Targeted resequencing of 9p in acute lymphoblastic leukemia yields concordant results with array CGH and reveals novel genomic alterations

Virinder Kaur Sarhadi a,1, Leo Lahti b,1, Ilari Scheinin a,c, Anne Tyybäkinova a, Suvi Savola a, Anu Usvasalo d, Riikka Räty e, Erkki Elonen e, Pekka Ellonen f, Ulla M. Saarinen-Pihkala d, Sakari Knuutila a,∗

a Department of Pathology, Haartman Institute and HUSLAB, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland
b Department of Veterinary Bioscience, University of Helsinki, Finland and Laboratory of Microbiology, Wageningen University, The Netherlands
c Department of Pathology, VU University Medical Center, Amsterdam, the Netherlands
d Children’s Hospital, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland
e Department of Hematology, Helsinki University Central Hospital, Helsinki, Finland
f Institute for Molecular Medicine Finland, University of Helsinki, Biomedicum Helsinki 2U, Helsinki, Finland

ARTICLE INFO

Article history:
Received 25 September 2012
Accepted 8 January 2013
Available online xxxx

Keywords:
Next generation sequencing
Chromosome 9
9p deletion
aCGH
Copy number variations

ABSTRACT

Genetic alterations of the short arm of chromosome 9 are frequent in acute lymphoblastic leukemia. We performed targeted sequencing of 9p region in 35 adolescent and adult acute lymphoblastic leukemia patients and sought to investigate the sensitivity of detecting copy number alterations in comparison with array comparative genomic hybridization (aCGH), and besides, to detect novel genomic anomalies. We found a high concordance of copy number variations (CNVs) as detected by next generation sequencing (NGS) and aCGH. By both methodologies, the recurrent deletion at CDKN2A/B locus was identified, whereas NGS revealed additional, small regions of CNVs, seen more frequently in adult patients, while aCGH was better at detecting larger CNVs. Also, by NGS, we detected novel structural variations, novel SNVs and small insertion/deletion variants. Our results show that NGS, in addition to detecting mutations and other genetic aberrations, can be used to study CNVs.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

In acute lymphoblastic leukemia (ALL), one of the most recurrently altered chromosomal regions is 9p containing numerous neoplasia-associated genes such as JAK2 at 9p24.1, PAX5 at 9p13.2, as well as CDKN2A, CDKN2B, MTAP, INI genes, MLL3, and PTPLAD2 at 9p21.3. These genes have been shown to be affected in ALL by homozygous and heterozygous deletions, loss of heterozygosity (LOH), mutations, promoter hypermethylation, and translocations [1–7]. Various mechanisms associated with these 9p alterations, such as the role of fragile sites, miRNAs and methylation clusters, and also sequence homology with antigen receptor genes in causal connection with rearrangement formation, have been widely studied [3–5]. The methods used for the previous studies include mainly karyotyping and fluorescence in situ hybridization (FISH) for fusion genes, array comparative genomic hybridization (aCGH) and SNP arrays for copy number changes, and direct sequencing and PCR for gene mutations.

Next generation sequencing (NGS) enables to study not only gene mutations but also copy number variations (CNVs), structural rearrangements and fusion genes in one experiment. Hence, utilizing a series of 35 patients with acute lymphoblastic leukemia, previously studied for copy number alterations by array comparative genomic hybridization (aCGH), we performed genomic analysis of the 9p region in these patients by 9p-targeted resequencing. Our principal aim was to compare the results of copy number changes by the two methods, and besides, to reveal novel copy number changes, gene mutations, fusions, and their frequencies in ALL.

2. Results

2.1. Copy number variations (CNV) by NGS

CNV analysis of 35 ALL cases showed copy number losses in 26 patients in the 9p21.3 region. The size of the deleted region varied considerably between the samples, ranging from smallest 0.3 kb (exon 2 of CDKN2A) in ALL-17 and ALL-34, to the entire target region. The most recurrent loss within 9p21.3, seen in 26 cases, was a 34 kb, (chr9:21,970,856-22,005,115) region (Fig. 1), harboring the CDKN2A/B gene. Of the 26 patients, 22 had homozygous deletions within this region (Table 1).

2.1.1. Comparison of NGS and aCGH data for CNV detection

The results observed with NGS and aCGH are presented in Table 1. In addition to the commonly deleted 9p21.3 region seen by both methods,
NGS analysis identified smaller CNVs not detectable by aCGH (Fig. 1). However, while NGS was more sensitive in detecting smaller deletions it was unable to detect large/whole-arm deletions, which were, in this study, determined on the basis of the aCGH data, as described in Section 5.4.

2.1.2. CDKN2A/B locus

By both NGS and aCGH methods, 26 out of a total of 35 (74%) showed similar results of copy number analysis at the 9p21.3 region (Table 1). Of these concordant cases, 6 patients had normal copy number while one patient had gain of 9p. In the remaining 19 patients, the deletion size of CDKN2A/B was similar by both methods, albeit two of these patients differed regarding the homozygous/heterozygous status of CDKN2A/B deletion. The results of 5 patients with large part of chromosome 9p deleted in aCGH data, were inconsistent between aCGH and NGS. On the other hand, in 3 patient, NGS analysis detected a small deletion (<50 kb) that was not seen with aCGH.

2.1.3. Recurrent CNVs seen only by NGS and mainly in adult ALL

Besides the frequently deleted CDKN2A/B region, there were other less recurrent CNVs, seen more frequently in adult ALL patients and only with NGS. The prominent regions included: chr:9:5,111,498-5,113,815 bp (R1, Fig. 1) with loss within this region seen in 8 patients (5 adult and 3 adolescent ALL) and a gain in 2 patients; loss within chr:9:19,102,475-19,121,201 region (R2, Fig. 1) in 10 patients (4 adolescent and 6 adult); loss within chr:9:33,288,771-33,524,540 (R3, Fig. 1) seen in 12 (2 adolescent; 10 adult ALL); a loss within chr:9:36,881,901-37,036,509 region (R4, Fig. 1) seen in 9 (3 adolescent, 6 adult ALL) and a gain in 3 patients.

2.2. Single nucleotide variations (SNVs)

Tumor DNA from patients showed, on an average, 1340 variants within the target region on 9p for each patient. SNVs observed in all 35 ALL patients were sorted, and those SNVs that were non-coding or resulted in synonymous change were recorded in dbSNP 135 were filtered out. Finally, 30 novel non-synonymous SNVs were detected (Table 2). While most of these variations were seen in individual patients, there were 4 recurrent SNVs seen in FAM214B, PIGO, PRSS3, and RUSC2 (present in 2 patients). Overall, the genes, ADAMTS11, PIGO or RUSC2 were affected in at least three patients by non-synonymous SNVs.

Please cite this article as: V.K. Sarhadi, et al., Targeted resequencing of 9p in acute lymphoblastic leukemia yields concordant results with array CGH and reveals novel genomic alterations, Genomics (2013), http://dx.doi.org/10.1016/j.ygeno.2013.01.001
2.3. Insertion/deletions

On average, 342 indel variants were observed in the target region of each patient. Compared to SNVs there were very few indels detected in the coding region of the genes and 9 novel non-synonymous indels (not recorded in dbSNP135) were observed (Table 3). Majority of the novel indels detected in ALL patients were in 3'-UTR of the genes. Seven novel indels were seen in 5'-UTR region of genes, of which 4 were seen in 5'-UTR of PAX5 gene. Four of the 9 novel indels lead to frameshift change. Two novel frameshift indels were identified in PAX5 gene.

2.4. Structural rearrangements/paired-end anomalies

We found 12 different kinds of paired-end anomalies mapping within the target region (Table 4). Two types of recurrent pericentric inversions, with a common break-point at 33.52 Mb, located 1 kb upstream of ANKRD18B gene in our 9p target region were detected. The other mate pair breakpoint was in 9q arm either at 66.49 Mb (8 patients) or at 99.98 Mb (13 patients). Four paracentric inversions of 9p were detected with the smallest insert size for paired-end anomaly of 100 kb and largest of 111.8 Mb. Three of these rearrangements included PAX5 gene, while one of them had a breakpoint in MIR31HG and included the often deleted CDKN2A gene. As for the 2 deletions within 9p region, the smaller one, which was detected in two patients (no. 6 in Table 4), was merely 30 kb and included only CDKN2A/CDKN2B genes, whereas the other one was a deletion of 7 Mb (no. 8, Table 4) involving many genes. Both deletions were verified by CNV analysis of NGS and aCGH data.

In our study we detected four inter-chromosomal rearrangements (Table 4), of which the most recurrent one was seen in 3 patients and involved chromosomes 9p and 7p, with a breakpoint in HAUS6 gene on 9p. Other chromosomal rearrangements involved FANCG on 9p and 16q (Fig. 2); 9p and NACC1 on 19p, as well as IFT74 on 9p and EDA on Xq.
The commonly deleted CDKN2A/B region detected by aCGH was seen by NGS in all except one (ALL12) patient. However, large deletions could not be detected by NGS since the large deletions covered nearly the entire sequenced region, making it challenging to control for within-sample variation. The limitation of NGS in not being able to identify large deletions is mainly because we studied only a small region of 32 Mb on chromosome 9 with a target capture of 1 Mb. CNVs larger than 13.85 Mb that could not be detected in our study, constituted nearly 43% of the target region. We anticipate that, in analogy to aCGH, inclusion of target regions from other chromosomes in target design could be used to improve the identification of larger deletions with NGS.

On the other hand, NGS had a higher sensitivity of detecting small size CNVs that could not be detected by aCGH (ALL17, ALL34 and ALL39, Table 1). For one of these patients (ALL17), we used 44 K array and for the other two (ALL34 and ALL39), 244 K arrays. In two of these patients (ALL17 and ALL34), NGS was able to detect single exon deletion of CDKN2A. The single exon deletions detected by NGS includes exon 2 of CDKN2A. This is the only coding region of this gene common to both, p16INK4 (CDK inhibitor) and p14ARF (stabilizer of p53) proteins encoded by this gene. Since the exon 2 is translated in different reading frames, deletion of exon 2 could lead to the disruption of both the reading frames and affect the function of both the genes. Even though these genes are structurally and functionally very different, they both are important for cell cycle G1 control by regulating CDK4 and p53 respectively.

In addition to the deletion of CDKN2A/B region, NGS revealed four regions of CNV (Fig. 1). These were not detected by aCGH. Notably,
many of the genes therein are involved in cell proliferation and hematopoiesis. The first region (R1) included part of the intronic region of *JAK2* gene and a pseudogene. According to the Database of Genomic Variations (DGV), this region has been affected by a CNV (both loss/gain) common in normal individuals. The second region (R2), included the 5' UTR and first exon of *HAUS6* as well as a complete *PLIN2* gene. A large 4.56 Mb region gain (Variation:52762), is reported in 2/2026 healthy individuals. The region (R3), which is centromeric to *CDKN2A/B*, harbored *NFIX, AQP7, AQP3, NOL6* and *SUGTI1P1* genes, with NFIX being the most frequently deleted gene within this region. Copy number gains and losses in genes *AQP7, AQP3 and NOL6* are reported in healthy individuals, especially more commonly in *AQP7* while there are no reports of CNV for *NFIX* gene. As for the most centromeric region (R4), it harbored *PAX5* and *MIR4476* genes. Copy number loss (chr9:36438586–37407033) corresponding to R4 region is reported in a single case from 1000 Genomes Project and somatic rearrangements involving *PAX5* are well known in ALL [7]. In our patients, *PAX5* deletion was more frequently seen among adult patients (46%) than in the group of adolescents (14%). Deletion of *PAX5* has been earlier reported in 29.7% of B-progenitor ALL [8], in 33% of adolescent and adult ALL [9], as well as in 60% of adult and childhood ALL patients with 9p abnormality [6]. In our study the frequency of CNV as whole was higher in adult than that in adolescent ALL. However because of lack of normal DNA available from these patients, it is difficult to know the acquired or germline nature of these variations.

Since 9p is frequently deleted in ALL, SNVs in hemizygous condition could have deleterious effects. Notably, our NGS analyses revealed several novel single nucleotide variations. However recurrent mutations in *CDKN2A* were not detected, which is in accordance with earlier studies [10]. In contrast, in *PAX5*, which is often mutated in ALL, we found 3 novel frameshift indels/non-synonymous mutations, in addition to novel indels in 5’-UTR region in 9 patients. Furthermore, novel mutations were seen in *RUSC2* (4 patients), *PIGO* (3 patients) and *PRSS3* gene (3 patients). *PIGO* is involved in GPI-anchor protein synthesis and is expressed in different kinds of blood cells [11] and *PRSS3* is implicated in cancer, specifically in metastasis [12–15].

As for structural rearrangements involving 9p, they showed, in our study, frequent pericentric inversions with breakpoint at 9p13.3 and at either 9q13 or 9q22.33. The constitutional pericentric inversion 9p11–12 and 9q13–21.1 is one of the most common balanced structural chromosomal variations known to occur in 1–4% of the normal population, and its predisposition to development of cancer and haematological disorders has been indicated. Acquired somatic form of 9p pericentric inversions with breakpoints at 9p13 and 9q12 is also reported in leukemia [16,17].

We detected in one patient structural rearrangement between chromosomes 9 and 16 involved exons 6–7 of *FANCC* (Fig. 2). It is worth mentioning that a higher incidence of leukemia is well known in patients with germ line mutations in *FANCC* compared to other *FANC* genes [18], but although mutations and indels are well known for this gene, structural variation involving this gene is a new finding. Furthermore, our analysis, revealed in 3 patients a structural rearrangement between chromosomes 9 and 7, involving *HAUS6* on chromosome 9. *HAUS6* is part of the augmin complex and is vital to mitotic spindle assembly and cytokinesis in human cells [19,20]. A region of CNV involving *HAUS6* was also seen in the three patients.

Our in depth analysis of 9p genetic alterations with targeted next generation sequencing highlights the 9p chromosomal fragility in ALL patients. Due to the unavailability of normal DNA from these patients, it is difficult to confirm constitutional vs. acquired nature of the genetic abnormalities. However, the fact that the novel variations are neither seen in any latest builds of variation databases nor reported in 1000 genomes data makes them more likely to be of acquired nature.

4. Conclusion

The comparison of CNV analysis of NGS and aCGH data showed similar variation pattern in the most frequently deleted *CDKN2A* region in our cohort of ALL patients. However, NGS had a higher sensitivity in CNV detection of small local variations while aCGH detected larger CNVs more accurately. We conclude that in addition to detecting mutations and other genetic aberrations, targeted NGS provides high resolution tools to detect local copy number variations. Our findings pinpoint novel genomic alterations in 9p and illustrate the high range of instability of this chromosome area in ALL.

5. Materials and methods

5.1. Patients

Altogether, 35 patients with ALL, who were diagnosed and treated at the Helsinki University Central Hospital, were included in the targeted resequencing analysis of 9p region. Of the 35 ALL patients, 22 were adolescent and young adults (age 10–25 yrs; median 16 yrs) with 14 males and 8 females, while the rest 13 were adult patients (age over 25 yrs; median 49 yrs) including 9 males and 4 females. The primary selection criteria of the patients, was, thus, the availability of aCGH results and included those from whom DNA was left for NGS analyses, comprising 25 patients with CNV and 10 patients without any CNV for the 9p region. As regards ALL subtypes, our cohort was, thus, not random but contained an excess of 9p deletion cases. The results of aCGH have been previously published elsewhere [3,7] and aCGH data is available at www.CanGEM.org (details in supplemental Table 2). This study was approved by the appropriate Institutional Review Boards and the National Authority for Medico-legal Affairs.

5.2. aCGH

DNA was extracted from the bone marrow samples and used for aCGH and NGS. ALL samples and reference DNA, pooled from peripheral blood of 4 healthy individuals, were differently labeled and hybridized to the Agilent 44 K and 244 K CGH microarrays (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocols. The arrays were scanned using the Agilent's microarray scanner G2565AA (Agilent Technologies), while Agilent's Feature Extraction software was used for image analysis and data was visualized using the Genomic Workbench Standard Edition (version 5.0.14). The most recurrently altered chromosomal region of 32 Mb was defined in 9p and selected for targeted sequencing.

5.3. Next generation sequencing

5.3.1. Target capture

Target regions with a total length of 1 Mb from a 32 Mb region of the chromosome 9p were sequenced. For capture of target regions, baits were custom designed to capture all exons, 3' and 5' untranslated region (UTR), and 1 kb upstream region of all genes within 9p24.2–9p13 (chr9:4848,246–37,036,509 in build GRch37 of the human reference genome). Custom designed baits were obtained from NimbleGen (Roche NimbleGen, Inc., Madison, WI, USA). A total of 2358 target regions, comprising of 191 protein coding and miRNA genes were included in the capture (supplementary Table 1 shows the gene list).

5.3.2. Targeted sequencing

DNA, isolated from bone marrow cells was fragmented, ligated to adapters and enriched for target regions using Roche NimbleGen's in-solution target capture and enrichment protocol. Paired-end sequencing of the target-enriched libraries was performed on the Illumina Genome Analyser Ix and HiSeq2000 sequencers (illumina, Inc, San Diego, CA, USA).

Please cite this article as: V.K. Sarhani, et al., Targeted resequencing of 9p in acute lymphoblastic leukemia yields concordant results with array CGH and reveals novel genomic alterations, Genomics (2013), http://dx.doi.org/10.1016/j.jgeno.2013.01.001
5.3.3. Primary analysis

Data obtained from sequencing was preprocessed with a variant-calling pipeline (VCP) developed at the Finnish Institute of Molecular Medicine (FIMM) [21]. Briefly, VCP included filtration of sequence reads for quality, alignment of paired-end reads to the reference genome with the Burrows-Wheeler Aligner [22], duplicate fragment removal by mdup algorithm, variant calling and identification of anomalously mapped paired-reads. For variant calling, VCP utilized SAMtools’ pileup [23]. SNV calling and read end anomaly calling, based on FIMM’s own developed algorithm, detection of indels with the Pindel [24], visualization of anomalously mapping paired-end (PE) reads with the Circos [25], were all included in the VCP. Results were visualized in the Integrative Genomic Viewer (IGV) [26]. On an average, >93% of the target region had more than 20-fold coverage for each patient.

5.4. Analysis of copy number variation

The reads mapping to target regions from the NGS were converted into discrete copy number calls by using standard copy number preprocessing techniques, including normalization, segmentation and calling [27]. To control target sequence bias between different regions, the comparison to reference sample was made in the raw data at the level of individual bins. Analogous probe-level analysis has been shown to improve comparability between multiple microarray platforms in the context of gene expression microarray studies [28,29]. The selected reference sample for CNV analyses with NGS (sample 4) did not show any copy number alterations in chromosome 9 by aCGH and had no mutation or structural rearrangement as studied by NGS.

After the bin-level preprocessing, the logarithmized raw signal was segmented using the GChallBiodiepacker package [30]. Since measurements from other chromosomes were not available in our targeted NGS study, whole-arm deletions or gains could not be determined solely on the basis of NGS observations. The whole-arm alteration status (loss/neutral/gain) was, therefore, first determined for all samples based on the aCGH data, and the median signal of the NGS segments was set accordingly to 1/0L, corresponding to whole-arm deletion, no change, and gain, respectively. Correction for the whole-arm alterations further improved the comparability between NGS and aCGH regarding the samples with whole-arm alterations and facilitated the separation between homo- and heterozygous alterations in the NGS data. Besides determining whole-arm alterations, the determination of local chromosomal changes along the 9p arm was otherwise independent between the NGS and aCGH data sets. The whole-arm alteration status for the NGS analysis could alternatively be determined through the results of cytogenetic analysis, which would make the NGS analysis completely independent of aCGH measurements. Finally, the continuous copy number signals from the segmentation step were discretized into five discrete levels of copy number call, denoted by 2/1/0/1/2 regarded as homo- and heterozygous loss, unaltered region, and low and high levels of chromosomal gain, respectively. The detection thresholds were set such that homo- and heterozygous deletions correspond to samples with <50% (<1/2) or <75% (<1.5/2) signal levels and low and high copy number gains correspond to >25% (>2.5/2) and >75% (>3.5/2) increase in the signal level with respect to the reference, respectively. The distinction between homo- and heterozygous deletions depends on the selection of the cutoff threshold. We also verified that the concordance between the NGS and aCGH copy number results was not particularly sensitive to the selection of the cutoff (data not shown).

Acknowledgments

This study was supported by research funding from the Sigrid Jusélius Foundation and the special governmental subsidy research funds appropriated to the Helsinki and Uusimaa Hospital District (HUS EVO), LL was partially supported by Alfred Kordelin Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.01.001.

References


