Iran J Public Health, Vol. 45, No.3, Mar 2016, pp.346-352

Original Article

Whole Exome Sequencing of Chronic Myeloid Leukemia Patients

Shaghayegh SABRI [1], ***Manouchehr KEYHANI*** [2], ****Mohammad Taghi AKBARI*** [1]

1. *Dept. of Medical Genetics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran*
2. *Hematology and Oncology Research Center, Vali Asr Hospital, Tehran University of Medical Sciences, Tehran, Iran*

***Corresponding Author:** Email: mtakbari@modares.ac.ir

(Received 22 Apr 2015; accepted 10 Sep 2015)

Abstract

Background: Previous studies have shown that leukemogenic chromosomal translocations, including fusions between Break point Cluster Region (BCR) and Abelson (ABL) are present in the peripheral blood of healthy individuals. The aim of this study was to gain insights into the genetic alterations other than BCR-Abl translocation in molecular level, which cause chronic myeloid leukemia (CML).

Methods: We performed whole-exome sequencing on four cases representative of BCR-ABL positive CML in chronic phase of the disease.

Results: We did not identify any pathogenic mutation in all known genes involved in CML or other cancers in our subjects. Nevertheless, we identified polymorphisms in related genes.

Conclusion: It is the first report of exome sequencing in Philadelphia chromosome positive CML patients. We did not identify any pathogenic mutation in known cancer genes in our patients who can be due to CML pathogenesis or technical limitations.

Keywords: CML, Whole exome sequencing, Iran

Introduction

"Human chronic myeloid leukemia (CML) is a myeloproliferative disorder (MPD) caused by the Philadelphia chromosome translocation, a t (9; 22) that generates the BCR/ABL fusion oncoprotein" (1).

The BCR-ABL fusion protein is a constitutively active tyrosine kinase. Normally, this kinase precisely regulates downstream genes, including c-Myc, Akt and Jun, all of which are major players to the proliferation and survival of normal cells. However, the hyperactivity of the BCR-ABL kinase upsets this fine balance and propels cells towards uncontrolled proliferation and survival, both of which provide a growth advantage to the malignant cells bearing this mutation, ultimately leading to CML (2).

Next generation sequencing has proven to be an effective tool to identify recurrent, specific mutations in solid tumors and leukemias. Although the genetic heterogeneity of cancer necessitates some warn in the interpretation and application of the NGS results (3, 4), high-throughput sequencing remains a powerful instrument to refine potentially cancer diagnosis and treatment (5).

The aim of this study was to gain insights into the genetic alterations other than BCR-Abl translocation in molecular level, which finally cause CML. We performed whole-exome sequencing of four

cases representative of BCR-ABL positive CML in chronic phase of the disease.

Material and Methods

This study was conducted in Tarbiat Modares University Tehran, Iran in 2014. We used exome sequencing technology to identify mutations in molecular level in four individuals with CML who had given informed consent for sample collection and analysis. CML diagnosis was suspected by the Complete Blood Count (CBC) testing and then confirmed by identifying BCR-Abl translocation by real-time PCR. The selected patients were in chronic phase of CML without any other interfering disease and they received no treatment before sampling. DNA was extracted from peripheral blood using the conventional salting-out method. The qualifying DNA samples were exome sequenced by BGI (Beijing Genomics Institute).

Exome sequencing procedures and data analysis

First, genomic DNA was randomly cleaved into a fragment library, purified and subsequently enriched by NimbleGen 2.1M-probe sequence capture array. The enriched library targeting the exome was sequenced on the Illumina HiSeq 2000 platform to acquire paired-end reads with a read length of 90 base pairs. After removing reads containing sequencing adapters and low-quality reads with more than five unknown bases, high-quality reads were aligned with the human genome reference sequence (hg19/GRCh37) using Bowtie2 software 27 with default parameters. The PCR duplicates detected from Alignment files were subsequently removed with Picard (http://picard.sourceforge.net/) to improve alignment accuracy. The Genome Analysis Toolkit (GATK) was then employed for base quality recalibration, local realignment around the potential insertion/deletion (Indel) sites and variant calling. The raw single nucleotide variants were filtered for low mapping quality, low coverage, SNP clusters, etc. Then, the filtered variants were annotated using ANNOVAR for the following parameters: function (exonic or splicing); gene; exonic function (synonymous, nonsynonymous, stop gain, nonframeshift or frameshift indels); amino acid change; conservation; dbSNP (version 135) reference number; allele frequency in 1000 Genomes Project (2012 Feb version).

Results

Data characteristics of exome sequencing of four samples are shown in Table 1. Statistics of annotated variants in four samples before and after filtering are listed in Table 2. We also prepared a list of all genes already involved in CML reported in publications summarized in Table 3.

Table 1: Whole exome sequencing characteristics

Items/samples	23022	23031	23652	23878
Total effective reads	51161535	51082069	50259766	50395087
Total effective yield (Mb)	4532.41	4522.97	4444.54	4470.17
Average read length (bp)	88.59	88.54	88.43	88.7
Average sequencing depth on target	56.7	56.12	56.14	56.66
Coverage of target region	99.70%	99.60%	99.70%	99.70%
Coverage of flanking region	94.20%	93.70%	93.90%	95.50%
Fraction of target covered with at least 20x	86.10%	85.10%	87.10%	89.10%
Fraction of target covered with at least 10x	95.70%	94.90%	96.50%	96.80%
Fraction of target covered with at least 4x	98.90%	98.60%	99.10%	99.00%
Fraction of flanking region covered with at least 20x	18.70%	18.70%	18.50%	20.40%
Fraction of flanking region covered with at least 10x	41.00%	40.80%	40.90%	44.30%
Fraction of flanking region covered with at least 4x	71.00%	70.50%	70.50%	74.50%
Mapping rate	99.49%	99.36%	99.37%	99.54%
Duplicate rate	5.51%	5.58%	5.38%	5.97%

Table 2: Whole exome sequencing data statistics

Items/samples	23022	23031	23652	23878
Total variants	84385	83618	85059	87055
SNPs variants	76627	76016	77235	78637
INDEL variants	7758	7602	7824	8418
Novel SNPs variants	2751	2750	2708	2778
Novel INDEL variants	1951	1875	1938	2155
Novel functional SNPs variants	441	434	432	389
Novel functional INDEL variants	63	56	59	58

Table 3: CML candidate genes (known to be involved in CML)

Gene	Description	Link
JAK 2	Janus kinase 2	http://www.ncbi.nlm.nih.gov/pubmed/25657500
STAP2	signal transducing adaptor family member 2	http://www.ncbi.nlm.nih.gov/pubmed/22231445
IKZF1	IKAROS family zinc finger 1	http://www.ncbi.nlm.nih.gov/pubmed/18408710
FANCD2	Fanconi anemia, complementation group D2	http://www.ncbi.nlm.nih.gov/pubmed/21203397
COPS5	COP9 signalosome subunit 5	http://www.ncbi.nlm.nih.gov/pubmed/21935931
SKP2	S-phase kinase-associated protein 2, E3 ubiquitin protein ligase	http://www.ncbi.nlm.nih.gov/pubmed/20717963
SHC1	SHC (Src homology 2 domain containing) transforming protein 1	http://www.ncbi.nlm.nih.gov/pubmed/10676660
GAB2	GRB2-associated binding protein 2	http://www.ncbi.nlm.nih.gov/pubmed/12124177
GRB2	growth factor receptor-bound protein 2	http://www.ncbi.nlm.nih.gov/pubmed/10887132
CRK	v-crk avian sarcoma virus CT10 oncogene homolog	http://www.ncbi.nlm.nih.gov/pubmed/8632906
DOK2	docking protein 2, 56kDa	http://www.ncbi.nlm.nih.gov/pubmed/15611294
DOK1	docking protein 1, 62kDa (downstream of tyrosine kinase 1)	http://www.ncbi.nlm.nih.gov/pubmed/15611294
NEDD9	neural precursor cell expressed, developmentally down-regulated 9	http://www.ncbi.nlm.nih.gov/pubmed/21848808
SGK223	homolog of rat pragma of Rnd2	http://www.ncbi.nlm.nih.gov/pubmed/20697350
RhoA	ras homolog family member A	http://www.ncbi.nlm.nih.gov/pubmed/22443473
LRRK1	leucine-rich repeat kinase 1	http://www.ncbi.nlm.nih.gov/pubmed/20697350
CBL	Cbl proto-oncogene, E3 ubiquitin protein ligase	http://www.ncbi.nlm.nih.gov/pubmed/9195915
TWIST-1	twist family bHLH transcription factor 1	http://www.ncbi.nlm.nih.gov/pubmed/21123820
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	http://www.ncbi.nlm.nih.gov/pubmed/23292937
INPPL1	inositol polyphosphate phosphatase-like 1	http://www.ncbi.nlm.nih.gov/pubmed/10194451
HCK	HCK proto-oncogene, Src family tyrosine kinase	http://www.ncbi.nlm.nih.gov/pubmed/12592324
LYN	LYN proto-oncogene, Src family tyrosine kinase	http://www.ncbi.nlm.nih.gov/pubmed/12509383
HoxA9	homeobox A9	http://www.ncbi.nlm.nih.gov/pubmed/20141430
RKIP	phosphatidylethanolamine binding protein 1	http://www.ncbi.nlm.nih.gov/pubmed/25015191
CDC42	cell division cycle 42	http://www.ncbi.nlm.nih.gov/pubmed/19718053
NOX4	NADPH oxidase 4	http://www.ncbi.nlm.nih.gov/pubmed/25928540
PHLPP1	PH domain and leucine rich repeat protein phosphatase 1	http://www.ncbi.nlm.nih.gov/pubmed/19261608
PHLPP2	PH domain and leucine rich repeat protein phosphatase 2	http://www.ncbi.nlm.nih.gov/pubmed/19261608
STAT5	signal transducer and activator of transcription 5	http://www.ncbi.nlm.nih.gov/pubmed/25170113
PIK3R2	phosphoinositide-3-kinase, regulatory subunit 2 (beta)	http://www.ncbi.nlm.nih.gov/pubmed/18704194
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha	http://www.ncbi.nlm.nih.gov/pubmed/18644865
MTOR	mechanistic target of rapamycin (serine/threonine kinase)	http://www.ncbi.nlm.nih.gov/pubmed/21715304

Table 3: Cond…

MYB	v-myb avian myeloblastosis viral oncogene homolog	http://www.ncbi.nlm.nih.gov/pubmed/2741649
USP18	ubiquitin specific peptidase 18	http://www.ncbi.nlm.nih.gov/pubmed/17374743
BACH2	BTB and CNC homology 1, basic leucine zipper transcription factor 2	http://www.ncbi.nlm.nih.gov/pubmed/11746976
SELE	selectin E	http://www.ncbi.nlm.nih.gov/pubmed/15674360
NOV	nephroblastoma overexpressed	http://www.ncbi.nlm.nih.gov/pubmed/19623482
NUDCD1	NudC domain containing 1	http://www.ncbi.nlm.nih.gov/pubmed/11416219
FOLR3	folate receptor 3 (gamma)	http://www.ncbi.nlm.nih.gov/pubmed/8110752
MSI2	musashi RNA-binding protein 2	http://www.ncbi.nlm.nih.gov/pubmed/12649177
RARA	retinoic acid receptor, alpha	http://www.ncbi.nlm.nih.gov/pubmed/8180390
NUP98	nucleoporin 98kDa	http://www.ncbi.nlm.nih.gov/pubmed/24971156
VPREB1	pre-B lymphocyte 1	http://www.ncbi.nlm.nih.gov/pubmed/23881307
SOCS6	suppressor of cytokine signaling 6	http://www.ncbi.nlm.nih.gov/pubmed/25172101
CSF3R	colony stimulating factor 3 receptor (granulocyte)	http://www.ncbi.nlm.nih.gov/pubmed/23656643
LHX2	LIM homeobox 2	http://www.ncbi.nlm.nih.gov/pubmed/14687986
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	http://www.ncbi.nlm.nih.gov/pubmed/25961029
ABCG2	ATP-binding cassette, sub-family G (WHITE), member 2	http://www.ncbi.nlm.nih.gov/pubmed/24123600
SMO	smoothened, frizzled class receptor	http://www.ncbi.nlm.nih.gov/pubmed/18772113
NUMB	numb homolog (Drosophila)	http://www.ncbi.nlm.nih.gov/pubmed/21084860
miR-31	microRNA 31	http://www.ncbi.nlm.nih.gov/pubmed/22511990
TEC	tec protein tyrosine kinase	http://www.ncbi.nlm.nih.gov/pubmed/22739199
miR-155	microRNA 155	http://www.ncbi.nlm.nih.gov/pubmed/22511990
RGS2	regulator of G-protein signaling 2	http://www.ncbi.nlm.nih.gov/pubmed/7643615
BLK	BLK proto-oncogene, Src family tyrosine kinase	http://www.ncbi.nlm.nih.gov/pubmed/22797726
NAT8	N-acetyltransferase 8 (GCN5-related, putative)	http://www.ncbi.nlm.nih.gov/pubmed/24556617
miR-564	microRNA 564	http://www.ncbi.nlm.nih.gov/pubmed/22511990
ALOX5	arachidonate 5-lipoxygenase	http://www.ncbi.nlm.nih.gov/pubmed/19503090
CD44	CD44 molecule	http://www.ncbi.nlm.nih.gov/pubmed/16998483
AXL	AXL receptor tyrosine kinase	http://www.ncbi.nlm.nih.gov/pubmed/7521695
FOXO3	forkhead box O3	http://www.ncbi.nlm.nih.gov/pubmed/18644865
AKAP13	A kinase (PRKA) anchor protein 13	http://www.ncbi.nlm.nih.gov/pubmed/8290273
AHI1	Abelson helper integration site 1	http://www.ncbi.nlm.nih.gov/pubmed/22183070
SETBP1	SET binding protein 1	http://www.ncbi.nlm.nih.gov/pubmed/22566606
IRF8	interferon regulatory factor 8	http://www.ncbi.nlm.nih.gov/pubmed/24242069
ETV6	ets variant 6	http://www.ncbi.nlm.nih.gov/pubmed/19480935
PDGFB	platelet-derived growth factor beta polypeptide	http://www.ncbi.nlm.nih.gov/pubmed/2660925
PDGFRA	platelet-derived growth factor receptor, alpha polypeptide	http://www.ncbi.nlm.nih.gov/pubmed/19175693
GATA2	GATA binding protein 2	http://www.ncbi.nlm.nih.gov/pubmed/19304323
VEGFC	vascular endothelial growth factor C	http://www.ncbi.nlm.nih.gov/pubmed/22169285
AKAP12	A kinase (PRKA) anchor protein 12	http://www.ncbi.nlm.nih.gov/pubmed/15287943
SLC22A1	solute carrier family 22 (organic cation transporter), member 1	http://www.ncbi.nlm.nih.gov/pubmed/23272163
PRKDC	protein kinase, DNA-activated, catalytic polypeptide	http://www.ncbi.nlm.nih.gov/pubmed/11264175
WNT	wingless-type MMTV integration site family	http://www.ncbi.nlm.nih.gov/pubmed/22823957

Then we checked all filtered variants in each individual for these known genes to find cancer related genes but we did not identify any pathogenic mutation. Nevertheless, we identified polymorphisms in related genes, some of listed in Table 4.

Discussion

In this study we performed exome sequencing as a high throughput technology to identify genetic alterations other than BCR-Abl translocation or those that lead to this cytogenetic translocation at molecular level which finally cause CML. We used public databases to prepare a list of cancer genes for further analysis; however, no pathogenic mutation was identified. Moreover, we analyzed functional variants (coding region and splice site variants) bioinformatically, but no pathogenic mutation was found. Logically, there are two main reasons for such results in our survey; disease nature and the technique characteristics.

A chromosomal translocation includes a DNA double strand break and repair more specifically, mis-repair.

Table 4: Identified polymorphism in this study

Gene	NM_ID	Variant	Function
AKAP12	NM_144497	rs3842128	Inframe insertion
		rs10872670	Missense
		rs3734799	Missense
SETBP1	NM_001130110	rs3085861	Frameshift insertion
		rs663651	Missense
		rs3744825	Missense
FOLR3	NM_000804	rs71891516	Frameshift insertion
PIK3R2	NM_005027	rs1011320	Missense
CD44	NM_001001389	rs9666607	Missense
		rs1467558	Missense
AXL	NM_001699	rs7249222	Missense
AKAP13	NM_006738	rs2061821	Missense
		rs2061822	Missense
		rs2061824	Missense
SLC22A1	NM_003057	rs683369	Missense
		rs628031	Missense
PDGFRA	NM_006206	rs35597368	Missense
SON	NM_032195	rs13433428	Missense
		rs13047599	Missense

Accordingly, all genes implicated in homologous recombination and non-homolegous end joining, as the two main DSB repair pathways, are putative candidate genes mutated before BCR-Abl translocation (6). Moreover, Alu elements have been involved in the pathogenesis of some complex translocations including BCR and ABL, but these are extremely rare (7).

Leukemogenic chromosomal translocations, including fusions between BCR and ABL are present in the peripheral blood of healthy individuals (8). It was controversial because for decades it had been proposed that these translocations unavoidably led to leukemia. There are important hints in these results. First, it forcefully implies that this oncogenic translocation is not adequate to produce malignancy, but it instead produces a "pre-malignant" clone that requires additional, complementary, events to transform fully the cell. On the other hand, this result shows that detection of an oncogenic translocation is not equivalent to detection of a malignancy (9, 10). Second, this result makes a possible explanation for the observation that mice manipulated to overexpress

an oncogenic fusion protein often do not grow leukemia. In these mice, one oncogenic mutation is integrated in the mouse germline, but leukemic transformation is not triggered until additional mutation(s) occur spontaneously as the mouse ages. However, most of these putative mutations have not been characterized (6) and we did not identify any pathogenic mutation in related genes as well.

In this study; however some polymorphic variants were identified among them; SNPs rs683369 and rs628031 in SLC22A1, found in all subjects, have previously been studied in relation to imatinib response.

"SNP rs683369 and advanced disease stage are correlated with a high rate of loss of cytogenetic response or treatment failure to imatinib in CML patients" (11). We cannot determine the effect of this variant due to the chronic phase of the disease in our patients.

Moreover, Chowbay et al. revealed a sub-haplotypic region encompassing one exonic SNP (rs628031) surrounded by two intronic SNPs [IVS6-878C.A (rs3798168) and IVS7+850C.T] that is significantly associated with imatinib clearance (12). Except rs628031, two other polymorphisms of this sub-haplotypic region were not detected in our subjects.

Exome sequencing has been a fast and cost-effective tool to identify recurrent, specific mutations in solid tumors and leukemias (13-15). Nevertheless, this recent technique has some limitations too. Two main technical limitations in NGS, which impress exome-sequencing results, are homologous sequences and guanine cytosine (GC) bias (16) which lead to alignment errors. Another technical consideration with exome sequencing is that variants located in UTRs, intronic, promoter, and intergenic regulatory regions are mostly missed. Although it is often difficult to interpret novel variants in such regions, there are known pathogenic variants in many genes that lie outside the exons.

Conclusion

It is the first report of exome sequencing in Philadelphia chromosome positive CML patients. We did not identify any pathogenic mutation in known cancer genes in our patients who can be due to CML pathogenesis or technical limitations.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

We are grateful to patients for their participation in this study.

References

1. Daley GQ (2004). Chronic myeloid leukemia: proving ground for cancer stem cells. *Cell,* 29, 119(3): 314-6.
2. Jabbour E1, Fava C, Kantarjian H (2009). Advances in the biology and therapy of patients with chronic myeloid leukaemia. *Best Pract Res Clin Haematol,* 22(3): 395-407.
3. J. Zhao and S. Grant (2011). Advances in whole genome sequencing technology. *Curr Pharm Biotechnol,* 12(2):293-305.
4. M. Meyerson, S. Gabriel, and G. Getz (2010). Advances in understanding cancer genomes through second-generation sequencing. *Nature Rev Genet*, 10, 685–696.
5. Piazza R1, Valletta S, Winkelmann N, Redaelli S, Spinelli R, Pirola A, Antolini L, Mologni L, Donadoni C, Papaemmanuil E et al. (2013). Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat Genet*, 45(1):18-24.
6. Peter D. Aplan (2006). Causes of oncogenic chromosomal translocation. *Trends Genet,* 22(1): 46–55.
7. D M Ross, M O'Hely, P A Bartley, P Dang, J Score, J M Goyne, M Sobrinho-Simoes, N C P Cross, J V Melo, T P Speed, T P Hughes and A A Morley (2013). Distribution of

genomic breakpoints in chronic myeloid leukemia: analysis of 308 patients. *Leukemia,* 27(10):2105-7.

8. Janz S, Potter M, Rabkin CS (2003). Lymphoma- and leukemia-associated chromosomal translocations in healthy individuals. *Genes Chromosomes Cancer*, 36 (3):211–223.
9. Mori H, Colman SM, Xiao Z, Ford AM, Healy LE, Donaldson C, Hows JM, Navarrete C, Greaves M (2002). Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci,* 99(12):8242–8247.
10. Izraeli S, Waldman D (2004). Minimal residual disease in childhood acute lymphoblastic leukemia: current status and challenges. *Acta Haematol,* 112 (1–2):34–39.
11. Kim DHD, Sriharsha L, Xu W, Kamel-Reid S, Liu X, Siminovitch K, Messner HA, Lipton JH (2009). Clinical relevance of a pharmacogenetic approach using multiple candidate genes to predict response and resistance to imatinib therapy in chronic myeloid leukemia. *Clin Cancer Res,* 15(14):4750–4758.
12. Singh O, Chan JY, Lin K, Heng CCT, Chowbay B (2012) SLC22A1-ABCB1 Haplotype Profiles Predict Imatinib Pharmacokinetics in Asian Patients with Chronic Myeloid Leukemia. *PLoS ONE,* 7(12): e51771.
13. Shah SP, Köbel M, Senz J, Morin RD, Clarke BA, Wiegand KC, Leung G, Zayed A, Mehl E, Kalloger SE et al. (2009). Mutation of *FOXL2* in granulosa-cell tumors of the ovary. *N Engl J Med*, 360:2719–2729.
14. Tiacci E, Trifonov V, Schiavoni G, Holmes A, Kern W, Martelli MP, Pucciarini A, Bigerna B, Pacini R, Wells VA et al (2011). *BRAF* mutations in hairy-cell leukemia. *N Engl J Med,* 364:2305–2315.
15. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD et al. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med*, 361:1058–1066.
16. Coonrod EM, DurtschiJd Fau - Margraf RL, MargrafRl Fau, Voelkerding KV (2013). Developing genome and exome sequencing for candidate gene identification in inherited disorders: an integrated technical and bioinformatics approach. *Arch Pathol Lab Med*, 137(3):415-33.