Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

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 Tyybäkinoja [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

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 genetic 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*, *IFN genes*, *MLLT3*, 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, (chr.9:21,970,856-22,005,115) 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,
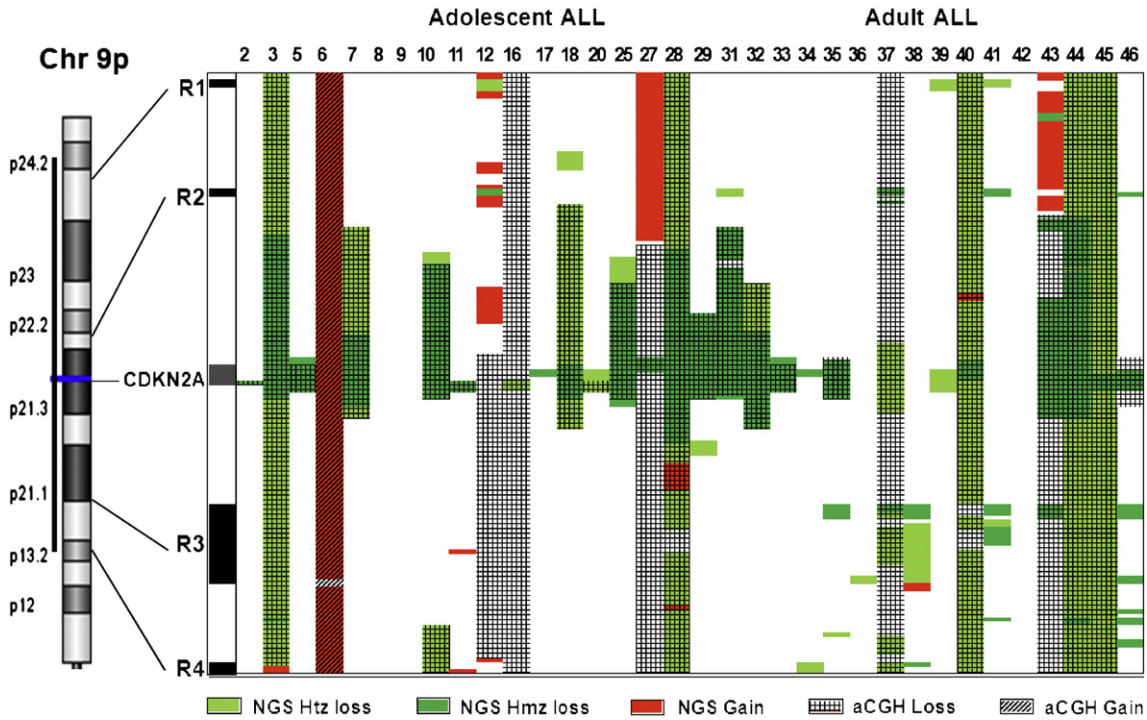
Fig. 1. Copy number variations in 9p detected by targeted NGS sequencing in adolescent (ALL 2-33) and adult (ALL 34-46) ALL patients. Regions with frequent CNVs detected by NGS are indicated by black bars marked R1 (chr 9:5111498–5113815), R2 (chr 9:19102475–19121201), R3 (chr 9:33288771–33524540), R4 (chr 9:36881901–37036509). Black vertical line along the ideogram shows the region studied by NGS. CNV shown are only from the target regions included in NGS. DNA sample from ALL-4 was used as reference for CNV by NGS. The colored region corresponds to CNVs detected with NGS and that marked with pattern correspond to that detected with aCGH. Region with color and no pattern correspond to CNV detected only by NGS and the region marked only with pattern and no color correspond to CNV seen only with aCGH.

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
Comparison of 9p region for CNV and loss of CDKN2A/B as seen with aCGH and NGS.

No of cases	Region of 9p loss		Loss of CDKN2A/B	
	aCGH	NGS	aCGH	NGS
Similar in both aCGH and NGS				
ALL-2, 5, 7, 10, 11, 18, 25, 29, 31, 32, 33	Region of loss at 9p21	Region of loss at 9p21	hmz	hmz
ALL-20	Region of loss at 9p21	Region of loss at 9p21	htz	htz
ALL-3, 28, 40, 44, 45	9p loss	target region loss	hmz	hmz
ALL-6	9p gain	target region gain		
ALL-4,* 8, 9, 36, 38, 41, 42	No imbalance at 9p	No imbalance at 9p		
ALL-35, 46	Region of loss at 9p21	Region of loss at 9p21	htz	hmz
Difference in size of deletion				
ALL-16,37	9p loss	Region of loss at 9p21	hmz	htz
ALL-27, 43	Loss at 9p21-9p13	Region of loss at 9p21	hmz	hmz
ALL-12	Loss at 9p21-9p13	No loss at 9p21-9p13	htz	
Loss detected only with NGS				
ALL-17,34	No imbalance at 9p	Region of loss at 9p21		hmz
ALL-39	No imbalance at 9p	Region of loss at 9p21		htz

hmz, homozygous; htz, heterozygous.
*ALL-4 was used as reference for CNV analysis by NGS.

(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.9:36,881901-37,036,509 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 or 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, *ADAMTSL1, PIGO* or *RUSC2* were affected in at least three patients by non-synonymous SNVs.

Table 2
Novel non-synonymous variations along with copy number variation observed at these loci in ALL patients.

Patient	Position on chromosome 9	CNV	Gene	Amino acid change
ALL-45	g.5231711 G > A	Loss	INSL4	Arg(63)His
ALL-38	g.5549463A > C	N	PDCD1LG2	Thr(164)Pro
ALL-25	g.12775965 G > T	N	LURAP1L	Gly(84)Val
ALL-11	g.14792848 G > A	N	FREM1	Arg(1292)Cys
ALL-32	g.15307246C > G	N	TTC39B	Gly(26)Arg
ALL-7	g.18675880 T > C	N	ADAMTSL1	Tyr(371)His
ALL-40	g.18775808 G > A	Loss	ADAMTSL1	Gly(822)Asp
ALL-5	g.18775844C > T	N	ADAMTSL1	Pro(834)Leu
ALL-43	g.21239667 G > T	Loss	IFNA14	Leu(90)Ile
ALL-27	g.26984513 G > A	N	IFT74	Val(141)Ile
ALL-8	g.33029927 T > C	N	DNAJA1	Tyr(119)His
ALL-36	g.33113553 T > A	N	B4GALT1	Met(366)Leu
ALL-46	g.33469243C > T	N	NOL6	Arg(275)His
ALL-3	g.33798543A > G	Loss	PRSS3	Lys(229)Glu
ALL-3	g.33798574 G > A*	Loss	PRSS3	Ser(239)Thr
ALL-40	g.33798574 G > A*	Loss	PRSS3	Ser(239)Thr
ALL-43	g.33986826 T > C	N	UBAP2	Glu(151)Gly
ALL-2	g.34372673 G > A	N	KIAA1161	Ala(90)Val
ALL-37	g.34521597C > G	N	ENHO	Ala(33)Pro
ALL-39	g.34637393 G > T	N	SIGMAR1	Ser(59)Tyr
ALL-32	g.35091321 T > C	N	PIGO	Ile(855)Val
ALL-3	g.35093527 T > C*	Loss	PIGO	Asp(277)Gly
ALL-44	g.35093527 T > C*	Loss	PIGO	Asp(277)Gly
ALL-34	g.35107658A > C*	N	FAM214B	Val(205)Gly
ALL-43	g.35107658A > C*	N	FAM214B	Val(205)Gly
ALL-45	g.35546739C > A	Loss	RUSC2	Pro(74)Gln
ALL-35	g.35547077A > C*	N	RUSC2	Thr(187)Pro
ALL-43	g.35547077A > C*	N	RUSC2	Thr(187)Pro
ALL-17	g.35555632C > A	N	RUSC2	Leu(864)Ile
ALL-34	g.35609323 G > T	N	TESK1	Ala(489)Ser
ALL-42	g.35819211 T > C	N	FAM221B	His(345)Arg
ALL-42	g.35819909C > A	N	FAM221B	Leu(277)Phe
ALL-41	g.36121624 G > A	N	RECK	Ser(878)Asn
ALL-28	g.37002702C > G	Loss	PAX5	Gly(183)Arg

CNV, copy number variation; N, normal copy number.
*variant present in two patients.

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

Table 3
Novel insertion/deletion (InDel) variants within coding region of genes on chromosome 9.

	Indel	Gene
1.	g.5922623_5922625delGCA	KIAA2026
2.	g.6610286_6610287insA	GLDC
3.	g.33264393_33264410delGGGTCAACTCCTCGCTCC	BAG1
4.	g.33797928_33797929insCC	PRSS3
5.	g.34977582_34977584delCAG	KIAA1045
6.	g.35906348_35906350delCTG	HRCT1
7.	g.35906560_35906562delCCC	HRCT1
8.	g.36840628_36840635delGGGAGCCT	PAX5
9.	g.36882005_36882006insACCATGGGCAA	PAX5

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.

3. Discussion

We have shown a high concordance of copy number variations (CNVs) between next generation sequencing (NGS) and aCGH in ALL. We used surplus frozen bone marrow aspirates not needed for diagnostic tests and the limited bone marrow sample available for our study restricted us from undertaking further validation studies. We were neither able to obtain normal bone marrow for reference sample for CNV analyses by NGS. Thus we used patient (ALL 4) bone marrow sample as a reference. As this bone marrow did not show any copy number alterations in chromosome 9 by aCGH and had no mutation or structural rearrangement as studied by NGS, we do not think that the selection of the reference had influence on the CNV results.

Table 4
Paired-end anomalies seen within 9p region showing small deletions, inversions and translocations.

|	Position Mb	Size Mb	Breakpoint	Type of rearrangement	Genes	No. of patients
1.	9p:33,52 9q:66,49		ANKRD18B (upstrream) PTGER4P2-CDK2AP2P2 (upstream)	Inversion pericentric		8
2.	9p:33,52 9q: 99,98		ANKRD18B (upstream) C9orf174	Inversion pericentric		13
3.	9p:21,47 9p:22,53	1.06	MIR31HG	Inversion paracentric	MIR31HG, MTAP, C9orf53, CDKN2A, CDKN2B-AS1, DMRTA1	1
4.	9p:33,57 9p:38,57	5.0	ANKRD18B/ ANKRD18A	Inversion paracentric	many, PAX5	2
5.	9p:37,03 9p:37,13	0.10	PAX5-ZCCHC7	Inversion paracentric	PAX5, NR_036592, ZCCHC7	1
6.	9p: 21,98 9p: 22,00	0.03	CDKN2A/ CDKN2B-AS1	Deletion	CDKN2A/2B	2
7.	9p: 33,52 9p:45,35	11.8	many, PAX5	Inversion paracentric	many,PAX5	2
8.	9p:33,03 9p: 40,07	7.04	DNAJA1	Deletion	many	1
9.	9p:19 7p		HAUS6-3′UTR	Inter-chromosomal		3
10.	9p:35,07 16q		FANCG	Inter-chromosomal		1
11.	9p:33,29 19p		NFX (upstream) NACC1-3′UTR	Inter-chromosomal		1
12.	9p:27, 0-Xq		IFT74-EDA	Inter-chromosomal		1
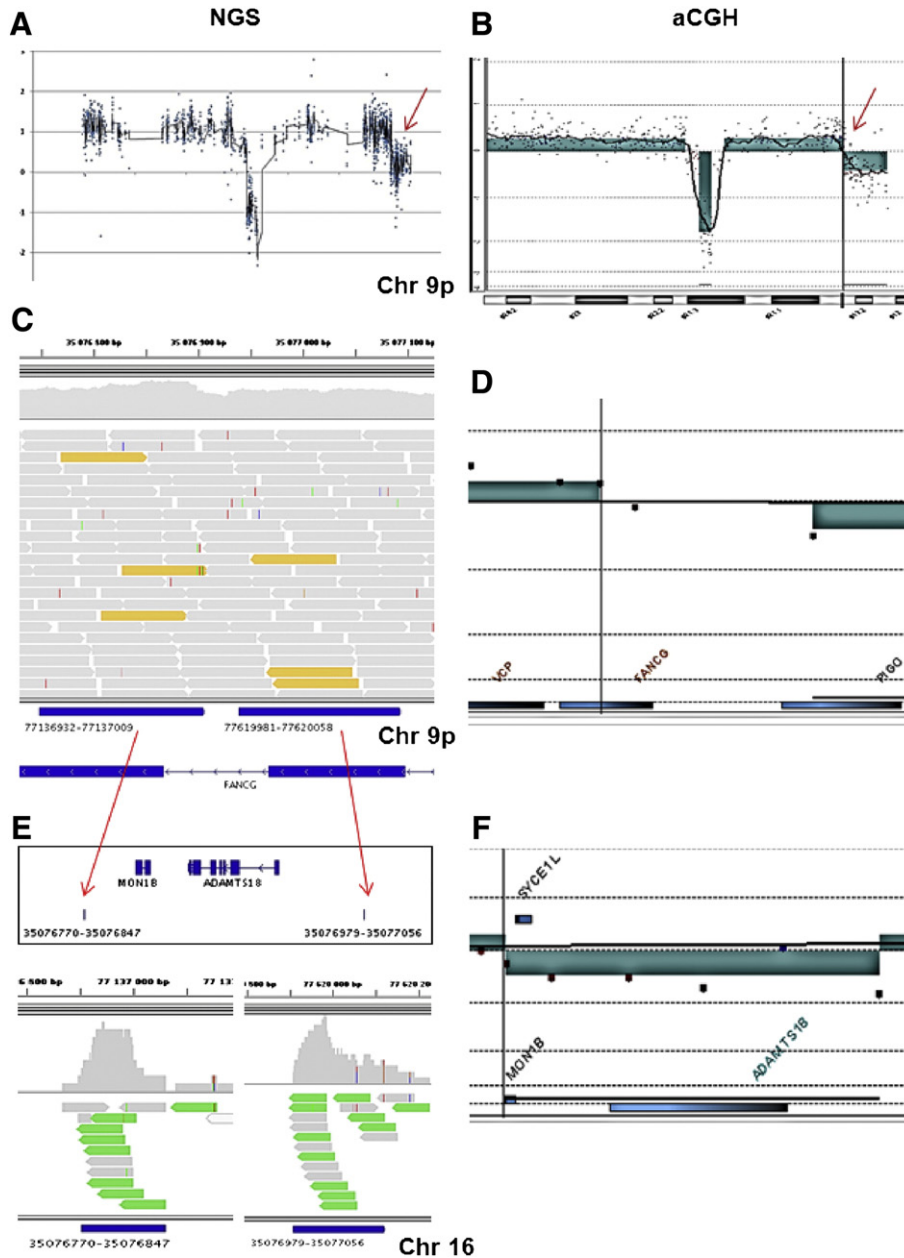
Fig. 2. Detection of CNV and structural rearrangement involving chr.9p13.3 and chr16q23.1 in ALL-10 patient by NGS and aCGH. A. Copy number profile of short arm of chromosome 9 as seen with NGS (represented as log2 ratio of ALL10/reference calculated from reads mapping to target regions and visualized with Agilent's Genomic bench). B. aCGH profile showing chr 9p in the same patient, with a loss at *CDKN2A/B* locus and another loss with a breakpoint (indicated by arrows) around chr.9:35076770. C. NGS data viewed in IGV shows anonymously mapped pairs (yellow colored), detected by NGS with breakpoint in *FANCG*. D. aCGH plot showing similar breakpoint in *FANCG*. E. NGS data showing position of the anonymously mapped pair at chr 16: 77136932–77137009 and Chr 16:77619921–77619998. F. aCGH results for the same coordinates on chr 16q23.1 showing two breakpoints at exactly the same location as in NGS and depicting a loss of *MON1B* and *ADAMTS18* genes within the region between the breakpoints, as seen with NGS.

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 thus lead to 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 hemato-poiesis. 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 *NFX1*, *AQP7*, *AQP3*, *NOL6* and *SUGT1P1* genes, with *NFX1* 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 *NFX1* 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 child-hood ALL patients with 9p abnormality [6]. In our study the fre-quency of CNV as whole was higher in adult than that in adolescent ALL. However because of lack of normal DNA available from these pa-tients, it is difficult to know the acquired or germline nature of these variations.

Since 9p is frequently deleted in ALL, SNVs in hemizygous condi-tion could have deleterious effects. Notably, our NGS analyses re-vealed 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 muta-tions, in addition to novel indels in 5′-UTR region in 9 patients. Fur-thermore, novel mutations were seen in *RUSC2* (4 patients), *PIGO* (3 patients) and in *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 struc-tural 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 *FANCG* (Fig. 2). It is worth mentioning that a higher incidence of leukemia is well known in patients with germ line mutations in *FANCG* 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. Further-more, our analysis, revealed in 3 patients a structural rearrangement be-tween 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 nei-ther 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 sim-ilar 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 muta-tions and other genetic aberrations, targeted NGS provides high resolu-tion tools to detect local copy number variations. Our findings pinpoint novel genomic alterations in 9p and illustrate the high range of instabil-ity 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 re-sults 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 Re-view 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 periph-eral blood of 4 healthy individuals, were differently labeled and hy-bridized 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 Ex-traction 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 (chr.9:4,848,246-37,036,509 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 re-gions, comprising of 191 protein coding and miRNA genes were included in the capture (supplemental 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 se-quencing of the target-enriched libraries was performed on the Illumina Genome Analyser IIx and HISeq2000 sequencers (Illumina, Inc., San Diego, CA, USA).

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 rmdup 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 CGHcall Bioconductor 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 $-1/0/1$, 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

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

[1] J.A. Irving, L. Bloodworth, N.P. Bown, M.C. Case, L.A. Hogarth, A.G. Hall, Loss of heterozygosity in childhood acute lymphoblastic leukemia detected by genome-wide microarray single nucleotide polymorphism analysis, Cancer Res. 65 (2005) 3053–3058.

[2] C.G. Mullighan, S. Goorha, I. Radtke, C.B. Miller, E. Coustan-Smith, J.D. Dalton, K. Girtman, S. Mathew, J. Ma, S.B. Pounds, X. Su, C.H. Pui, M.V. Relling, W.E. Evans, S.A. Shurtleff, J.R. Downing, Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia, Nature 446 (2007) 758–764.

[3] A. Usvasalo, S. Ninomiya, R. Raty, J. Hollmen, U.M. Saarinen-Pihkala, E. Elonen, S. Knuutila, Focal 9p instability in hematologic neoplasias revealed by comparative genomic hybridization and single-nucleotide polymorphism microarray analyses, Genes Chromosomes Cancer. 49 (2010) 309–318.

[4] A. Usvasalo, S. Savola, R. Raty, K. Vettenranta, A. Harila-Saari, P. Koistinen, E.R. Savolainen, E. Elonen, U.M. Saarinen-Pihkala, S. Knuutila, CDKN2A deletions in acute lymphoblastic leukemia of adolescents and young adults: an array CGH study, Leuk. Res. 32 (2008) 1228–1235.

[5] F. Novara, S. Beri, M.E. Bernardo, R. Bellazzi, A. Malovini, R. Ciccone, A.M. Cometa, F. Locatelli, R. Giorda, O. Zuffardi, Different molecular mechanisms causing 9p21 deletions in acute lymphoblastic leukemia of childhood, Hum. Genet. 126 (2009) 511–520.

[6] E. Coyaud, S. Struski, N. Prade, J. Familiades, R. Eichner, C. Quelen, M. Bousquet, F. Mugneret, P. Talmant, M.P. Pages, C. Lefebvre, D. Penther, E. Lippert, N. Nadal, S. Taviaux, B. Poppe, I. Luquet, L. Baranger, V. Eclache, I. Radford, C. Barin, M.J. Mozziconacci, M. Lafage-Pochitaloff, H. Antoine-Poirel, C. Charrin, C. Perot, C. Terre, P. Brousset, N. Dastugue, C. Broccardo, Wide diversity of PAX5 alterations in B-ALL: a Groupe Francophone de Cytogenetique Hematologique study, Blood 115 (2010) 3089–3097.

[7] I. Iacobucci, A. Lonetti, F. Paoloni, C. Papayannidis, A. Ferrari, C.T. Storlazzi, M. Vignetti, D. Cilloni, F. Messa, V. Guadagnuolo, S. Paolini, L. Elia, M. Messina, A. Vitale, G. Meloni, S. Soverini, F. Pane, M. Baccarani, R. Foa, G. Martinelli, The PAX5 gene is frequently rearranged in BCR-ABL1-positive acute lymphoblastic leukemia but is not associated with outcome. A report on behalf of the GIMEMA Acute Leukemia Working Party, Haematologica 95 (2010) 1683–1690.

[8] C.G. Mullighan, Genomic profiling of B-progenitor acute lymphoblastic leukemia, Best Pract. Res. Clin. Haematol. 24 (2011) 489–503.

[9] K. Paulsson, J.B. Cazier, F. Macdougall, J. Stevens, I. Stasevich, N. Vrcelj, T. Chaplin, D.M. Lillington, T.A. Lister, B.D. Young, Microdeletions are a general feature of adult and adolescent acute lymphoblastic leukemia: unexpected similarities with pediatric disease, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 6708–6713.

[10] S. Sulong, A.V. Moorman, J.A. Irving, J.C. Strefford, Z.J. Konn, M.C. Case, L. Minto, K.E. Barber, H. Parker, S.L. Wright, A.R. Stewart, S. Bailey, N.P. Bown, A.G. Hall, C.J. Harrison, A comprehensive analysis of the CDKN2A gene in childhood acute lymphoblastic leukemia reveals genomic deletion, copy number neutral loss of heterozygosity, and association with specific cytogenetic subgroups, Blood 113 (2009) 100–107.

[11] Y. Hong, Y. Maeda, R. Watanabe, N. Inoue, K. Ohishi, T. Kinoshita, Requirement of PIG-F and PIG-O for transferring phosphoethanolamine to the third mannose in glycosylphosphatidylinositol, J. Biol. Chem. 275 (2000) 20911–20919.

[12] G. Jiang, F. Cao, G. Ren, D. Gao, V. Bhakta, Y. Zhang, H. Cao, Z. Dong, W. Zang, S. Zhang, H.H. Wong, C. Hiley, T. Crnogorac-Jurcevic, N.R. Lemoine, Y. Wang, PRSS3 promotes tumour growth and metastasis of human pancreatic cancer, Gut 59 (2010) 1535–1544.

[13] A. Hockla, D.C. Radisky, E.S. Radisky, Mesotrypsin promotes malignant growth of breast cancer cells through shedding of CD109, Breast Cancer Res. Treat. 124 (2010) 27–38.

[14] C.J. Marsit, M.R. Karagas, H. Danaee, M. Liu, A. Andrew, A. Schned, H.H. Nelson, K.T. Kelsey, Carcinogen exposure and gene promoter hypermethylation in bladder cancer, Carcinogenesis 27 (2006) 112–116.

[15] C.J. Marsit, C. Okpukpara, H. Danaee, K.T. Kelsey, Epigenetic silencing of the PRSS3 putative tumor suppressor gene in non-small cell lung cancer, Mol. Carcinog. 44 (2005) 146–150.

[16] J.L. Betz, A.S. Behairy, P. Rabionet, B. Tirtorahardjo, M.W. Moore, P.D. Cotter, Acquired inv(9): what is its significance? Cancer Genet. Cytogenet. 160 (2005) 76–78.

[17] A.M. Udayakumar, A.V. Pathare, D. Dennison, J.A. Raeburn, Acquired pericentric inversion of chromosome 9 in acute myeloid leukemia, J. Appl. Genet. 50 (2009) 73–76.

[18] L. Faivre, P. Guardiola, C. Lewis, I. Dokal, W. Ebell, A. Zatterale, C. Altay, J. Poole, D. Stones, M.L. Kwee, M. van Weel-Sipman, C. Havenga, N. Morgan, J. de Winter, M. Digweed, A. Savoia, J. Pronk, T. de Ravel, S. Jansen, H. Joenje, E. Gluckman, C.G. Mathew, Association of complementation group and mutation type with clinical outcome in fanconi anemia. European Fanconi Anemia Research Group, Blood 96 (2000) 4064–4070.

[19] S. Lawo, M. Bashkurov, M. Mullin, M.G. Ferreria, R. Kittler, B. Habermann, A. Tagliaferro, I. Poser, J.R. Hutchins, B. Hegemann, D. Pinchev, F. Buchholz, J.M. Peters, A.A. Hyman, A.C. Gingras, L. Pelletier, HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity, Curr. Biol. 19 (2009) 816–826.

[20] R. Uehara, R.S. Nozawa, A. Tomioka, S. Petry, R.D. Vale, C. Obuse, G. Goshima, The augmin complex plays a critical role in spindle microtubule generation for mitotic progression and cytokinesis in human cells, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 6998–7003.

[21] A.M. Sulonen, P. Ellonen, H. Almusa, M. Lepisto, S. Eldfors, S. Hannula, T. Miettinen, H. Tyynismaa, P. Salo, C. Heckman, H. Joensuu, T. Raivio, A. Suomalainen, J. Saarela, Comparison of solution-based exome capture methods for next generation sequencing, Genome Biol. 12 (2011) R94.

[22] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform, Bioinformatics 26 (2010) 589–595.

[23] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.

[24] K. Ye, M.H. Schulz, Q. Long, R. Apweiler, Z. Ning, Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads, Bioinformatics 25 (2009) 2865–2871.

[25] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: an information aesthetic for comparative genomics, Genome Res. 19 (2009) 1639–1645.

[26] J.T. Robinson, H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, Nat. Biotechnol. 29 (2011) 24–26.

[27] M.A. van de Wiel, F. Picard, W.N. van Wieringen, B. Ylstra, Preprocessing and downstream analysis of microarray DNA copy number profiles, Brief Bioinform. 12 (2011) 10–21.

[28] L.L. Elo, L. Lahti, H. Skottman, M. Kylaniemi, R. Lahesmaa, T. Aittokallio, Integrating probe-level expression changes across generations of Affymetrix arrays, Nucleic Acids Res. 33 (2005) e193.

[29] L. Lahti, L.L. Elo, T. Aittokallio, S. Kaski, Probabilistic analysis of probe reliability in differential gene expression studies with short oligonucleotide arrays, IEEE/ACM Trans. Comput. Biol. Bioinform. 8 (2011) 217–225.

[30] W.N. van Wieringen, M.A. van de Wiel, B. Ylstra, Normalized, segmented or called aCGH data? Cancer. Inform. 3 (2007) 321–327.