

A Sequence Motif Associated with Intrinsic Mutation Hot-Spots in Human Cancers

Isar Nassiri, Esmaeel Azadian and Ali Masoudi-Nejad*

Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

Abstract

Mutability varies significantly along nucleotide sequences, and mutations occur at higher frequencies at certain positions of a genome. The high rate of mutation in some regions is independent from the function and structure of corresponding regions in protein products. This raises the possibility of DNA sequence and cis-elements as causes of mutations. We used computational methods to examine the surrounding region of 20 mutation hotspots related to different human genetic disorders, and combinatorial patterns of gene mutations in cancers. We introduced A(C/G)AA(C/G)(A/T) as an associated sequence motif in fourteen hotspot of mutations in different cancers. These observations are support the correlation between DNA motifs with hotspots of mutation in cancer and promising marker for selection of suitable regions for genes mutation screening.

Keywords: Co-occurrence; DNA motif; Intrinsic mutation; Mutation hot spot

Introduction

Mutations in specific genes are the etiology of many human diseases. These mutations are induced by external environmental abuse (like chemicals and radiation) or spontaneously. Examples of the mechanisms of induced and spontaneous mutagenesis are interaction between DNA and mutagens and deficiency in action of the replication machineries, respectively. The frequency of mutations are significantly vary along genomic sequences [1]. The mutation frequency in some nucleotide positions is particularly high [2]. These positions are called mutation hotspots [3]. Some properties of intrinsic mutation hotspots such as repetitive DNA sequences are related to mutagenesis processes [4]. Computer analysis of nucleotide sequence context of identified hotspots can provide information about the biomarkers and their usage for prediction of suitable regions for mutation screening [5].

DNA motifs are recurring and short patterns with a biological function [6]. They often show fingerprint of interactions between DNA and many regulatory and modification enzymes [7]. The association of DNA motifs with hotspots of intrinsic mutation in genome can be evidence of such interactions [6]. Normally, the nucleotide DNA motifs are short (5 to 20 bp) and is believed to reoccur in different genes or several times within a gene [8]. Many algorithms exist to detect and analyze DNA motifs [9]. Motif finding algorithms have been categorized in word-based and probabilistic sequence models. Word-based algorithms are mostly relying on exhaustive comparing oligo-nucleotide frequencies. In probabilistic algorithms the model parameters are estimated using maximum-likelihood principle or Bayesian inference [9]. Most of these algorithms have been designed to find motifs by incorporate parameters in higher organisms including human [7].

Comparative studies introduced several common short sequence features within recombination hotspots [10-14]. We hypothesized the involvement of DNA motifs in local high rate of mutation in cancer genes. In support of this hypothesis, we used computational methods to examine the surrounding region of hotspots in human genes and combinatorial mutational patterns.

Materials and Methods

In this study, we collected DNA motifs in adjacent of mutation hotspots, and subsequently searched for motif(s) that is common

on multiple motif sets. In first step, based on a careful inspection of the available literatures and Catalog of Somatic Mutations in Cancer (COSMIC) database, mutation hotspots in 20 cancer genes were selected (Table 1, Supplementary file 1) [15,16]. We considered the 660 nucleotide around a mutation hotspot by most common motif finder programs including AlignACE, ANN-Spec, BioProspector, Improbizer, MDScan, MEME, MITRA, Motif Sampler, SPACE, Weeder, and Trawler [17]. After analyzing the selected regions of human genome, huge numbers of DNA motifs were collected. To recognize the common motifs between sets of motifs, a C++ program was prepared and used. In this algorithm, the input is list of unique motifs as a text file and output is common motifs between clusters (Figure 1).

For each mutation hotspots, we selected an identically sized cold-spot region on the same gene, where there was no evidence of high frequency of mutation in human cancers [15]. For each hot-spot and cold-spot pairs, we recorded the number of A(C/G)AA(C/G)(A/T) motif, and compared the frequency of this motif in two groups by unpaired t-test [5].

Results

In first step, we considered the frequency of A(C/G)AA(C/G)(A/T) motif in hot and cold-spots of mutations, as recognition site of Phenylacetyl nitrogen mustard. We found this DNA motif in association with hotspot of mutations in different human cancers including breast cancer, blood cancer, large intestine cancer, lung cancer, kidney cancer, glioma, liver cancer, stomach cancer, etc. (Table 1). In all cases, hotspot of mutation and motif were on the same strand of DNA. We observed that frequency of A(C/G)AA(C/G)(A/T) were

***Corresponding author:** Ali Masoudi-Nejad, Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran, Tel: +98-21-6695-9256; Fax: +98-21-6640-4680; E-mail: amasoudin@ibb.ut.ac.ir

Received July 24, 2013; **Accepted** September 04, 2013; **Published** September 06, 2013

Citation: Nassiri I, Azadian E, Masoudi-Nejad A (2013) A Sequence Motif Associated with Intrinsic Mutation Hot-Spots in Human Cancers. J Proteomics Bioinform 6: 183-186. doi:10.4172/jpb.1000279

Copyright: © 2013 Nassiri I, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Row	Gene Name	Cancer Tissue & Disease Names	Gene Position on Human Genome	Hotspot Position on cDNA	Number of motif	
					Hot-spot	Cold-spot
1	<i>PIK3CA</i>	large intestine, breast, endometrium	Chromosome 3: 180,349,005-180,435,189 forward strand	c.3140A>G	3	1
2	<i>APC</i>	large intestine, Stomach	Chromosome 5: 112,101,483-112,209,834 forward strand	c.4348C>T	2	1
3	<i>CDKN2A</i>	haematopoietic and lymphoid tissue, central nervous system, lung	Chromosome 9: 21,957,751-21,984,490 reverse strand	c.238C>T	2	1
4	<i>ATM</i>	haematopoietic and lymphoid tissue	Chromosome 11: 107,598,769-107,745,036 forward strand.	c.9023G>A	3	2
5	<i>RB1</i>	Lung, eye	Chromosome 13: 47,775,884-47,954,027 forward strand.	c.958C>T	0	2
6	<i>VHL</i>	kidney	Chromosome 3: 10,158,319-10,168,744 forward strand	c.241C>T	1	0
7	<i>PTEN</i>	Endometrium, glioma, breast	Chromosome 10: 89,613,175-89,718,511 forward strand	c.388C>T	4	0
8	<i>FLT3</i>	haematopoietic and lymphoid tissue	Chromosome 13: 28576811:28675329:1	c.2503G>T	1	0
9	<i>MSH6</i>	central nervous system, large intestine, stomach	Chromosome 7: 48009621:48034692:1	c.3656C>T	2	0
10	<i>SMARCB1</i>	soft tissue, central nervous system	Chromosome 22: 24,129,150-24,176,701 forward strand.	c.601C>T	0	1
11	<i>ALK</i>	autonomic ganglia	Chromosome 2: 29,415,640-30,144,432	c.3824G>A	0	1
12	<i>CTNNB1</i>	Liver, soft tissue	Chromosome 3: 41,236,328-41,281,939 forward strand.	c.121A>G & c.134C>T	4	2
13	<i>FGFR3</i>	Achondroplasia, Skin, Urinary tract	Chromosome 4: 1,795,560-1,810,598	c.746C>G	3	0
14	<i>CdK6</i>	Skin	Chromosome 7: 92,072,175-92,301,148 reverse strand.	c.588C>T	0	3
15	<i>PDGFRA</i>	soft tissue, stomach	Chromosome 4: 55,095,264-55,164,412 forward strand.	c.2525A>T	0	0
16	<i>JAK2</i>	haematopoietic and lymphoid tissue	Chromosome 17: 7,512,445-7,531,642 reverse strand.	c.1849G>T	2	1
17	<i>KIT</i>	haematopoietic and lymphoid tissue, soft tissue	Chromosome 4: 55,524,095-55,606,879 forward strand.	c.2447A>T	5	0
18	<i>MAP2K4</i>	urinary tract, testis	Chromosome 17: 11,864,866-11,987,865 forward strand.	c.551C>T	2	0
19	<i>RET</i>	Thyroid, adrenal gland	Chromosome 10: 42,892,523-42,945,805 forward strand.	c.2753T>C	0	0
20	<i>P53</i>	Breast, Large intestine, Upper aerodigestive tract	Chromosome 17: 7572927-7579912 forward strand.	c.743G>A	2	0

Table 1: Selected genes for analysis of common sequence motifs associated with intrinsic mutation hot spots.

significantly different for hot-spot and cold-spot regions ($t(4.81)=2.62$, $P<0.05$) (Table 1).

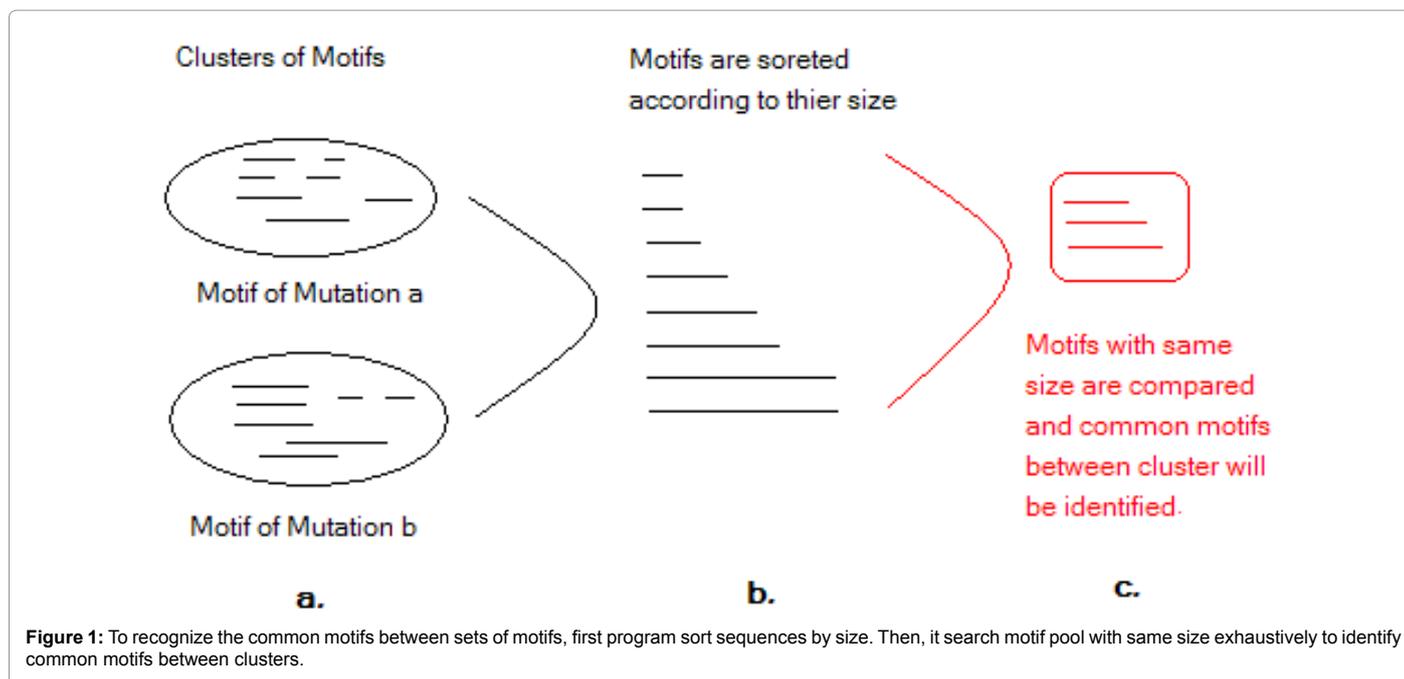
The association of specific DNA motif with hot spots of mutation can be indication of same sequence dependent mutagenesis mechanism. We examined the authenticity of this theory by study of combinatorial mutational patterns in fourteen different human cancer genes containing A(C/G)AA(C/G)(A/T) motif [16]. The co-occurrence of mutations have been detected in *PI3K* with *ACP*, *RB1* with *PTEN*, *KIT* with *CTNNB1*, *PDGFR* and *FLT3*, and *P53* with *RB1*, *PTEN*, *PIK3* and *APC* genes [18].

In next step, we analyzed selected regions of human genome around the hotspot of mutations by MotifVoter to find more motifs that are common between them. We compared hundreds of motifs in produced sets, but any statistically significant common motif did not detect.

Discussion

Mutation screening is one of the most popular methods for genetic diagnosis and confirming the genetic cause of diseases. The numerous exon of genes and limitation in length of sequence that can be analyzed with common sequencing machines are two major difficulties in screening of mutations [19]. For example, the *APC* gene is composed of 15 exons, nevertheless, about two thirds of the total mutations are clustered in the 5' region of exon 15 [20]. We hypothesized the DNA motif(s) can result in local high rate of mutation and this motif can be a marker for selection of most probable regions for mutation screening [21]. In support of this hypothesis, we used computational analysis method to examine surrounding regions of mutation hotspots in 20 human cancer genes.

According to the results of this study, we introduced A(C/G)



AA(C/G)(A/T) as an associated DNA motif with fourteen hotspots of mutations. In addition, the simultaneous of mutations in different human genes containing A(C/G)AA(C/G)(A/T) motif could be the evidence of same mutagenesis mechanisms.

The A(C/G)AA(C/G)(A/T) motif is glucocorticoid receptor (GR) binding sequence and cyclic polyamide recognition site [22]. The glucocorticoid receptor response element (GRE) consists of palindromic GR binding sequence to which a hormone-bound GR dimer can be bind. GR behave as a ligand-inducible regulator that binds to DNA as a monomer at GRE half-sites (GRE1/2). GRE1/2s can act without associated elements [23]. For example, in mammary tumor virus (MMTV) enhancer, the GRE half-site can be bound by GR [24]. The glucocorticoid recruits ATP-dependant chromatin remodeling complex positioned nucleosomes over the GRE half-sites. The absence of this chromatin remodeling prevents the binding of NF-1, Oct and TBP to the DNA. In contrast, nucleosomes do not hinder GR [25]. It is likely, some features of chromatin structure have effect on mutational frequency of neighbor sequences [26]. The A(C/G)AA(C/G)(A/T) motif is also the recognition site of Phenylacetyl nitrogen mustard (PNA) and hairpin polyamides [27,28]. PNA show specificity for its match sequence (5'-AGAACT-3') in minor groove of DNA. The replacement of sulfur by nitrogen from mustard gas produces PNA as an alkylating agent. They formerly used as toxicants and powerful mutagens, but now use to preventing development of neoplasms. PNA and hairpin polyamides exhibit exquisite AGAACT sequence specificity, and PNA conjugates exploited to direct modification of DNA targets [29,30].

Some previous researches characterized the role of nucleotide sequences and hotspot of mutations [19,21,31]. One well-known example of intrinsic mutation hotspots is CpG di-nucleotides [32]. Another example of mutational hotspots is signal of retroposable elements (LINEs and SINEs) in mammals [33]. The (CG)₄ motif was found to be hotspot of Spontaneous frame shifts in *Salmonella typhimurium* [34]. GC is hotspots of frame shifts during DNA synthesis *in vitro* by *Sulfolobus solfataricus* DNA polymerase IV [35]. The

CATCGCTTRRT motif is signal of recombination in *Bacillus subtilis mal* gene [36]. Although these studies have characterized the role of DNA motifs in replication, recombination, and mutagenesis at the sequence level, none have investigated the relation between sequence motifs and hotspot of mutations in human cancer genes [21].

Conclusion

Sequence motifs as biomarkers hold promise for the uniform, molecular, and database-driven predictor of mutation hotspots in coding region of candidate human disease genes. In addition, the connection between genes mutations and specific DNA motif as a target of small molecules is a proper opportunity from serendipity capable of reprogramming the cellular machinery to the challenge of treating genetic diseases.

Supplementary File

It includes the histogram of mutation frequency for 20 cancer genes and identically sized regions around their cold and hotspot.

References

- Lundemo S, Stenøien HK, Savolainen O (2010) Investigating the effects of topography and clonality on genetic structuring within a large Norwegian population of *Arabidopsis lyrata*. *Ann Bot* 106: 243-254.
- Benzer S (1961) On The Topography Of The Genetic Fine Structure. *Proc Natl Acad Sci U S A* 47: 403-415.
- Rogozin IB, Babenko VN, Milanese L, Pavlov YI (2003) Computational analysis of mutation spectra. *Brief Bioinform* 4: 210-227.
- Laskov R, Yahud V, Hamo R, Steinitz M (2011) Preferential targeting of somatic hypermutation to hotspot motifs and hypermutable sites and generation of mutational clusters in the IgVH alleles of a rheumatoid factor producing lymphoblastoid cell line. *Mol Immunol* 48: 733-745.
- Rogozin IB, Pavlov YI (2003) Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat Res* 544: 65-85.
- D'haeseleer P (2006) What are DNA sequence motifs? *Nat Biotechnol* 24: 423-425.
- Stormo GD (2000) DNA binding sites: representation and discovery. *Bioinformatics* 16: 16-23.

8. Rombauts S, Déhais P, Van Montagu M, Rouzé P (1999) PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res* 27: 295-296.
9. Das MK, Dai HK (2007) A survey of DNA motif finding algorithms. *BMC Bioinformatics* 8 Suppl 7: S21.
10. Smith AD, Sumazin P, Xuan Z, Zhang MQ (2006) DNA motifs in human and mouse proximal promoters predict tissue-specific expression. *Proc Natl Acad Sci U S A* 103: 6275-6280.
11. Myers S, Freeman C, Auton A, Donnelly P, McVean G (2008) A common sequence motif associated with recombination hot spots and genome instability in humans. *Nat Genet* 40: 1124-1129.
12. Mani P, Yadav VK, Das SK, Chowdhury S (2009) Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination. *PLoS One* 4: e4399.
13. McVean G (2010) What drives recombination hotspots to repeat DNA in humans? *Philos Trans R Soc Lond B Biol Sci* 365: 1213-1218.
14. Zheng J, Khil PP, Camerini-Otero RD, Przytycka TM (2010) Detecting sequence polymorphisms associated with meiotic recombination hotspots in the human genome. *Genome Biol* 11: R103.
15. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, et al. (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 39: D945-D950.
16. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, et al. (2007) A map of human cancer signaling. *Mol Syst Biol* 3: 152.
17. Wijaya E, Yiu SM, Son NT, Kanagasabai R, Sung WK (2008) MotifVoter: a novel ensemble method for fine-grained integration of generic motif finders. *Bioinformatics* 24: 2288-2295.
18. Yeang CH, McCormick F, Levine A (2008) Combinatorial patterns of somatic gene mutations in cancer. *FASEB J* 22: 2605-2622.
19. Rosenberg SM (2001) Evolving responsively: adaptive mutation. *Nat Rev Genet* 2: 504-515.
20. Crow JF (1997) The high spontaneous mutation rate: is it a health risk? *Proc Natl Acad Sci U S A* 94: 8380-8386.
21. Rogozin IB, Malyarchuk BA, Pavlov YI, Milanese L (2005) From context-dependence of mutations to molecular mechanisms of mutagenesis. *Pac Symp Biocomput* .
22. Meijnsing SH, Pufall MA, So AY, Bates DL, Chen L, et al. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* 324: 407-410.
23. Segard-Maurel I, Rajkowski K, Jibard N, Schweizer-Groyer G, Baulieu EE, et al. (1996) Glucocorticosteroid receptor dimerization investigated by analysis of receptor binding to glucocorticosteroid responsive elements using a monomer-dimer equilibrium model. *Biochemistry* 35: 1634-1642.
24. Aumais JP, Lee HS, DeGannes C, Horsford J, White JH (1996) Function of directly repeated half-sites as response elements for steroid hormone receptors. *J Biol Chem* 271: 12568-12577.
25. Archer TK, Lee HL (1997) Visualization of multicomponent transcription factor complexes on chromatin and nonnucleosomal templates in vivo. *Methods* 11: 235-245.
26. Blomquist P, Li Q, Wrange O (1996) The affinity of nuclear factor 1 for its DNA site is drastically reduced by nucleosome organization irrespective of its rotational or translational position. *J Biol Chem* 271: 153-159.
27. Nielsen PE, Egholm M, Berg RH, Buchardt O (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254: 1497-1500.
28. Dervan PB, Edelson BS (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol* 13: 284-299.
29. Kim KH, Nielsen PE, Glazer PM (2006) Site-specific gene modification by PNAs conjugated to psoralen. *Biochemistry* 45: 314-323.
30. O'Hare CC, Uthe P, Mackay H, Blackmon K, Jones J, et al. (2007) Sequence recognition in the minor groove of DNA by covalently linked formamido imidazole-pyrrole-imidazole polyamides: effect of H-pin linkage and linker length on selectivity and affinity. *Biochemistry* 46: 11661-11670.
31. Radman M (1999) Enzymes of evolutionary change. *Nature* 401: 866-867, 869.
32. Cupples CG, Cabrera M, Cruz C, Miller JH (1990) A set of lacZ mutations in *Escherichia coli* that allow rapid detection of specific frameshift mutations. *Genetics* 125: 275-280.
33. Jurka J, Klonowski P (1996) Integration of retroposable elements in mammals: selection of target sites. *J Mol Evol* 43: 685-689.
34. Isono K, Yourno J (1974) Chemical carcinogens as frameshift mutagens: *Salmonella* DNA sequence sensitive to mutagenesis by polycyclic carcinogens. *Proc Natl Acad Sci U S A* 71: 1612-1617.
35. Kokoska RJ, Bebenek K, Boudsocq F, Woodgate R, Kunkel TA (2002) Low fidelity DNA synthesis by a γ family DNA polymerase due to misalignment in the active site. *J Biol Chem* 277: 19633-19638.
36. Lopez P, Espinosa M, Greenberg B, Lacks SA (1984) Generation of deletions in pneumococcal mal genes cloned in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 81: 5189-5193.