromosomal regions that included well-documented copy number variants (CNVs) common in the general population were not included in the analysis with two exceptions: biallelic losses involving TCR gamma alternate reading frame protein (*TCRG*) in 7p14.1 and in TCR alpha (*TCRA*) in 14q11.2. Although neither of these findings is considered clinically significant, one or both of these biallelic losses were shown to be present in the majority of the leukemic samples, and thus served as a quality control measure to document that the abnormal clone was being identified. CNAs that involved only intronic or non-gene encoding sequences were also not included as significant for these analyses. All array platforms and genomic positions were based on hg19 human genome sequence build from February, 2009.

Validation by NEXUS—Raw data files were uploaded and validated using NEXUS 6.0 software (BioDiscovery, El Segundo, CA) at UMMC using the mosaicism filter. For arrays originally run on the Illumina and Affymetrix platforms a 25-probe cutoff was used for CNA calls, while for arrays originally run on Agilent, a 3-probe cutoff was used for CNA calls.

NEXUS was utilized to permit consistency across cases in analysis, creation of a single database, and to identify discordances between platforms and/or software. Three independent measures were evaluated in both the local laboratory's analysis and in the validation with NEXUS: CNA, CNLOH, and allelic imbalance/balance.

Group review of findings and resolution of discordant findings—Primary review of the microarray data for each of the four laboratories was performed by the respective laboratory director. NEXUS validation was performed at the UMMC, and any discordant or questionable findings between the NEXUS and local laboratory's findings were discussed by video conference with directors and members of the four laboratories. Consensus was reached among all four directors for all final data entries. For visualization of G-band and array findings, data were prepared in Circos format according to Krzywinski, et al. [15].

Results

ETV6-RUNX1

Twenty cases of *ETV6-RUNX1* B-ALL (Table 1) which met the acceptance criteria by the COG cytogenetics committee were included. In order to ensure that the karyotypes represent the leukemic clone, acceptance of the case in the COG database requires that secondary abnormalities be identified by G-banding or, in cases without G-band detected abnormalities, that the *ETV6-RUNX1* fusion was proven by metaphase FISH to be present in cells that were normal by G-banding. Both the G-band and microarray data are presented in the Circos plot in Figure 1A. Because of the cryptic nature of the *ETV6-RUNX1* fusion, G-banding cannot distinguish between the presence of an extra copy of a normal chromosome 21 and the presence of an extra copy of the derivative chromosome 21, whereas microarray clearly distinguishes these chromosomes. As shown in Figure 1B, among the 10 cases with trisomy 21 identified by G-banding, array analysis confirmed five as having trisomy 21, three as having partial trisomy 21 resulting from the presence of an extra copy of the der(21)t(12;21), and one with partial trisomy 21 due to the presence of an extra copy of 21 with a deletion of its distal long arm. In the remaining case (case 12), the trisomy 21 identified by G-banding was not detected by array, due to the low percentage (10% among metaphases) of cells comprising the subclone with the gain of 21 (Figure 1B and Supplementary Figure 1). Two of the cases with the extra der(21) resulted, as expected, in partial gains of 12p (Figure 2A); the third case had, in addition to the extra der(21), an unbalanced t(9;12) resulting in no net gain for distal 12p. The cases with an extra der(21) had breakpoints in *ETV6* and *RUNX1*, and thus provided information on the intragenic breakpoints (Figure 2A and Supplementary Figure 1). Two of the three cases involved breakpoints between exons 5 and 6 of *ETV6* (Figure 2) and all three had breakpoints between exons 2 and 3 of *RUNX1*.

Among all *ETV6-RUNX1* cases, microarray analysis revealed a total of 114 CNAs with an average of 5.7 CNAs per case. Thirty-five of these were copy number gains, including 13 whole chromosome gains (chromosomes 21,10,16,22, and Y in order of highest to lowest frequency), and 22 were partial gains ranging in size from 90 Kb to 67 Mb.

The majority of the CNAs, (79 of 114), were copy number losses, including one whole chromosome loss of an X. As confirmed by FISH, the most common losses (13 of 20 cases, 65%) involved deletions of *ETV6* on the chromosome 12 homologue that was not involved in the t(12;21) (Figure 2A). These *ETV6* deletions ranged in size from 39 Kb to 27 Mb with heterogeneous breakpoints. The second most common CNA, identified in 7 of 20 *ETV6-RUNX1* cases, (35%) was loss of 9p21.3. Monoallelic losses were identified in all cases, and ranged in size from 129 Kb to 34.3 Mb. Monoallelic losses in *PAX5* were identified in 6 of 20 cases ranging in size from 105 Kb to 808 Kb. Three cases had biallelic losses ranging from 77 Kb to 2.4 Mb, which included *CDKN2A* and/or *CDKN2B*, and were nested within a larger monoallelic loss region (Figure 2B). Biallelic losses within *TCRG* (7p14.1) and *TCRA* (14q11.2) were identified in 80% of cases (Figure 1A). Notably, the single case with loss of an X chromosome also had a very small loss (298 Kb) within Xp22.33, generating a fusion between *CRLF2* and *P2RY8* (Figure 2C).

Three of the *ETV6/RUNX1* positive cases also had what appeared by G-banding to be balanced reciprocal translocations (Table 1). Patient 2 had a t(11;19)(q13;q13.3), patient 8 had a t(10;12)(q22.1;p13) and t(12;13)(q24.3;q14), and patient 9 had three translocations, t(1;3)(p34;q21), t(3;12)(q12;p13) and t(15;16)(q11.2;q22). However, by microarray all of these translocations were associated with losses at one or both of the breakpoints demonstrating that they were actually unbalanced rearrangements (Table 2).

CNLOH was not common among the *ETV6-RUNX1* cases (Figure 1A): 8 had partial regions of CNLOH ranging in size from 5.2 Mb to 20.9 Mb (1p32.3-1p33, 5p13.3-5p14.3, 5q34, 8q11.1-8q11.21, 12p11.21 - 12p13.2 and 12p13.2-12p13.33, both involving a portion of *ETV6*, 17q23.2-17q24.2, and Xp11.3-Xp21.1). There were no cases with whole chromosome CNLOH (Figure 1A).

High hyperdiploidy with trisomy 4 and 10

Twenty cases of favorable prognosis, high hyperdiploid B-ALL (Table 1) with modal numbers of 53-67 and including trisomies of chromosomes 4 and 10 were analyzed by microarray. The frequency distribution of the whole chromosome gains is depicted in Figure 3A and B. The most common gained chromosomes (excluding 4 and 10), were 21 (20 of 20), X (19 of 20), 6 (18 of 20 cases), 14 (18 of 20), 18 (18 of 20) and 17 (14 of 20) (Figure 3B). Recurrent tetrasomies were observed for chromosomes 21 (19 of 20 cases), 14 (5 of 20 cases), 18 (4 of 20 cases), X (3 of 20 cases) and 10 (2 of 20 cases). Fifty-five percent of cases had, in addition to whole chromosome gains or losses, structural abnormalities that were visible by G-banding and 40% had additional CNA abnormalities visible by array only. 55 CNA with an average of 2.8 CNAs per case were identified (not including whole chromosome gains): 33 deletions (not including *TCRG* and/or *TCRA*), 21 gains, and one chromosomal amplification involving a region of 16p encompassing the *ERCC4 (FANQ)* gene (Figure 3A). Similar to the *ETV6-RUNX1* category, twelve (60%) of the hyperdiploid cases had biallelic losses within *TCRG* and/or *TCRA* (Figure 3A).

Among the monoallelic losses, notable were two patients with losses that encompassed *IKZF1*: one spanned 17.9 Mb from band 7p12.1 to 7p14.2 and the other 57.5 Mb, from band 7p11.2 to 7p22.3. Three patients had monoallelic (2 of 20) or biallelic (1 of 20) loss of

CDKN2A/B ranging in size from 39 Kb to 536 Kb for the monoallelic loss and 132 Kb for the biallelic loss.

Analysis of the SNP data tracks from the microarrays yields information on allelic balance. The genotype of a normal diploid cell is generally heterozygous and maintained in allelic balance (i.e. two different homologues represented) whereas loss or gain of a single allele and CNLOH are forms of allelic imbalance (unequal representation of two different homologues). Trisomy for a chromosome typically shows allelic imbalance whereas tetrasomy can be present in either allelic balance or imbalance. Among the trisomy 4 and 10 hyperdiploid cases, 34 tetrasomies (involving chromosomes 21, 14, 18, X, 10 and 4) were detected; 31 were in allelic balance (two copies each of two homologues) and three (two cases of tetrasomy 21, and one of tetrasomy 10) were in allelic imbalance (one chromosome preferentially gained) (Figure 3A).

CNLOH was common among the favorable hyperdiploid group (Figure 3A, C); 25% of these cases had one to three whole chromosomes demonstrating CNLOH. Chromosomes 9, 15 and 16 were most frequently involved (Figure 3C).

Hyperdiploidy without trisomy 4 and 10

Five cases were classified as variant hyperdiploid because the clones had modal numbers of 49 to 63, but did not include both trisomy 4 and trisomy 10 (Table 1). The most common trisomy observed in this group was a gain of chromosome 17, which was identified in all five cases, followed by gains of chromosomes 18, X and 21 in 4 of the 5 cases. The frequency distribution of chromosomal gains in this group is presented in Figure 3B. Similar to the hyperdiploid trisomy 4 and 10 group, two cases had tetrasomy of chromosome 21 in allelic balance, one case had tetrasomy of chromosome 10, one case had tetrasomy of chromosome 14 and one had tetrasomy of chromosome X. In addition to the whole chromosome gains, a total of 25 CNA were observed including 10 gains, 13 monoallelic losses, 2 biallelic losses (not including *TCRG* or *TCRA*), 5 whole chromosome CNLOH and one partial chromosomal region of CNLOH with a range of zero to four whole chromosome CNLOH per specimen (Figure 3C). Similar to the trisomy 4 and 10 hyperdiploid leukemias, the chromosomes with CNLOH (i.e. chromosomes 2, 9, 13, 15 and 22) were not those commonly seen in trisomic state.

Three (60%) cases had *CDKN2A* and/or *CDKN2B* losses; one with a monoallelic loss (1.8 Mb), one with only a biallelic loss (2.0 Mb) and one with a biallelic loss of *CDKN2A* and *CDKN2B* (890 kb) nested within a 7.4 Mb loss of 9p21.2 to 9p22.1. As was seen in the *ETV6-RUNX1* cases, a single case had a region of loss (349 Kb) within Xp22.33 generating a *CRLF2-P2RY8* fusion. Biallelic losses within *TCRG* and/or *TCRA* were common (3 of 5 cases). The distribution of losses in this category is presented in Figure 4.

Normal karyotypes by G-banding and FISH

Eight cases were classified as normal by G-banding and had negative interphase FISH results for *BCR-ABL1*, *ETV6-RUNX1*, *MLL* and trisomies 4 and 10 (Table 1). All eight cases had abnormalities identified by the array (Figure 5A) with an average of 5.6 CNAs per case. Only one region of CNLOH was found, involving 9p13.3 to 9p21.3, proximal to the

loss of *CDKN2A/B*. Of the total 45 CNAs detected, 41 (91.1%) were losses including 34 monoallelic and seven biallelic (not including *TCRG* or *TCRA*) and four (8.9%) were gains (Figure 5A). Four (50%) of the eight cases had losses involving *IKZF1* (Figure 5A-B and Figure 6). These losses included all or portions of *IKZF1*; in one case with loss within 7p12.1 to 7p12.2, neighboring genes including *FIGNL1*, *DDC*, *GRB10* and *COBL* were also deleted encompassing a total region of 917 Kb (Figure 5B). Another case had a biallelic loss (77 kb) within *IKZF1* embedded within a larger (87 Kb) region of monoallelic loss. The remaining monoallelic losses ranged in size from 124 Kb to 222 Kb with an average of 173 Kb within 7p12.2. Thus, these losses were not due to cytogenetically visible deletions of the short arm of chromosome 7 or monosomy 7. The profile of losses among cases with *IKZF1* deletions and those without differed (Supplementary Table 2). On average, each case with an *IKZF1* deletion had an additional 6.5 monoallelic and/or biallelic losses (not including *TCRG* and *TCRA*) whereas each case with no *IKZF1* deletion had only 2.5 monoallelic and/or biallelic losses (not including *TCRG* or *TCRA*)

The second most common monoallelic loss involved *PAX5* (3 of 8 cases) ranging in size from 270 Kb to 439 Kb. Similar to the *ETV6-RUNX1* group, four (50%) of the cases had monoallelic (2 of 8) or biallelic (2 of 8) loss of *CDKN2A* and *CDKN2B*. All eight of the normal G-band cases had biallelic losses within *TCRG* and/or *TCRA* (Figure 4A).

iAMP21

There were four cases of iAMP21. By G-banding, the iAMP21 manifested as an abnormal chromosome 21 usually described by the local laboratory as an add(21)(q22) with the exception of one that appeared to be an isoderivative (21)(p11.2) (Table 1, Figure 7A). For these cases, FISH showed amplification of the *RUNX1* signal ranging from 5 to 20 copies (Figure 7B-C). As depicted in Figure 7D, the amplification was not limited to the *RUNX1* gene, but rather spanned a contiguous region ranging in size from 19 Mb to 32 Mb. The start points were diverse, with case 44 showing the most proximal start bp 14,338,286 in 21q11.2 and case 41 showing the most distal start at bp 22,559,711 in 21q21.1. The degree of amplification was not equal across the region; however, in all cases, the peak amplification included *RUNX1*. The array did not provide an exact estimate of the number of copies of *RUNX1*. Notable in each of these iAMP21 cases was the associated deletion of the distal long arm of the involved chromosome 21. This deletion extended from 41 Mb-46 Mb (21q22.2-22q22.3) to the 21q telomere, with recurring breakpoints between base pair 41,406,543-45,922,853 (Figure 7D) involving genes *DSCAM* (case 41), *AIRE* (case 43) and *TSPEAR* (case 44). Deletions within proximal 21q were also observed in case 43. In addition to the amplification of *RUNX1* and deletion of distal 21q, there were two cases with monoallelic loss of *RBI* ranging in size from 149 Kb to 181 Kb, and one case with monoallelic loss of *CDKN2A*, *CDKN2B* and *PAX5* in a region of loss spanning 17.7 Mb (Figure 4A). All four of the iAMP21 cases had biallelic losses within *TCRG* and/or *TCRA*. A single region of CNLOH was found, involving 12q24.12-12q24.32.

Doubled near haploid

Three cases had modal numbers of 46 to 61 that represented doubling of near haploid clones. The cases were determined to represent doubling of a near haploid clone rather than

a hyperdiploid clone by the high number of whole chromosome LOH present and, in some cases, the unusual pattern of gains that included 4, 9, 10, 14, 18, 21 and X. The chromosomes present in at least two copies always included chromosomes 14, 18, 21 and X (Figure 8A) and these chromosomes did not demonstrate CNLOH nor did chromosomes 4, 9 or 10 (Figure 8B). However, all of the other chromosomes displayed CNLOH, a hallmark of near haploid cases, as depicted in Figure 8D using NEXUS software. Thus, the chromosomes that are most frequently gained (4, 10, 14, 18, 21 and X) in the favorable +4 and +10 high hyperdiploid leukemias are not among those prone to CNLOH in the doubled near haploid cases. Each of these near haploid cases had, in addition to whole chromosome losses and gains, a total of 6 CNA including loss of 10p13-10p15.3 (12.5 Mb), gain of 1q21.1-1q32.2 (64.7 Mb), and four biallelic losses (not including *TCRG* or *TCRA*) of 5q31.3 (18 Kb), 8q21.13 (113 Kb), 15q26.2 (123 Kb) and 17q11.2 (19 Kb including the *NFI* gene) (Figure 4A). No focal losses of *CDKN2A/B* or *TP53* were identified. While two of the three cases had two chromosome 7s and all three cases had a gain of two extra chromosome 14s, none had biallelic deletions of *TCRG* or *TCRA* (Figure 4A).

Karyotypes with atypical abnormalities

Five cases had abnormalities identified by G-banding that did not represent the major recurring primary abnormalities in pediatric ALL and all of these were negative by FISH for *ETV6-RUNX1* or *MLL* cryptic rearrangements (Table 1, cases 61-65). They were thus designated as “atypical abnormalities” for this study. Among these cases, 33 CNA were identified with a range of 3 to 11 CNA per case, and an average of 6.6 CNA per case. These included 6 gains, 24 monoallelic losses, 2 biallelic losses (not including *TCRG* or *TCRA*) and one region of multi copy gain of approximately 4 additional copies within 9p21.3 to p22.3. Three cases had losses within 9p involving *CDKN2A/B* or *PAX5*. The region of multi copy gain within 9p was located proximal to a biallelic region of loss (6 Mb) that included *CDKN2A/B* embedded within a larger monoallelic 21 Mb loss of 9p12-p21.3 (21 Mb) that included *PAX5*. Two cases had losses involving *IKZF1*, one 71 Kb loss limited to 7p12.2 resulting in loss of exons 4-7 and the other spanning 11.2 Mb in bands 7p11.2 to 7p12.3, including all of *IKZF1*. As shown in the Circos diagram in Figure 4 and Table 3, the two cases with *IKZF1* deletions each had other CNA including loss of *CDKN2A/B*, *PAX5* and *TP53* in one case, and 21q22.2 including *ERG* in the other. In addition to the CNA, 17 regions of partial chromosome CNLOH were identified, within one case with multiple regions of CNLOH within chromosome X. Four of the 5 cases had biallelic losses within *TCRG* and/or *TCRA*.

DISCUSSION

The primary goal of this study was to evaluate the feasibility of combining genomic microarray results derived from assays performed and analyzed in multiple laboratories with different array platforms and software, with integration of other cytogenetic data from G-banding and FISH assays. Each of the three array platforms from Agilent, Affymetrix and Illumina successfully detected a range of CNAs and regions of CNLOH, from small deletions, such as those involving only portions of *IKZF1* or *ETV6*, to whole chromosome CNLOH. Each platform had different strengths and weaknesses; for example, the Agilent

platform had superior copy number calls when the clonal fraction was greater than 30% while Affymetrix and Illumina were superior at detecting lower clonal fractions by utilizing the B allele frequency. Critical to the detection of these abnormalities was the establishment of analytic criteria that differed significantly from those utilized in constitutional (germline) work. This primarily involved lowering the thresholds (Supplementary Table 1) required for calling an aberration based on the percent involvement by the abnormal clone, and by the number of probes used to make an individual call, based on both the size of the genes and the number of probes targeting that gene in a given platform. For example, a partial deletion of *IKZF1* could easily be missed if applying “constitutional” algorithms to the evaluation of these cases due to the small size of the region (the gene spans only 128 Kb) and involvement of less than 100% of cells harboring the deletion. For lower density arrays, it is critical for the laboratory to know the copy number and SNP probe coverage of specific regions.

The unbiased validation of all of the cases in the present study using NEXUS was helpful in revealing both the strengths and limitations of the different platforms. Overall, there was high consensus of calls made by NEXUS and by the platform specific software tools. In some cases, NEXUS was more helpful in identifying biallelic losses of *TCRG*, whereas these were not highlighted by the manufacturer's specific software. This is because the *TCRG* region was masked in the local Affymetrix software. In contrast, NEXUS failed to call a CNA involving BTG1 (patient 14) that the local Illumina software made by relying on the SNP tracks and the allelic ratios. With the exception of the *TCRG* region, the failure of NEXUS to call the CNA of the BTG1 gene represented the only discrepancy between NEXUS and the local software.

Although evaluation of the allele pattern and log2 values can help distinguish somatic versus germline alterations, many of the aberrations identified in these analyses were assumed to be somatic. When the clonal fraction of the tumor approaches 100%, it is difficult to distinguish between somatic versus germline alterations. In some situations germline analysis of a noninvolved tissue, or analysis of blood at the time of remission may be necessary to rule out a potential rare germline aberration.

The concordance of patterns of secondary abnormalities observed within the *ETV6-RUNX1*, +4,+10 hyperdiploid, variant hyperdiploid, and normal subgroups support the conclusion that such multi-laboratory generated data are both reliable and consistent and should be integrated with other forms of cytogenetic testing data to permit more comprehensive and more accurate characterization of leukemic clones. Ideally, such analyses should be performed concurrently with the G-band and FISH studies, to enable full integration at the time of initial diagnosis. Review of the karyotypes and corresponding array results revealed several hyperdiploid cases with misclassified derivative and whole chromosomes by G-banding, even when those aberrations encompassed more than 18 Mb of DNA. This is largely reflective of the poor chromosome morphology that is characteristic of some of the leukemic clones.

There are now multiple large published series of investigations that have addressed the presence of secondary abnormalities in some of the same cytogenetic subgroups investigated here. The patterns of secondary abnormalities generated in the present study confirm the

patterns observed in previously reported single institution or single platform studies [4-10], reinforcing the conclusion that data collected from different platforms with different software can yield consistent and reliable results. For example, the frequency of gained chromosomes among the +4,+10 hyperdiploids, is essentially the same as that observed in prior studies [16, 17]. Similarly, the pattern of amplification with accompanying deletion for the iAMP21 cases reported here is consistent with those reported in the literature [18-25] and the pattern of CNAs including *IKZF1* deletions seen in the present “normal” karyotype cases is concordant with the pattern of gains previously reported in similar patients [6].

FISH, genomic microarray, and karyotype analysis serve complementary roles in the characterization of leukemic clones

For all primary abnormalities that can be detected by FISH (e.g. *ETV6-RUNX1*, +4,+10 hyperdiploid, iAMP21), the interphase FISH frequencies were very helpful in establishing the expected ratios (Supplementary Table 1). The interphase FISH data provided a quantitative measure of the percentage of tumor cells in the specimen. Although estimates of tumor burden can be obtained from the report of the hematopathology examination of the bone marrow and/or immunophenotyping, the percentage of leukemic cells present in the sample received in the cytogenetics laboratory can be significantly higher or lower, depending on whether the cytogenetics sample was obtained on the first, second, or later draw of the aspirate from the patient. The true percentage of leukemic cells provides guidelines for array analysis, as thresholds may need to be adjusted from the -1.0 for losses and +0.58 for gains (on a log2 scale) that are expected if the CNAs are present in 100% of cells. Supplementary Table 1 provides a simple guide for converting the percentage of leukemic cells to expected ratios of loss and gain.

FISH also proved particularly valuable in determining amplification levels of *RUNX1* in the iAMP21 cases. High amplification levels cannot be accurately estimated using microarrays, particularly when there is heterogeneity from cell to cell. In contrast, the array data provided a more sensitive method for detecting deletions, as the small sizes of some of these deletions were below the sensitivity of detection with commercial FISH probes.

As depicted in Figure 2A, 13 cases with an *ETV6-RUNX1* fusion had deletions that encompassed all or portions of the homologous *ETV6* gene. The size of the deletions ranged from 39 Kb to 27 Mb and encompassed a region as small as exon 1 to the entire *ETV6* gene, up to and including the *KRAS* gene. Commercially available probes such as the *ETV6-RUNX1* ES fusion probe from Abbott Molecular that effectively detects the fusion rearrangement, spans approximately 350 Kb in and around the *ETV6* gene. Thus, detection of small deletions encompassing only portions of *ETV6* will not be possible with this probe set. The same is true for *CDKN2A* and *CDKN2B* deletions which were detected in all the primary subgroups, with the exception of the near haploid doubled cases. For these genes, monoallelic losses spanned 129 Kb to 34.3 Mb, and biallelic deletions from 77 Kb to 2.4 Mb. The commercial probe from Abbott Molecular for example, spans 222 Kb, and thus the small losses will not be detectable with these reagents. In contrast, they were readily detected by all of the array platforms used in this study.

Standard karyotype analysis using G-banding adds particular value in those cases in which there are small subclones. For example, an evolved clone with trisomy 21 in the *ETV6-RUNX1* case #12, and the clone with the t(2;16)(p11.2;p11.2) in case #61 were both present in only 10% of metaphases. If the clone with this t(2;16) had a proliferative advantage, its frequency among metaphase cells could be significantly higher than among interphase cells. Karyotypes are therefore helpful in determining subclone hierarchy and composition that can frequently be difficult to determine from the array data. Furthermore, G-banding analysis provides information regarding the structural abnormality that leads to the CNA. Such is the case with the gains of proximal 21q and distal 12p generated by the presence of the extra derivative chromosome 21 in some of the *ETV6-RUNX1* cases (Supplemental Figure 1). As the short arms of the acrocentric chromosomes are not included on arrays, due to the highly repetitive DNA encompassed within them, G-banding is informative for characterizing derivative chromosomes involving these regions.

Most of the apparent balanced translocations identified by G-banding were unbalanced

Recurrent translocations in the hematologic malignancies typically result in the juxtaposition of two genes that causes deregulation of one of these genes, or generates a chimeric gene with an abnormal gene product. As suggested by the observation of unique (not known recurrent) translocations in the *ETV6-RUNX1* group, the significant ramifications may be in the loss of genes contiguous to the breakpoints rather than the formation of a critical fusion product. The genes encompassed by the losses associated with the secondary translocations in the *ETV6-RUNX1* group included *ARID5B*, *ETV6* and *LTK*. Of note, one other apparently balanced translocation was detected in this study; a t(2;16)(p11.2;p11.2) in the subgroup with atypical cytogenetic abnormalities. No loss or gain near the breakpoint regions was detected by array, which could reflect it being a truly balanced rearrangement or the losses being too small or being present at frequencies too low to be detected by array.

There is an inverse relationship between the identity of chromosomes gained in the hyperdiploid cases and the constellation involved in CNLOH

As described in the results, and depicted in Figure 3, among the +4,+10 hyperdiploid cases, the chromosomes that most frequently appear in trisomic or tetrasomic state (chromosomes 4, 6, 10, 14, 17, 18 and 21) did not appear in CNLOH state. Conversely, chromosomes such as 9, 15, and 16 which were the most frequent chromosomes displaying CNLOH were rarely, if at all, in a trisomic state in the +4,+10 hyperdiploid cases. Of particular interest in this investigation were the “variant” hyperdiploid cases without the double trisomies for 4 and 10. As illustrated in Figure 3B, the pattern of gains was similar but notably, chromosome 17 was the most commonly gained among the variant hyperdiploid cases. In the cases with doubling of a near haploid, the same autosomes that most frequently occurred in trisomic or tetrasomic state in the +4,+10 hyperdiploid cases, also were present in heterozygous state in the doubled near-haploid leukemias (Figure 8A). Similar to the +4,+10 hyperdiploid cases, the chromosomes that were gained were not those most commonly observed with CNLOH.

The significance of multiple whole chromosomes with CNLOH among non-doubled hypodiploid cases is not known

It has now been well documented [26-28] that the presence of a near haploid or doubled near haploid clone is a high risk, poor prognostic indicator in pediatric B-lineage ALL. However, the significance of the presence of multiple whole chromosome CNLOH in a hyperdiploid clone that is not a doubled near haploid leukemia is unknown. Is the total number of CNLOH of significance, or rather the specific chromosomes involved in the CNLOH? Does the presence of multiple whole chromosome CNLOH underlie any of the heterogeneity in outcome that is observed among the patients with +4,+10 hyperdiploid clones? Array analyses on a larger series of hyperdiploid cases, including both variant and +4,+10 hyperdiploid cases are needed to address this clinically relevant question.

***IKZF1* deletions vary in frequency, size, and associated constellation of CNAs between subgroups**

IKZF1 deletions have been identified as a negative prognostic indicator in pediatric ALL [29]. In the present analysis, *IKZF1* deletions were identified in two (10%) of the +4, +10 hyperdiploids cases, 4 (50%) of the cases with normal karyotypes, and two (40%) of the cases with atypical abnormalities (Table 3). None of these cases had CNLOH of chromosome 7 or other chromosomes. The deletions observed in the karyotypically normal cases were smaller than those in the hyperdiploid cases, in that they were relatively focal, ranging in size from 77 Kb to 917 Kb, whereas the hyperdiploid *IKZF1* losses ranged in size from 17.9 Mb to 57.5 Mb encompassing all of the gene and contiguous regions of 7p. Of the two *IKZF1* deletions detected among the cytogenetically atypical leukemias, one was 71 Kb in size resulting in a deletion of exons 4 to 7 and the other was 11.2 Mb that included *IKZF1* and contiguous proximal and distal genes. Interestingly, in the normal cases, the *IKZF1* losses were typically seen in constellation with an average of 7 other abnormalities, that frequently included monoallelic losses of *PAX5* (2/4) and *CDKN2A/B* (2/4). In the +4, +10 hyperdiploid cases, the *IKZF1* losses were typically seen with an average of 2 other abnormalities that did not include *PAX5* nor *CDKN2A/B*. Similarly, the smaller 71 kb *IKZF1* deletion identified in the group with atypical cytogenetics, also had losses of *CDKN2A* and *PAX5* as well as a loss in *ERG* [30] whereas the larger 11.2 Mb deletion did not. Deletions of *IKZF1* have been, in some cases, associated with mutations of tyrosine kinase receptor and cell signaling pathways [11, 31], and/or the gene expression profile termed “*BCR-ABL* like” [8, 13]. Both the size of the *IKZF1* deletion and the constellation of associated CNAs may be predictive of such mutations and gene expression profiles and thus have relevance for potential targeted therapies.

The most frequent genes involved in CNAs (other than *TCRG* and *TCRA*) were *ETV6*, *CDKN2A* or *CDKN2B*, *PAX5*, *IKZF1*, *CRLF2*, *P2RY8* and *RB1*

As expected, loss of *ETV6* was most frequent among the *ETV6-RUNX1* cases (Table 4), as this is well documented to be the most common secondary abnormality for this subtype of pediatric ALL [32, 33]. Lower frequencies of *ETV6* deletions were seen among all other subgroups, with the exception of the near haploid cases, which may simply be due to the fact that only three near haploid cases were analyzed in this cohort. The variant hyperdiploid and

cytogenetically normal cases had the highest frequency of *CDKN2A/CDKN2B* deletions and the atypical and cytogenetically normal cases had the highest frequency of *PAX5* loss. In contrast, as discussed previously, *IKZF1* deletions were not seen among all subgroups, neither were those involving *CRLF2* and *P2RY8* or *RB1* (Table 4).

Whole chromosomes displaying CNLOH were almost exclusively found in the doubled near haploid cases, and in the hyperdiploid cases. Partial chromosome CNLOH that involved regions > 10 Mb were observed in all groups except for the doubled near haploid tumors. One case each of the iAMP21, the atypical, and the cytogenetically normal cases had such CNLOH regions. These regions were more frequent among the *ETV6-RUNX1* (4 cases) and hyperdiploid (both +4,+10 (5 cases) and the variant hyperdiploids (1 case) tumors. With one exception, all of these cases had one to three different chromosomes involved in these CNLOH regions; the one exception was the single case from the atypical group that had 11 different chromosomes with 17 total regions of CNLOH involving 5 – 55 Mb of the genome. The mechanism giving rise to such multiple regions is not known. It may be that some of these were germline, generated by identity by descent.

The chromosomes most frequently involved in whole chromosome CNLOH were 15,16,19 and 3, while partial chromosomal regions of CNLOH (>10 Mb) most frequently involved chromosomes 9 and 12.

Contiguous regions of CNLOH and copy number loss recur in leukemic cells

Of interest, in 11 cases (Figure 6), were the findings of the stretches of CNLOH juxtaposed to regions of copy number loss, and the embedding of small regions of biallelic loss within larger regions of monoallelic loss. As illustrated in Figure 6, this most frequently occurred in the short arm of chromosome 9 involving biallelic deletion of *CDKN2A/CDKN2B* and the adjacent regions of monoallelic loss and CNLOH. Such regions are likely generated by mitotic crossover events. Similarly, in case #5, there was a 12.0 Mb region of CNLOH encompassing 12p13.2 to 12ptel adjacent to a region of monoallelic loss encompassing bands 12p11.22 to 12p13.2 with breakpoints in *ETV6*. FISH showed a signal pattern consistent with the presence of an *ETV6-RUNX1* fusion and two derivative #12 chromosomes. The CNLOH likely resulted from a somatic crossover event near the centromeres of the derivative #12 and the normal chromosome 12, generating CNLOH for the short arm of 12, and monoallelic loss of the region of 12p distal to the *ETV6* fusion breakpoint. (Of note, the loss is not biallelic because there is a derivative chromosome 21 that contains one copy of this distal 12p.) This finding of juxtaposed CNLOH and copy number loss is not a common occurrence in constitutional (germline) cases with chromosomal deletions or UPD.

Microarrays reveal the presence of non-favorable findings among otherwise highly favorable subgroups

Among cases with favorable cytogenetic findings, there were some with secondary abnormalities identified by array that have been proposed to have negative prognostic value. These included deletions within the pseudoautosomal region of X that resulted in *CRLF2-P2RY8* fusions, and deletions in 7p12 involving all or part of *IKZF1*. The *CRLF2-P2RY8*

fusion, which has been shown to deregulate *CRLF2*, has been suggested to be associated with more high risk features and constitutive JAK signaling [11, 34, 35]. Similarly, deletions of *IKZF1* have been shown to be associated with increased rates of relapse among patients with high risk features [29, 36]. What is the prognostic significance of these abnormalities when they are identified in a case that by other cytogenetic and clinical criteria would be classified as low risk? Only with larger sample sizes and outcome data will these questions be addressed.

Analyzing microarray data concomitantly with G-banding and FISH enables a more complete interpretation of each technique's results than could be achieved independently. As noted above, this can lead to identification of clinically relevant secondary abnormalities that are unexpected, or novel. It can elucidate patterns of association between primary and secondary abnormalities, and between different secondary abnormalities that would otherwise not be detectable. As such, utilization of these techniques together may identify factors that contribute to heterogeneity in outcome among groups classified as low risk as well as those classified as high risk. None of the current primary cytogenetic subgroups of pediatric ALL demonstrate 100% concordance with patient outcome. Further, despite the highly informative findings of large genomic studies of pediatric ALL, there remain important unanswered questions. The results of the present study support integrating microarray, G-band and FISH data being collected as part of diagnostic evaluation by local laboratories, to more accurately and comprehensively characterize leukemic clones and enable the accrual of larger numbers of cases to address the clinically relevant questions raised in the present study. The authors believe there is significant value added by genomic cancer array analyses, and encourage the incorporation of such data, when available, into cooperative group and/or public cancer databases with integration of cytogenetic, FISH and molecular sequencing data. Such comprehensive genomic findings will facilitate patient care and basic and translational research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by grants to the Children's Oncology Group U10 CA98543 (Chair's grant) and to NIH/NCI P30 (CA77598) University of Minnesota Comprehensive Cancer Center: Cytogenomics Shared Resource Core Facility. K.R.R. is a St. Baldrick's Foundation Scholar. We thank Donna Wilmoth and Laura Tooke for technical assistance.

References

1. Greaves M. The causation of childhood leukemia: a paradox of progress? *Discovery medicine*. 2006; 6:24–8. [PubMed: 17234124]
2. Hunger SP, Loh ML, Whitlock JA, Winick NJ, Carroll WL, Devidas M, Raetz EA, Committee COGALL. Children's Oncology Group's 2013 blueprint for research: acute lymphoblastic leukemia. *Pediatric blood & cancer*. 2013; 60:957–63. [PubMed: 23255467]
3. Mullighan CG. Molecular genetics of B-precursor acute lymphoblastic leukemia. *The Journal of clinical investigation*. 2012; 122:3407–15. [PubMed: 23023711]

4. Kuiper RP, Schoenmakers EF, van Reijmersdal SV, Hehir-Kwa JY, van Kessel AG, van Leeuwen FN, Hoogerbrugge PM. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia*. 2007; 21:1258–66. [PubMed: 17443227]
5. Strefford JC, Worley H, Barber K, Wright S, Stewart AR, Robinson HM, Bettney G, van Delft FW, Atherton MG, Davies T, Griffiths M, Hing S, Ross FM, Talley P, Saha V, Moorman AV, Harrison CJ. Genome complexity in acute lymphoblastic leukemia is revealed by array-based comparative genomic hybridization. *Oncogene*. 2007; 26:4306–18. [PubMed: 17237825]
6. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui CH, Relling MV, Evans WE, Shurtleff SA, Downing JR. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007; 446:758–64. [PubMed: 17344859]
7. Kawamata N, Ogawa S, Zimmermann M, Kato M, Sanada M, Hemminki K, Yamamoto G, Nannya Y, Koehler R, Flohr T, Miller CW, Harbott J, Ludwig WD, Stanulla M, Schrappe M, Bartram CR, Koeffler HP. Molecular allelokaryotyping of pediatric acute lymphoblastic leukemias by high-resolution single nucleotide polymorphism oligonucleotide genomic microarray. *Blood*. 2008; 111:776–84. [PubMed: 17890455]
8. Harvey RC, Mullighan CG, Wang X, Dobbin KK, Davidson GS, Bedrick EJ, Chen IM, Atlas SR, Kang H, Ar K, Wilson CS, Wharton W, Murphy M, Devidas M, Carroll AJ, Borowitz MJ, Bowman WP, Downing JR, Relling M, Yang J, Bhojwani D, Carroll WL, Camitta B, Reaman GH, Smith M, Hunger SP, Willman CL. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*. 2010; 116:4874–84. [PubMed: 20699438]
9. Simons A, Stevens-Kroef M, El Idrissi-Zaynoun N, van Gessel S, Weghuis DO, van den Berg E, Waanders E, Hoogerbrugge P, Kuiper R, van Kessel AG. Microarray-based genomic profiling as a diagnostic tool in acute lymphoblastic leukemia. *Genes, chromosomes & cancer*. 2011; 50:969–81. [PubMed: 21882283]
10. Dougherty MJ, Wilmoth DM, Tooke LS, Shaikh TH, Gai X, Hakonarson H, Biegel JA. Implementation of high resolution single nucleotide polymorphism array analysis as a clinical test for patients with hematologic malignancies. *Cancer genetics*. 2011; 204:26–38. [PubMed: 21356189]
11. Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X, Chen SC, Payne-Turner D, Churchman ML, Harvey RC, Chen X, Kasap C, Yan C, Becksfort J, Finney RP, Teachey DT, Maude SL, Tse K, Moore R, Jones S, Mungall K, Birol I, Edmonson MN, Hu Y, Buetow KE, Chen IM, Carroll WL, Wei L, Ma J, Kleppe M, Levine RL, Garcia-Manero G, Larsen E, Shah NP, Devidas M, Reaman G, Smith M, Paugh SW, Evans WE, Grupp SA, Jeha S, Pui CH, Gerhard DS, Downing JR, Willman CL, Loh M, Hunger SP, Marra MA, Mullighan CG. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer cell*. 2012; 22:153–66. [PubMed: 22897847]
12. Mullighan CG, Zhang J, Harvey RC, Collins-Underwood JR, Schulman BA, Phillips LA, Tasian SK, Loh ML, Su X, Liu W, Devidas M, Atlas SR, Chen IM, Clifford RJ, Gerhard DS, Carroll WL, Reaman GH, Smith M, Downing JR, Hunger SP, Willman CL. JAK mutations in high-risk childhood acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:9414–8. [PubMed: 19470474]
13. Den Boer ML, van Slegtenhorst M, De Menezes RX, Cheok MH, Buijs-Gladdines JG, Peters ST, Van Zutven LJ, Beverloo HB, Van der Spek PJ, Escherich G, Horstmann MA, Janka-Schaub GE, Kamps WA, Evans WE, Pieters R. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *The lancet oncology*. 2009; 10:125–34. [PubMed: 19138562]
14. Kearney HM, Kearney JB, Conlin LK. Diagnostic implications of excessive homozygosity detected by SNP-based microarrays: consanguinity, uniparental disomy, and recessive single-gene mutations. *Clinics in laboratory medicine*. 2011; 31:595–613, ix. [PubMed: 22118739]

15. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome research*. 2009; 19:1639–45. [PubMed: 19541911]
16. Paulsson K, Forestier E, Lilljebjorn H, Heldrup J, Behrendtz M, Young BD, Johansson B. Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:21719–24. [PubMed: 21098271]
17. Heerema NA, Raimondi SC, Anderson JR, Biegel J, Camitta BM, Cooley LD, Gaynon PS, Hirsch B, Magenis RE, McGavran L, Patil S, Pettenati MJ, Pullen J, Rao K, Roulston D, Schneider NR, Shuster JJ, Sanger W, Sutcliffe MJ, van Tuinen P, Watson MS, Carroll AJ. Specific extra chromosomes occur in a modal number dependent pattern in pediatric acute lymphoblastic leukemia. *Genes, chromosomes & cancer*. 2007; 46:684–93. [PubMed: 17431878]
18. Sinclair PB, Parker H, An Q, Rand V, Ensor H, Harrison CJ, Strefford JC. Analysis of a breakpoint cluster reveals insight into the mechanism of intrachromosomal amplification in a lymphoid malignancy. *Human molecular genetics*. 2011; 20:2591–602. [PubMed: 21487021]
19. Rand V, Parker H, Russell LJ, Schwab C, Ensor H, Irving J, Jones L, Masic D, Minto L, Morrison H, Ryan S, Robinson H, Sinclair P, Moorman AV, Strefford JC, Harrison CJ. Genomic characterization implicates iAMP21 as a likely primary genetic event in childhood B-cell precursor acute lymphoblastic leukemia. *Blood*. 2011; 117:6848–55. [PubMed: 21527530]
20. Heerema NA, Carroll AJ, Devidas M, Loh ML, Borowitz MJ, Gastier-Foster JM, Larsen EC, Mattano LA Jr, Maloney KW, Willman CL, Wood BL, Winick NJ, Carroll WL, Hunger SP, Raetz EA. Intrachromosomal amplification of chromosome 21 is associated with inferior outcomes in children with acute lymphoblastic leukemia treated in contemporary standard-risk children's oncology group studies: a report from the children's oncology group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2013; 31:3397–402. [PubMed: 23940221]
21. Robinson HM, Harrison CJ, Moorman AV, Chudoba I, Strefford JC. Intrachromosomal amplification of chromosome 21 (iAMP21) may arise from a breakage-fusion-bridge cycle. *Genes, chromosomes & cancer*. 2007; 46:318–26. [PubMed: 17243167]
22. Moorman AV, Richards SM, Robinson HM, Strefford JC, Gibson BE, Kinsey SE, Eden TO, Vora AJ, Mitchell CD, Harrison CJ. Party UKMRCNCRICLW. Prognosis of children with acute lymphoblastic leukemia (ALL) and intrachromosomal amplification of chromosome 21 (iAMP21). *Blood*. 2007; 109:2327–30. [PubMed: 17095619]
23. Strefford JC, van Delft FW, Robinson HM, Worley H, Yiannikouris O, Selzer R, Richmond T, Hann I, Bellotti T, Raghavan M, Young BD, Saha V, Harrison CJ. Complex genomic alterations and gene expression in acute lymphoblastic leukemia with intrachromosomal amplification of chromosome 21. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:8167–72. [PubMed: 16702559]
24. Harrison CJ, Moorman AV, Schwab C, Carroll AJ, Raetz EA, Devidas M, Strehl S, Nebral K, Harbott J, Teigler-Schlegel A, Zimmerman M, Dastuge N, Baruchel A, Soulier J, Auclerc MF, Attarbaschi A, Mann G, Stark B, Cazzaniga G, Chilton L, Vandenberghe P, Forestier E, Haltrich I, Raimondi SC, Parihar M, Bourquin JP, Tchinda J, Haferlach C, Vora A, Hunger SP, Heerema NA, Haas OA. An international study of intrachromosomal amplification of chromosome 21 (iAMP21): cytogenetic characterization and outcome. *Leukemia*. 2013
25. Li Y, Schwab C, Ryan SL, Papaemmanuil E, Robinson HM, Jacobs P, Moorman AV, Dyer S, Borrow J, Griffiths M, Heerema NA, Carroll AJ, Talley P, Bown N, Telford N, Ross FM, Gaunt L, McNally RJ, Young BD, Sinclair P, Rand V, Teixeira MR, Joseph O, Robinson B, Maddison M, Dastugue N, Vandenberghe P, Haferlach C, Stephens PJ, Cheng J, Van Loo P, Stratton MR, Campbell PJ, Harrison CJ. Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature*. 2014
26. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L, Payne-Turner D, Churchman M, Andersson A, Chen SC, McCastlain K, Becksfort J, Ma J, Wu G, Patel SN, Heatley SL, Phillips LA, Song G, Easton J, Parker M, Chen X, Rusch M, Boggs K, Vadodaria B, Hedlund E, Drenberg C, Baker S, Pei D, Cheng C, Huether R, Lu C, Fulton RS, Fulton LL, Tabib Y, Dooling DJ, Ochoa K, Minden M, Lewis ID, To LB, Marlton P, Roberts AW, Raca G, Stock W, Neale G, Drexler HG,

- Dickins RA, Ellison DW, Shurtleff SA, Pui CH, Ribeiro RC, Devidas M, Carroll AJ, Heerema NA, Wood B, Borowitz MJ, Gastier-Foster JM, Raimondi SC, Mardis ER, Wilson RK, Downing JR, Hunger SP, Loh ML, Mullighan CG. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nature genetics*. 2013; 45:242–52. [PubMed: 23334668]
27. Safavi S, Forestier E, Golovleva I, Barbany G, Nord KH, Moorman AV, Harrison CJ, Johansson B, Paulsson K. Loss of chromosomes is the primary event in near-haploid and low-hypodiploid acute lymphoblastic leukemia. *Leukemia*. 2013; 27:248–50. [PubMed: 22889820]
  28. Harrison CJ, Moorman AV, Broadfield ZJ, Cheung KL, Harris RL, Reza Jalali G, Robinson HM, Barber KE, Richards SM, Mitchell CD, Eden TO, Hann IM, Hill FG, Kinsey SE, Gibson BE, Lilleyman J, Vora A, Goldstone AH, Franklin IM, Durrant J, Martineau M. Childhood, Adult Leukaemia Working P. Three distinct subgroups of hypodiploidy in acute lymphoblastic leukaemia. *British journal of haematology*. 2004; 125:552–9. [PubMed: 15147369]
  29. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, Ma J, Liu W, Cheng C, Schulman BA, Harvey RC, Chen IM, Clifford RJ, Carroll WL, Reaman G, Bowman WP, Devidas M, Gerhard DS, Yang W, Relling MV, Shurtleff SA, Campana D, Borowitz MJ, Pui CH, Smith M, Hunger SP, Willman CL, Downing JR, Children's Oncology G. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *The New England journal of medicine*. 2009; 360:470–80. [PubMed: 19129520]
  30. Clappier E, Auclerc MF, Rapon J, Bakkus M, Caye A, Khemiri A, Giroux C, Hernandez L, Kabongo E, Savola S, Leblanc T, Yakouben K, Plat G, Costa V, Ferster A, Girard S, Fenneteau O, Cayuela JM, Sigaux F, Dastugue N, Suciu S, Benoit Y, Bertrand Y, Soulier J, Cave H. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia*. 2014; 28:70–7. [PubMed: 24064621]
  31. Loh ML, Zhang J, Harvey RC, Roberts K, Payne-Turner D, Kang H, Wu G, Chen X, Becksfort J, Edmonson M, Buetow KH, Carroll WL, Chen IM, Wood B, Borowitz MJ, Devidas M, Gerhard DS, Bowman P, Larsen E, Winick N, Raetz E, Smith M, Downing JR, Willman CL, Mullighan CG, Hunger SP. Tyrosine kinome sequencing of pediatric acute lymphoblastic leukemia: a report from the Children's Oncology Group TARGET Project. *Blood*. 2013; 121:485–8. [PubMed: 23212523]
  32. Forestier E, Andersen MK, Autio K, Blennow E, Borgstrom G, Golovleva I, Heim S, Heinonen K, Hovland R, Johannsson JH, Kerndrup G, Nordgren A, Rosenquist R, Swolin B, Johansson B, Nordic Society of Pediatric H, Oncology, Swedish Cytogenetic Leukemia Study G, Group NLCS. Cytogenetic patterns in ETV6/RUNX1-positive pediatric B-cell precursor acute lymphoblastic leukemia: A Nordic series of 245 cases and review of the literature. *Genes, chromosomes & cancer*. 2007; 46:440–50. [PubMed: 17285576]
  33. van Delft FW, Horsley S, Colman S, Anderson K, Bateman C, Kempinski H, Zuna J, Eckert C, Saha V, Kearney L, Ford A, Greaves M. Clonal origins of relapse in ETV6-RUNX1 acute lymphoblastic leukemia. *Blood*. 2011; 117:6247–54. [PubMed: 21482711]
  34. Harvey RC, Mullighan CG, Chen IM, Wharton W, Mikhail FM, Carroll AJ, Kang H, Liu W, Dobbin KK, Smith MA, Carroll WL, Devidas M, Bowman WP, Camitta BM, Reaman GH, Hunger SP, Downing JR, Willman CL. Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. *Blood*. 2010; 115:5312–21. [PubMed: 20139093]
  35. Mullighan CG, Collins-Underwood JR, Phillips LA, Loudin MG, Liu W, Zhang J, Ma J, Coustan-Smith E, Harvey RC, Willman CL, Mikhail FM, Meyer J, Carroll AJ, Williams RT, Cheng J, Heerema NA, Basso G, Pession A, Pui CH, Raimondi SC, Hunger SP, Downing JR, Carroll WL, Rabin KR. Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia. *Nature genetics*. 2009; 41:1243–6. [PubMed: 19838194]
  36. Kuiper RP, Waanders E, van der Velden VH, van Reijmersdal SV, Venkatachalam R, Scheijen B, Sonneveld E, van Dongen JJ, Veerman AJ, van Leeuwen FN, van Kessel AG, Hoogerbrugge PM. IKZF1 deletions predict relapse in uniformly treated pediatric precursor B-ALL. *Leukemia*. 2010; 24:1258–64. [PubMed: 20445578]

Figure 1. Cytogenetic aberrations within *ETV6-RUNX1* cases

A. Abnormalities identified by G-banding are shown in track A (one copy gain, blue) and track B (one copy loss, red). Abnormalities identified by microarray shown in track C (CNLOH, yellow), track D (one copy gain, blue), track E (one copy loss, red) and track F (two copy loss, black). The *t*(12;21) is depicted by the purple curved line in the center of the diagram. B. Enlarged Circos plot of chromosomes 20, 21, 22, X and Y.

Figure 2. CNAs within chromosome arms 12p and 9p of *ETV6-RUNX1* cases
 UCSC genome browser images of CNAs within *ETV6* in A and *CDKN2A/B* in B. 1 copy (monoallelic) loss, red; 2 copy (biallelic) loss, green; 1 copy gain, blue; CNLOH, yellow. C. Microarray results of the monoallelic and biallelic loss of Xp22.33.

Figure 3. Cytogenetic aberrations within trisomy 4 and 10 hyperdiploid leukemias

A. Circos plot of the cytogenetic aberrations within the trisomy 4 and 10 hyperdiploid patient group. Abnormalities identified by microarray shown in track A (CNLOH, yellow), track B (chromosomal amplification, purple), track C (two copy gain in allelic balance, green), track D (two copy gain in allelic imbalance, orange), track E (one copy gain, blue), track F (one copy loss, red) and track G (two copy loss, black). B. Comparison of the whole chromosome gains and C whole chromosome CNLOH within the trisomy 4 and 10 hyperdiploid cases (black bars) and the variant hyperdiploid without trisomy 4 and 10 cases (grey bars).

Figure 4. Comparison of monoallelic and biallelic losses between patient groups

A. Circos plot of the losses identified by microarray within the *ETV6-RUNX1*, hyperdiploid, variant hyperdiploid, normal G-banding, Atypical, iAMP21 and near haploid patient groups. Tracks are labeled within the Circos figure. Monoallelic losses with the exception of those in chromosomes 9p and 7p are red, biallelic losses are black, monoallelic losses within 9p are purple and monoallelic losses within 7p are orange.

Figure 5. Cytogenetic aberrations within normal G-banding patient groups

A. Abnormalities identified by G-banding are shown in track A (one copy gain, blue) and track B (one copy loss, red). Abnormalities identified by microarray shown in track C (CNLOH, yellow), track D (one copy gain, blue), track E (one copy loss, red) and track F (two copy loss, black). B. UCSC genome browser images of CNAs within *IKZF1*.

Figure 6. Multiple regions of CNAs are consecutive within chromosomes 7, 9 and 12
UCSC genome browser images of CNAs within *IKZF1* (top), *CDKN2A/B* (middle) and *ETV6* (bottom).

Figure 7. iAMP21 cases have unique G-banding, FISH and array findings

A. G-banding of chromosome 21 from four iAMP21 cases. Interphase FISH in B and metaphase FISH in C using *ETV6-RUNX1* ES probe (*ETV6*, green; *RUNX1*, red) and microarray in D shows chromosome 21 with blue (gains) and red (losses).

Figure 8. Recurring copy whole chromosome gains and CNLOH in near haploid cases
Whole chromosome gains in A and the whole chromosome CNLOH in B within the near haploid cases. Whole genome view of a trisomy 4 and 10 hyperdiploid case in C and a near haploid case in D.

Table 1

B-ALL patients included in this study.

Cytogenetic abnormality	ISCN	Abnormal clone %	Patient
<i>ETV6-RUNX1</i>	47,XY,add(6)(q25),+21[7]/46,XY[13]	64%	1
	46,XXY,t(11;19)(q13;q13.3)[5]/46,XY[15]	84%	2
	46,XX,add(13)(q14)[15]/46,XX[6]	71%	3
	47,XX,+21[2]/48,idem,+10[2]/46,XX[8]	86%	4
	46,XY,der(9;12)(q10;q10),+21[3]/46,XY[17]	72%	5
	46,XXY,del(12)(p12)[3]/46,idem,+21[1]/46,XY[16]	95%	6
	45,X,-X[13]/46,XX[2]	90%	7
	46,XXY,der(10)t(10;12)(q22.1;p13),der(12)t(10;12)(q22.1;p13)t(12;13)(q24.3;q14),der(13)t(12;13)(q24.3;q14)[5]/46,XY[15]	80%	8
	48,XY,t(1;3)(p34;q21)t(3;12)(q12;p13),+10,t(15;16)(q11.2;q22),+21[cp16]/46,XY[4]	75%	9
	46,XXY,i(6)(p10)[3]/46,XY[17]	41%	10
	46,XX,+21,der(21;22)(q10;q10)x2,+22[6]/47,XX,add(21)(p11.2),+mar[2]/46,XX[12]	91%	11
	47,XX,+21[3]/46,XX[27]	87%	12
	46,XXY,del(9)(p13p24)[2]/46,XY[10]	52%	13
	47,XXY,?add(12)(q11.2)[3],+21[cp4]/46,XY[35]	50%	14
	48,XX,+10,+21[5]	38%	15
	46,XX,-9,add(12)(p11.2),+mar[cp19]/46,XX[12]	88%	16
	46,XX,add(5)(q33),add(6)(p21),add(13)(q14)[17]/46,XX[8]	80%	17
	46,XXY,del(12)(p11.2)[8]/46,XY[12]	86%	18
	46,XXY,t(12;21)(p13;q22),+der(21)t(12;21)[10]/46,XY[13]	98%	19
	46,XXY,t(12;21)(p13;q22)/46,XY[13]	70%	20
	55,XXY,X,+4,+6,+8,+10,+14,+18,der(19)(1;19)(q21;p13.3),+21,+21[6]/55,idem,del(6)(q15q25),+19,-der(19)[3]/46,XY[11]	84%	21
	65-67,XXX,-3,-7,+8,-9,+10,-13,-15,-17,+21,+2mar[cp4]/46,XX[16]	81%	22
	55,XXY,X,der(1)?dup(1)(q12q42)?inv(1)(p36q21),+4,+6,add(9)(p?1),+10,+14,+17,+18,+21,+21[4]/56,idem,+18[2]/46,XY[14]	97%	23
	65,XX,+X,-Y,del(1)(p21p36.3),-2,-3,+4,-9,-13,-19,-20,+21[11]/46,XY[9]	53%	24
	58,XXY,dup(1)(q21.1q32.2),+4,+6,+8,+10,+11,+14,+17,add(17p11.2),+18,+18,+21,+21[10]/46,XY	42%	25
	55,XXY,X,+4,+6,+10,+11,+14,+18,+21,+21[2]/46,XY[20]	64%	26

Cytogenetic abnormality	ISCN	Abnormal clone %	Patient
	53.XX,+4,+6,+10,+17,+18,+21,-21[6]/46,XX[14]	72%	27
	56.XY,+X,+4,+6,+10,+17,+21,+21,+3mar[5]/46,XY[15]	95%	28
	57-58,XX,+X,+4,+6,+7,+8,+10,+14,+17,+18,+19,+21,+21[cp7]/46,XX[11]	96%	29
	57.XY,+X,+Y,+4,+6,+9,+10,+11,+14,+18,+21,+21[11]	85%	30
	57.XY,+X,+4,+5,+6,+8,+10,del(10)(q24)+12,+14,+17,+18,-21[2]/46,XY[4]	96%	31
	46.XY[14]/60-61,XY,+X,+4,+del(5)+6,+7,+9,+10,+11,+12,+14,+15,+16,+17,+18,+21,+21[6]	91%	32
	55-59<2n>,XX,+X,+X,+4,+5,+6,i(7)(q10),+9[7],+10[6],+11[3],add(12)(p11.2),+14,+17,+18,+21,-21,mar[4][cp10]/46,XX[10]	86%	33
	56.XY,+X,+4,+6,+8,+10,+14,+17,+21,+21,+mar[15]/46,XY[5]	67%	34
	54.XY,+X,+4,+6,?del(7)(p13p15),+10,+14,+17,+21,-21[10]/46,XY[10]	85%	35
	58,XX,+X,+X,+4,+6,+8,+10,+10,+11,+14,add(17)(p11.2),+18,+21,+21[7]/46,XX[21]	17%	36
	56.XY,+X,+4,+8,+9,+10,+14,+17,+18,+21,+21[9]/46,XY[16]	89%	37
	55.XY,+X,add(1)(q41),+4,+6,+10,+14,+14,+17,+18,+21[3]/56,idem,add(5)(q35),+21[9]/57,idem,add(5)(q35),+8,+21[2]/46,XXY[9]	97%	38
	58,XX,+X,+4,+5,+6,+10,+11,-12,+14,+17,+18,+21,+21[4]/57,idem,del(11)(q14.1),-12[13]	85%	39
	59,XX,+X,+X,+4,+5,+6,+10,+14,+14,+17,+18,+21,+21,+22[7]/60,idem,+9[11]/46,XX[2]	98%	40
iAmp21	46,XX,r(19)(p13.3q13.4),der(21)add(21)(q22)amp(RUNX1)[17]/46,XX[3]	98%	41
	46.XY,add(21)(q22)amp(RUNX1)[8]/46,XY[12]	57%	42
	46,XX,der(21)add(21)(p11.2)add(21)(q22)amp(RUNX1)[12]/46,XX[8]	91%	43
	46,XY,del(7)(q22q32),t(12;14)(q24.1q11.2),add(21)(q22)amp(RUNX1)[16]/46,XY[4]	95%	44
	46,XY[30]	NA	45
Normal	46,XX[26]	NA	46
	46,XX[17]	NA	47
	46,XY[20]	NA	48
	46,XX(30)	NA	49
	46,XY[20]	NA	50
	46,XX[20]	NA	51
	46,XX	NA	52
	59-63<2n>,XX,+X,+X,+3,+5,+6,+10,+10,+11,+12,+14,+16,+17,+18,+21,+21,+1-3mar[4][cp7]/46,XX[23]	87% (RUNX1)	53
	49,XX,+X,+17,+21[7]/46,XX[13]	77% (RUNX1)	54
	53.XY,+X,+4,+6,+14,der(15)t(5;15)(q32;q26),+17,+18,+21[7]/54,idem,+21[4]/46,XY[9]	95% (RUNX1)	55

Cytogenetic abnormality	ISCN	Abnormal clone %	Patient
	54,XX,+X,+4,+6,+14,+17,+18,+21,+21[10]/55,idem,+14[5]/56,idem,+9,+14[6]/46,XX[3]	96% (<i>RUNX1</i>)	56
	55-57,XY,+X,add(1)(q42),+4,+7,+8,+10,+12,+14,+17,+18,+19,+21,+21[8]/46,XY[112]	77% (<i>RUNX1</i>)	57
Doubled Near haploid	58(31,X,+X,+4,+8,+9,+10,+14,+18,+21)x2,-4,-8,-9,-18[3]/58,idem,del(10)(q23.3)[11]/46,XX[9]	NA	58
	27<n>,X,+X,+14,+18,+21[9]/50,idemx2,-X,-X,-18[10]/28,idem,+9[2]/56,idemx2,+9,+9[6]/46,XX[4]	NA	59
	61(31,X,+X,t(2;10)(q31;q21.2)c,+4,+9,+10,+14,+17,+18,+21)x2,der(X)t(X;1)(q13;q21),del(1)(p32),-6,der(9)(9;17)(q12;q11.2)[11]/46,XX,t(2;10)(q31;p1.2)c[18]	NA	60
	46,XY,t(2;16)(p11.2;p11.2)[3]/46,XY[27]	NA	61
Atypical	46,XY,dup(14)(q32q32)[6]/46,XY[29]	NA	62
	46,XX,der(19)t(12;19)(p13;p13.3)[8]/46,XX[12]	NA	63
	46,XX,+X,psu dic(7;9)(p12;p13),del(12)(q13q15)(17)(q10,add(20)(q11.2)[10]/47,idem,+mar[2]/46,XX[8]	NA	64
	46,XX,der(5)del(5)(q11.2q12.1)del(5)(q22.1q32)[8]/46,XX[13]	NA	65

Percentage of abnormal clone identified by FISH based on *ETV6-RUNX1* fusion probe for the *ETV6-RUNX1* and iAMP categories, CEP 4 and CEP 10 FISH for the +4,+10 Hyperdiploid category. Percentage of abnormal clone by FISH was not reported for the Normal, Doubled Near Haploid and Atypical categories.

Table 2

Genomic location and potential genes of interest near the translocation breakpoints.

Case	Translocation	Location	Genomic coordinates	Size	Genes
2	t(11;19)(q13;q13.3)	11q12.3q13.1	63051354-65326832	2.3 Mb	<i>MEN1</i>
		19q13.32q13.33	46763711-48907252	2.1 Mb	
8	t(10;12)(q22.1;p13)	10q21.2q21.3	62483228-64889397	2.4 Mb	<i>ARID5B</i>
		12p12.3p13.1	13673704-15362296	1.7 Mb	Proximal to <i>ETV6</i>
		12q24.32	126203446-127640666	1.4 Mb	
		t(12;13)(q24.3;q14)	50191673-53108361	2.9 Mb	
		13q14.2q14.3			
9	t(3;12)(q12;p13)	12p13.2	11999075-12188503	189 Kb	<i>ETV6</i>
		15q15.1	40873226-42150645	1.3 Mb	<i>LTK</i>
		16q22.1	67544084-68343839	800 Kb	

Table 3

Comparison of monoallelic and biallelic losses with normal G-banding cases with or without *IKZF1* deletions.

Primary abnormality group	IKZF1* chr7:50344378-50472798	Size	Monoallelic loss	Biallelic loss	Gains	CN LOH	Case
+4, +10 Hyperdiploid	Complete (M)	7p12p14.2 (17.9 Mb)	3p22.3 (<i>ARPP-21</i>)	-	-	-	30
	Complete (M)	7p11.2p22.3 (57.5 Mb)	12p13.1p13.31 (<i>ETV6</i>)				
			22q12.2 (<i>NF2</i>)	-	7q	-	33
			3p14.2 (<i>PTPRG</i>)	12p13.2 (<i>ETV6</i>)			
			6q25.1 (<i>PLEKHG1</i>)	13q14.2 (<i>Rb1</i>)	-	-	47
			7q11.21				
			9p21.3 (<i>CDKN2A/B</i>)				
	Complete (M)	7p12.2 (222 Kb)	9p11.2p13.1				
			9q12q21.11				
			10q25.1 (<i>ADD3</i>)				
			20q11.21q11.22 (<i>DNMT3B</i>)				
			Xp22.33 (<i>CRLF2-P2RY8</i>)				
			Xp11.3 (<i>EFHC2</i>)				
Normal G-banding			2q34 (<i>IKZF2</i>)				
	Complete (M)	7p12.1p12.2 (917 Kb)	4q25 (<i>LEF1</i>)	-	2q13	-	48
			7p15.2 (<i>SNXI10</i>)		6q22.31 (<i>STL</i>)		
	Complete (M)	7p12.2 (124 Kb)	9p13.2 (<i>PAX5</i>)	-	-	-	49
			3p26.1 (<i>EDEM1</i>)	3q13.2 (<i>BTLA/CD200</i>)			
			9p21.3 (<i>CDKN2A</i>)	9p21.3 (<i>CDKN2A</i>)			
			9p13.2 (<i>PAX5</i>)	13q14.11 (<i>SERP2</i>)	-	-	50
	Deletion of exons 2-3 (B)	7p12.2 (M) (87 Kb) 7p12.2 (B) (77 Kb)	10q25.1 (<i>ADD3</i>)				
			11q14.1 (<i>RAB30</i>)				
			12p12.1p13.2 (<i>ETV6</i>)				
Atypical			13q14.11 (<i>SMIM2</i>)				
	Deletion of exons 4-7 (M)	7p12.2 (M) (71 Kb)	14q32.33				
			21q22.2 (<i>ERG</i>)	-	7p21.3	-	62

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Primary abnormality group	IKZF1* chr7:50344378-50472798	Size	Monoallelic loss	Biallelic loss	Gains	CN LOH	Case
	Complete (M)	7p11.2p12.3 (M) (11.2 Mb)			14q32.2q32.33		
			17p (<i>TP53</i>)	9p22.3p23	17q		
			7p21.3p22.3	9p21.2p21.3 (<i>CDKN2A/B</i>)	9p21.3p22.3	-	64
			7p15.3p21.3				
			9p23p24.3 (<i>AK2</i>)				
			9p12p21.3 (<i>CDKN2A/B,PAX5</i>)				

IKZF1 transcript variant 1*, monoallelic deletion (M), biallelic deletion (B)

Table 4

Comparison of frequently occurring CNAs within ALL subgroups.

Primary abnormality group	Percentage of losses (monoallelic and biallelic) and gains						
	<i>ETV6</i>	<i>CDKN2A or CDKN2B</i>	<i>PAX5</i>	<i>IKZF1</i>	<i>CRLF2/P2RY8</i>	<i>RB1</i>	1q gain
ETV6-RUNX1 (n=20)	13/20 (65%)	7/20 (35%)	6/20 (30%)	0/20 (0%)	1/20 (5%)	0/20 (0%)	0/20 (0%)
Hyperdiploid (n=20)	1/20 (5%)	3/20 (15%)	0/20 (0%)	2/20 (10%)	0/20 (0%)	0/20 (0%)	5/20 (25%)
Variant hyperdiploid (n=5)	0/5 (0%)	3/5 (60%)	0/5 (0%)	0/5 (0%)	1/5 (20%)	0/5 (0%)	1/5 (20%)
Normal G-banding (n=8)	2/8 (25%)	4/8 (50%)	3/8 (38%)	4/8 (50%)	1/8 (13%)	1/8 (13%)	0/8 (0%)
iAMP21 (n=4)	1/4 (25%)	1/4 (25%)	1/4 (25%)	0/4 (0%)	1/4 (25%)	2/4 (50%)	0/4 (0%)
Near Haploid (n=3)	0/3 (0%)	0/3 (0%)	0/3 (0%)	0/3 (0%)	0/3 (0%)	0/3 (0%)	1/3 (33%)
Atypical (n=5)	1/5 (20%)	2/5 (40%)	2/5 (40%)	2/5 (40%)	0/5 (0%)	0/5 (0%)	0/5 (0%)