



Comparison of hematopoietic cancer stem cells with normal stem cells leads to discovery of novel differentially expressed SSRs



Batool Hosseinpour^a, Mohammad Reza Bakhtiarizadeh^{b,*},
Seyedeh Maryam Mirabbassi^a, Esmail Ebrahimie^{c,d,**}

^a Iranian Research Organization for Science and Technology, Tehran, Iran

^b Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran

^c Institute of Biotechnology, Shiraz University, Shiraz, Iran

^d School of Molecular and Biomedical Science, The University of Adelaide, Adelaide, Australia

ARTICLE INFO

Article history:

Received 26 December 2013

Received in revised form 2 July 2014

Accepted 29 July 2014

Available online 30 July 2014

Keywords:

Cancer stem cell

GAT triplet

Gene Ontology

SSR

Tandem repeat

NF-kappa

ABSTRACT

Tandem repeat expansion in the transcriptomics level has been considered as one of the underlying causes of different cancers. Cancer stem cells are a small portion of cancer cells within the main neoplasm and can remain alive during chemotherapy and re-induce tumor growth. The EST-SSR background of cancer stem cells and possible roles of expressed SSRs in altering normal stem cells to cancer ones have not been investigated yet. Here, SSR distributions in hematopoietic normal and cancer stem cells were compared based on the expressed EST-SSR. One hundred eighty nine and 223 EST-SSRs were identified in cancer and normal stem cells, respectively. The EST-SSR expression pattern was significantly different between normal and cancer stem cells. The frequencies of AC/GT and TA/TA EST-SSRs were about 10% higher in cancer than normal stem cells. Remarkably, the number of triplets in cancer stem cells was 1.5 times higher than that in normal stem cells. GAT EST-SSR was frequent in cancer stem cells, but, conversely, normal stem cells did not express GAT EST-SSR. We suggest this EST-SSR as a novel triplet in cancer stem cell induction. Translating EST-SSRs to amino acids demonstrated that Asp and Ile were more abundant in cancer stem cells compared to normal stem cells. Finally, Gene Ontology (GO) enrichment analysis was carried out on genes containing triplet SSRs and showed that SSRs intentionally visit some specific GO classes. Interestingly, a NF-kappa (nuclear factor-kB) binding transcription factor was significantly hit by SSR instability which is a hallmark for leukemia stem cells. NF-kappa is an over represented transcription factor during cancer progression. It seems that there is a crosstalk between the NF-kB transcription factor and expressed GAT tandem repeat which negatively regulate apoptosis. In addition to better understanding of tumorigenesis, the findings of this study offer new DNA markers for diagnostic purposes and identifying at risk populations. In addition, a new approach for gene discovery in cancer by target analysis of differentially expressed EST-SSRs between cancer and normal stem cells is presented here.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Identification of the cell type capable of initiating and sustaining growth of the neoplastic clone in vivo has received an increased attention in cancer research. It has been suggested that a rare type of cells, so called cancer stem cells, is responsible for maintenance of the neoplasm. Unlike other cells in a tumor, these cells can be saved and remain alive during chemotherapy and induce tumor growth recurrently. There are some similar characteristics between normal stem cells and cancer

stem cells. Both have the capability of self-renewal and proliferation. The best proof of existence of cancer stem cells has been derived from analysis of hematological malignancies (Soltysova et al., 2005).

Hematopoietic cancer stem cells were first isolated by Bonnet and Dick (1997). Previous reports strongly provided evidences for the stem cell origin of acute and chronic myeloid leukemia (Bhatia et al., 2003; Graham et al., 2002; Guan et al., 2003; Holyoake et al., 1999; Hope et al., 2004; Jaiswal et al., 2003). Hope et al. (2004) reported that acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes. Furthermore, highly quiescent stem cells were isolated from chronic myeloid leukemia by Holyoake et al. (1999). Cancer stem cells have also been identified in solid tumors such as breast, brain and lung cancers (Al-Hajj et al., 2003; Hemmati et al., 2003; Kim et al., 2005; Singh et al., 2003).

Considerable development in genomics and functional genomics has led to generation of large-scale biological data collections deposited in

Abbreviations: ESTs, expressed sequenced tags; SSR, single sequence repeat; GO, Gene Ontology; CSCs, cancer stem cells.

* Corresponding author.

** Correspondence to: E. Ebrahimie, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia.

E-mail addresses: mrbakhtiari@ut.ac.ir (M.R. Bakhtiarizadeh), esmaeil.ebrahimie@adelaide.edu.au (E. Ebrahimie).

databanks. Meta-analysis of such data has been used for functional study of biological events such as diseases. Expressed sequenced tags (ESTs) are incomplete sequenced cDNAs which developed from cDNA libraries to study transcriptome under developmental or environmental conditions (Bakhtiarizadeh et al., 2012; Shamloo-Dashtpaderdi et al., 2013; Toth et al., 2000). ESTs of different organisms obtained from cells, organs or tissues under different biological conditions have been collected and deposited in known databanks such as NCBI (<http://www.ncbi.nlm.nih.gov/blast>) and Gene Index-DFCI (The Gene index Project, <http://compbio.dfci.harvard.edu/tgi/tgipage.html>). SSR motifs or microsatellites as a kind of DNA markers can be identified in ESTs. They are more advantageous compared to other DNA markers because of their simplicity, abundance, ubiquity, variation, co-dominance, and multi-allele nature among genomes (Bakhtiarizadeh et al., 2011, 2012; Sharma et al., 2009). It has been reported that EST-SSRs are mostly detected in transcribed active regions rather than in interons or intergenic regions (Slate et al., 2007). Additionally, there are evidences that EST-SSRs are not distributed randomly throughout the genome probably due to their effects on chromatin organization, regulation of gene activity, recombination, DNA replication, cell cycle, and mismatch repair (Bakhtiarizadeh et al., 2011; Li et al., 2004).

Dinucleotide repeats are commonly more abundant than the other types of tandem repeats. Some dinucleotides are functional. The (CT)_n element can bind to the GAGA transcription factor. This binding can modify the chromatin structure being hypersensitive to DNase I and lead to realignment of nucleosomes at heat shock gene promoters (Seema, 2004).

Trinucleotide repeats, as a subset of SSRs, draw more attention because of their role in cancers and human neurodegenerative disorders (Bacolla et al., 2008; Bakhtiarizadeh et al., 2011; Haddad et al., 1998). For example, the role of (CAG)_n repeat expansions in spinobulbar muscular atrophy has been proved in myotonic dystrophy (Brouwer et al., 2009). Microsatellite instability reflects replication errors induced by a defective function of mismatch repair genes which result in appearance of novel, non-inherited alleles in tumor cells. Consequently, co-expressed SSRs with ESTs can be correlated with clinicopathological features of cancer and its formation and tumor development. SSRs have been used in cancer genetics and indirect cancer diagnosis to help unraveling the genetic basis of tumor formation and cancer progression. Microsatellite instability has been reported in colorectal cancer (Oda et al., 2005), breast cancer (Janatova and Pohlreich, 2004; Miller and Liu, 2007), ovarian cancer, lung cancer and other cancers (Alvarez and Lokeshwar, 2007; Bakhtiarizadeh et al., 2011; Deng et al., 2007; Grady and Carethers, 2008; Tinelli et al., 2010; Venkatesan et al., 2006; Vergara et al., 2010).

However, up to now, there has been no report on the identification of EST-SSRs in hematopoietic cancer stem cells. To this aim, the distribution and type of EST-SSRs in normal and cancerous hematopoietic stem cells were compared. Then the amino acids were derived from differential ESTs and their frequencies were compared between the two stem cells. Finally, the putative functions of SSR-containing genes were analyzed in both types of stem cells based on Gene Ontology (GO) enrichment analysis.

2. Methods

2.1. EST libraries

EST libraries of normal hematopoietic stem cells and cancer stem cells were downloaded from the EST collection (The Gene Index Project) of Harvard University (<http://compbio.dfci.harvard.edu/tgi/tgipage.html>). Three EST libraries of normal hematopoietic stem cells (Cat No. 6Q0 containing 5458 ESTs; Cat No. 6PV containing 3162 ESTs; Cat No. 6L8 containing 8684 ESTs) and two EST libraries of cancer stem cells (Cat No. 5DU containing 7633 ESTs; Cat No. 5PS containing 3162 ESTs) were selected.

2.2. Comparison of EST-SSRs between normal stem cell and CSC libraries

EST-SSRs were identified using SSR Locator software. Micro- and minisatellites were retrieved from DNA sequences by this software. To analyze the distribution of EST-SSRs between two cells, EST sequences of two normal and cancer stem cells were pooled. Next, the EST sequences of normal and cancerous libraries were scanned for SSR motifs ranging in length from 2 to 7 nucleotides with dinucleotide repeat numbers ≥ 7 , trinucleotide repeat numbers ≥ 6 , tetranucleotide repeat numbers ≥ 5 , pentanucleotide repeat numbers ≥ 5 , hexanucleotide repeat numbers ≥ 5 and heptanucleotide repeat numbers ≥ 3 .

2.3. Comparison of EST-SSRs within cancer stem cell libraries

Two EST libraries of cancer stem cells (Cat No. 5DU and Cat No. 5PS) were used to test whether the EST-SSR distribution is similar between cancerous tissues. The EST sequences from the cancerous libraries were searched with SSR Locator for SSR motifs ranging from 2 to 6 nucleotides in length. The repeat number parameters were as follows: ≥ 7 for dinucleotides, ≥ 6 for trinucleotides, ≥ 5 for tetranucleotides, ≥ 5 for pentanucleotides, ≥ 5 for hexanucleotides and ≥ 3 for heptanucleotides.

2.4. Primer designing for EST-SSRs

For each microsatellite-containing EST, primers were designed using Primer3 (<http://frodo.wi.mit.edu/primer3/>) by running the software in a batch mode with the assistance of the SSR Locator interface module. The primer design function was used to determine if the sequences had sufficient flanking sequences for designing primers. The major parameters for primer design were as follows: PCR product size of 100–300 bp, primer length of 18–25 bp with 20 bp as the optimum, optimum annealing temperature of 58–63 °C with 60 °C as the optimum, and a minimum GC content of 30%, with 50% being the optimum.

2.5. Amino acid distributions of ESTs with trinucleotide repeats

The type of amino acids and their distributions in normal and cancer stem cells were predicted for ESTs with trinucleotide repeats using SSR Locator software. Translating the EST-SSRs to their corresponding amino acids provides some clues about the differences and similarities between normal and cancer stem cells at the protein level.

2.6. Statistical analysis

To understand the similarities and differences between normal and cancer stem cells in generation of EST-SSRs, Fisher's exact test was employed to compare the number of expressed SSRs in each class of EST-SSRs (dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, hexanucleotides and heptanucleotides) between the normal and cancer stem cell libraries. To this end, the IDEG6 web application (Romualdi et al., 2003) was used as described previously (Shamloo-Dashtpaderdi et al., 2013). In addition, to evaluate the co-linearity of cancerous and normal tissues in generating different EST-SSRs, Pearson correlations were calculated using MINITAB16 software (www.minitab.com).

After predicting the amino acid composition of ESTs with trinucleotide tandem repeats, the number of amino acid loci and number of amino acid repeats were compared between normal and cancer stem cells by Fisher's exact test. Furthermore, to test the equal distribution of amino acids between the cancer stem cells and normal libraries, Bartlett's test (Manly, 2004) was performed by MINITAB16 to examine the equality of variance.

In addition, the expressed SSRs in each class of EST-SSRs were compared by the Fisher's exact test between cancer stem cell libraries to examine their distribution in cancer stem cells by IDEG6 web application.

2.7. Annotation of SSR-containing sequences

To shed light on the putative functions of SSR-containing genes in normal and cancer stem cells, Fasta files of all identified EST-SSRs in cancer and normal stem cells were subjected to Blast2GO (<http://www.blast2go.org/>) software and were run against the non-redundant (nr) protein database of the NCBI (<http://www.ncbi.nlm.nih.gov/blast>). The obtained hits were compiled. EST-SSRs with the best e-value of 10^{-6} or lower were assigned a putative identity.

3. Results

3.1. Frequency and distribution of EST-SSRs in normal and cancer stem cells

In total, 28,099 ESTs were analyzed by SSR Locator in this study, including 17,304 ESTs of normal stem cells and 10,795 ESTs of cancer stem cells (Table 1). Analyzing a large number of ESTs contributes in increasing the accuracy of the results. The average length of ESTs was 463 and 523 bp in normal and cancer stem cells, respectively. This length difference clearly shows that when normal stem cells turn into cancer stem cells, a shift in alternative splicing occurs in the whole genome, thereby producing longer ESTs and proteins.

As presented in Table 1, 197 SSRs were found in 189 ESTs in cancer stem cells and 223 SSRs were identified in 214 ESTs in normal stem cells. Fig. 1 compares the frequencies of dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, hexanucleotides and heptanucleotides in different libraries of normal and cancer stem cells. Also, Table 2 represents the average percentages of SSRs in normal and cancer stem cells. As it can be inferred from Fig. 1 and Table 2, trinucleotides have higher frequencies in cancer stem cells whereas dinucleotides and tetranucleotides are more frequent in normal stem cells. The average percentage of trinucleotides in cancer stem cells is 45.1% while this ratio decreases to 21.1% in normal stem cells. Fisher's exact test confirms that the observed difference is statistically significant at $p = 0.01$ (Table 2). Higher frequencies of trinucleotides compared to normal conditions have been previously reported in neurodegenerative disorders and lung cancer (Bacolla et al., 2008; Bakhtiarizadeh et al., 2011; Haddad et al., 1998). Distributions of pentanucleotides, hexanucleotides and heptanucleotides were not significant between the two types of stem cells (Fig. 1).

If the observed differences between distributions of EST-SSRs are actually related to the differences between cancer stem cells and normal stem cells, no statistically significant difference should be detected within the libraries of each type of stem cells. Fig. 2 shows that the distributions of EST-SSRs within the libraries of cancer stem cells or libraries of normal stem cells are similar, but distribution is different between cancer and normal stem cell libraries. To increase confidence on the observed differences, the distributions of the different classes of SSRs (di-, tri-, tetra-, penta-, hexa- and hepta-nucleotides) on ESTs within normal stem cell libraries as well as within cancer stem cell libraries were examined by Bartlett's test (F-test) as well as Levene's

test (Supplementary 1). Both tests confirmed the equal distribution of EST-SSRs within cancer libraries or within normal libraries assuring us that the observed differences are caused by the different nature of cancer stem cells versus normal stem cells.

3.2. Comparative analysis of EST-SSR types between normal stem cells and cancer stem cells

The frequencies of the sequences within each class of repeat units (dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, hexanucleotides or heptanucleotides) were compared between normal and cancer stem cells (Supplementary 2). Fig. 3 represents different types of dinucleotides and their frequencies in both normal and cancer stem cells. Six dinucleotides including AC/GT, AG/CT, AT/AT, TA/TA, CA/TG, and GA/TC were identified in both types of stem cells. Interestingly, the distribution of dinucleotide sequences was not similar between the two types of stem cells. In cancer stem cells, 33% and about 13% of the dinucleotides belonged to AC/GT and TA/TA respectively. Based on the results, the frequencies of AC/GT and TA/TA in cancer stem cells were about 10% higher than those in normal stem cells. In contrast, the frequencies of CA/TG, AT/AT and AG/CT were higher in normal stem cells. Therefore, AC/GT and TA/TA can be a discrepancy between the two types of stem cells (cancer and normal).

Twenty one types of triplet repeats were detected in cancer stem cells, while 15 types of triplets were found in normal stem cells (Supplementary 2). Some triplets were found only in a single cancer stem cell library or normal stem cell library. However, the GAT triplet was frequently expressed in cancer stem cells (31% and 42%, 5DU and 5PS), while this SSR was not expressed in the normal stem cells (Supplementary 2).

The numbers of EST-SSRs in each class of repeat units were compared between cancer stem cells and normal stem cells (Table 2). Cancer stem cells statistically had a greater number of trinucleotide tandem repeats ($p = 0.01$) based on Fisher's exact test. This finding confirms the probable role of trinucleotide EST-SSRs in induction of cancer stem cells for the first time. The reported role of trinucleotide EST-SSRs in cancers such as lung and breast (Bakhtiarizadeh et al., 2011; Kozłowski et al., 2010) supports this finding. As presented in Table 2, dinucleotide EST-SSRs were more abundant in normal stem cells than in cancer stem cells.

3.3. Disagreement of cancer stem cells with normal stem cells in generation of different types of expressed SSRs

Pearson correlation between normal and cancer stem cells in generation of different types of expressed SSR sequences in each class of tandem repeats (dinucleotides, trinucleotides, and tetranucleotides) is presented in Table 3. Interestingly, there is no correlation found on the expression of trinucleotide SSRs between cancer and normal stem cell libraries (Table 3). In contrast, the correlation between normal and cancer stem cell libraries was positive (Table 3). It can be concluded that in line with the alteration of normal stem cells to cancer stem cells, the expression profile of trinucleotide EST-SSRs significantly alters. Expressed trinucleotide tandem repeats can be considered as reliable candidates for detecting and predicting cancer stem cells in future studies.

3.4. Virtual PCR

In the pooled cancerous library, 144 EST-SSRs had proper flanking regions for primer design. Consequently, 63, 81, 65, 12 and 81 primers were designed for 5DU, 5PS, 6L8, 6PV and 6Q0 libraries, respectively (Supplementary 3). When virtual PCR was run with the SSR Locator software, 110 out of 144 primers produced suitable fragments. Based on the proper flanking regions, 158 primers were identified for 158 EST-SSRs in the normal library, and 105 of these primers produced

Table 1
A summary of the EST and EST-SSR distribution in cancer and normal stem cell libraries.

EST and EST-SSR parameters	Cancer stem cells	Normal stem cells
Total number of ESTs	10,795	17,304
Average length of EST sequences	523 bp	463 bp
Total number of identified SSRs	197	223
Total number of SSR-derived ESTs (number of ESTs containing SSR)	189 (1.75%)	214 (1.23%)
Number and frequency of EST sequences containing one SSR	181 (95.77%)	206 (96.26%)
Number and frequency of EST sequences containing two SSRs	8 (4.23%)	7 (3.27%)
Number and frequency of EST sequences containing three SSRs	0 (0.00%)	1 (0.47%)

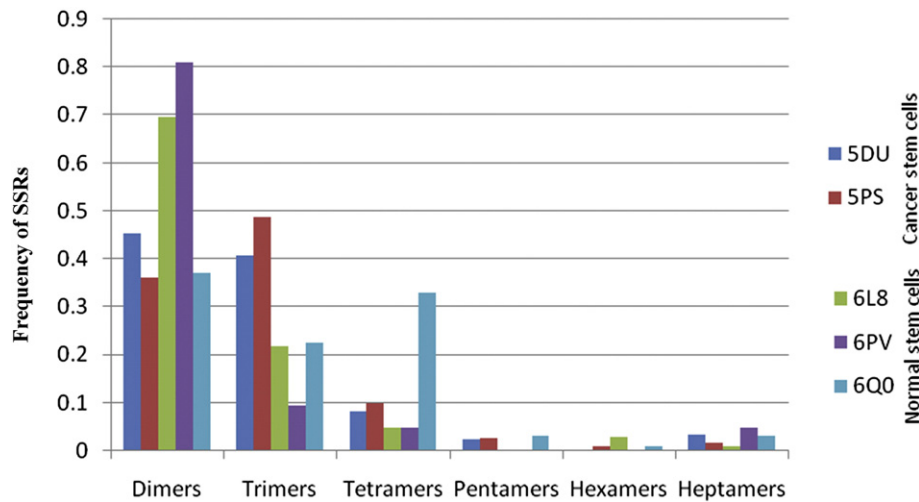


Fig. 1. The distribution of EST-SSRs in normal and cancer stem cells. The frequencies of dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, hexanucleotides and heptanucleotides have been compared between cancer and normal stem cell libraries. Trinucleotide SSR repeats are abundant in the ESTs of cancer stem cells.

SSR fragments during virtual PCR. The primer sequences are presented in Supporting information S2.

3.5. GO enrichment analysis of EST-SSRs

To explore the functions of the EST sequences with SSRs in both normal and cancer stem cells, BLAST2GO was used to search the annotation of EST-SSR containing genes in the non-redundant (nr) protein databank of NCBI. A total of 170 out of 189 EST-SSRs in cancer stem cells and 58 out of 214 sequences in normal stem cells had significant hits (Supporting information S3).

A comparative functional annotation of EST-SSRs between normal and cancer stem cells is presented in Tables 4 and 5. The complete list is presented in Supplementary 4. In addition, annotated proteins for trinucleotide EST-SSRs have been compared between normal and cancer stem cells in Supplementary 4. Presented results in Tables 4 and 5 highlight that most of the identified GO terms in normal and cancer stem cells are associated with regulatory mechanisms and cytosolic process respectively. Interestingly, genes targeted by expressed SSRs significantly ($p = 0.05$) contribute to two apoptosis related GOs: negative regulation of programmed cell death and regulation of apoptosis (GO:0043069 and GO:0042981). GO analysis highlights the possible function of expressed SSRs in cancer cells in regulating apoptosis.

Genes such as the nuclear factor- κ B (NF- κ B) transcription factor, nucleophosmin, zinc finger homeobox protein 4, cysteine-rich pdz-binding protein, inhibitor of DNA binding 2 protein and ubiquitin-like protein contained expressed SSRs on their sequences.

Table 2

Distribution comparison of different types of EST-SSR in each of dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide and heptanucleotide repeat levels between normal and cancer stem cells by Fisher's exact test. Means of SSR distributions in cancer and normal stem cell libraries were calculated for each SSR level.

SSR level*	Cancer stem cell libraries (percentage average)	Normal stem libraries (percentage average)	p-Value of difference**
Dinucleotides	40.1%	56.5%	$p = 0.01$
Trinucleotides	45.1%	21.1%	$p = 0.01$
Tetranucleotides	9.1%	17.0%	$p = 0.05$
Pentanucleotide	2.5%	1.3%	NS
Hexanucleotide	0.5%	1.7%	NS
Heptanucleotide	2.5%	2.2%	NS

* Types of SSRs in each level are presented in Supplementary 2.

** NS: not significant, $p = 0.05$: significant at the 5% level, and $p = 0.01$: highly significant at the 1% level.

3.6. Amino acid distribution of ESTs containing trinucleotide tandem repeats

Regarding the finding of this study on the importance of expressed triplet repeats in cancer stem cells, the types of predicted amino acids and their distributions in ESTs with trinucleotide repeats were analyzed in normal and cancerous libraries. In line with the observed difference in the mRNA level, the expression pattern of EST-SSRs at the amino acid level was quite different (Table 6). F-test (Bartlett's test) statistically confirmed the different amino acid patterns at $p = 0.05$.

The number of amino acid repeats was approximately 1.5 times higher in cancer stem cells than in normal stem cells (588 versus 393 repeats, respectively, Table 6). In addition, the type of expressed amino acids was notably different between cancer stem cells and normal stem cells (Table 6). Asp (37%) and Ile (20%) were the most abundant amino acids in cancer stem cells. Noticeably, these amino acids were not found in normal stem cells. In contrast, Arg (25%) was more abundant in normal stem cells, compared to cancer stem cells.

4. Discussion

The normal hematopoietic stem cells differentiate into the hematopoietic lineage giving rise to erythrocytes, platelets, leukocytes, and granulocytes. However, these normal stem cells or their progenitors sometimes undergo unwanted mutations and change to cancer stem cells. Cancer stem cells are similar to normal stem cells in the view of self-renewal and high proliferation. The origin of cancer stem cells has been the subject of considerable research in recent years for therapeutic issues and eradication of cancer stem cells.

Here, we analyzed the functional alteration of genes between normal stem cells and cancer stem cells at the level of transcriptome by identifying SSRs in their ESTs and tracing the consequences in the amino acid level. To this end, the number of different types of SSRs and their distributions were identified in both normal stem cell and cancer stem cell libraries. Then, the amino acid composition of translated EST-SSRs and functional annotation of generated EST-SSRs and their GO enrichments were determined in both normal stem cells and cancer stem cells.

Two dinucleotides, including AC/GT and TA/TA, were higher in cancer stem cells than normal stem cells. There is some evidence that AC/GT and TA/TA contents and repeat expansion are involved in human cancers and genetic disorders. Davis and Russell (1993) reported a length polymorphism of TA dinucleotides in the 3' untranslated region of the SRD5A2 gene (5- α reductase). SRD5A2 is an enzyme expressed

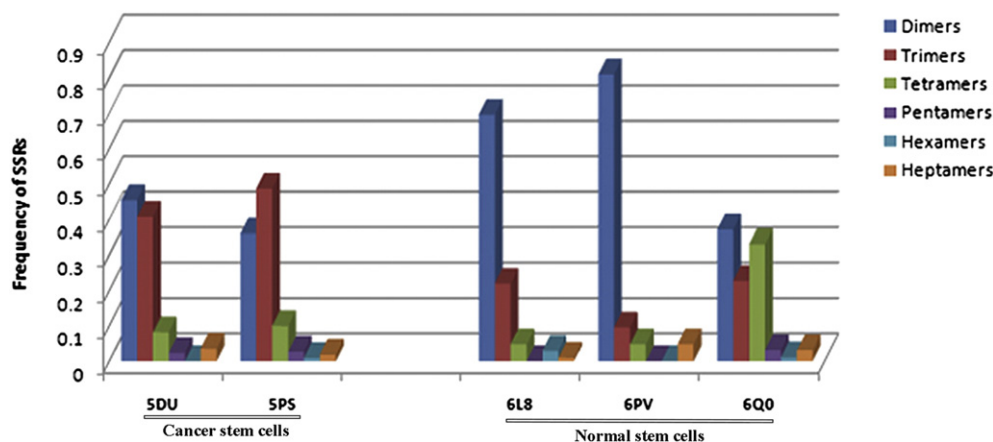


Fig. 2. The distribution of EST-SSRs within normal and cancer stem cell lines. The frequencies of dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, hexanucleotides and heptanucleotides have been compared within cancer and normal stem cell libraries. Trinucleotide SSR repeats are abundant in the ESTs of cancer stem cells.

in androgen-dependent tissues which may play a significant role in breast cancer (Wilson et al., 1993). Variations in the length of these dinucleotide repeats have been reported to affect the enzymatic activity of SRD5A2 and prostate-specific antigen concentration (Bharaj et al., 2000). AC/GT polymorphism is also demonstrated in intron 27 of the human neurofibromatosis type 1 gene (Lazaro et al., 1994).

GAT was a remarkably frequent trinucleotide in cancer stem cells. Interestingly, this triplet has not been reported previously in any other cancer. Previous studies have shown that most of the cancers and genetic disorders in animals and plants are rich in GC content. For example, CGG repeats are frequent in chronic myeloid leukemia and acute lymphoblastic leukemia (Horwitz et al., 1996; Panzer et al., 1995), and CAG repeats have high frequency in familial leukemia. This motif is common in some transcriptional activators and helicases and encodes polyglutamine or polyserine (Horwitz et al., 1996; Muchardt and Yaniv, 1993; Seipel et al., 1994). Kozłowski et al. (2010) found a frequency of 1.1% for GAT in human exons. They also showed that the frequencies of GAT in ORF and 3'UTR are 76% and 21%, respectively. However, GAT was not found significantly at 5'UTR. Toth et al. (2000) analyzed the abundance of 501 SSR motifs in different eukaryotic taxonomic groups including human as a major representative species in primates. The distribution of SSR triple repeats demonstrated that GAT was infrequent in exons, introns and intragenic regions. However, EST-SSR analysis on cancer stem cells in this study identified a novel triplet GAT in cancers but not CGG or CAG. Our finding strongly suggests that

microsatellite polymorphism in the primary stage of cancer growth, i.e. at the level of cancer stem cells, may be completely different from the progressive level. Consequently, this finding not only supports stem cells' origin in acute and chronic myeloid leukemia but also gives important novel clues regarding the genes involved in an earlier stage of cancer stem cell formation based on the hit genes by identified SSRs and GO enrichment analysis results.

GO analysis of the identified genes by EST-SSR discerned a number of common key GOs in cancer (Missiaglia et al., 2004) including cell proliferation, signal transduction, cell growth, anti-apoptosis and cell cycle. However, there was a particular GO that verified the results of our EST-SSR analysis: NF- κ B transcription factors (a hallmark of leukemia stem cells). NF- κ B transcription factors have different roles in integrating multiple stress stimuli and regulating innate and adaptive immune responses (Karin et al., 2006). Experimental evidences have unraveled the specific mechanisms by which NF- κ B influences self-sufficiency in growth signals, insensitivity to growth-inhibitors, evasion of apoptosis, limitless replication potential, tissue invasion and metastasis, and sustained angiogenesis in cancer. Activation of NF- κ B has been observed in many cancers, including breast cancer (Chua et al., 2007), melanoma (Yang et al., 2007), lung cancer (Tew et al., 2008), colon cancer (Scartozzi et al., 2007), multiple myeloma (Annunziata et al., 2007), pancreatic cancer (Weichert et al., 2007), esophageal adenocarcinoma (Izzo et al., 2007), and various types of leukemia (Fabre et al., 2007; Rae et al., 2007; Vilimas et al., 2007) and lymphoma (Zhang et al.,

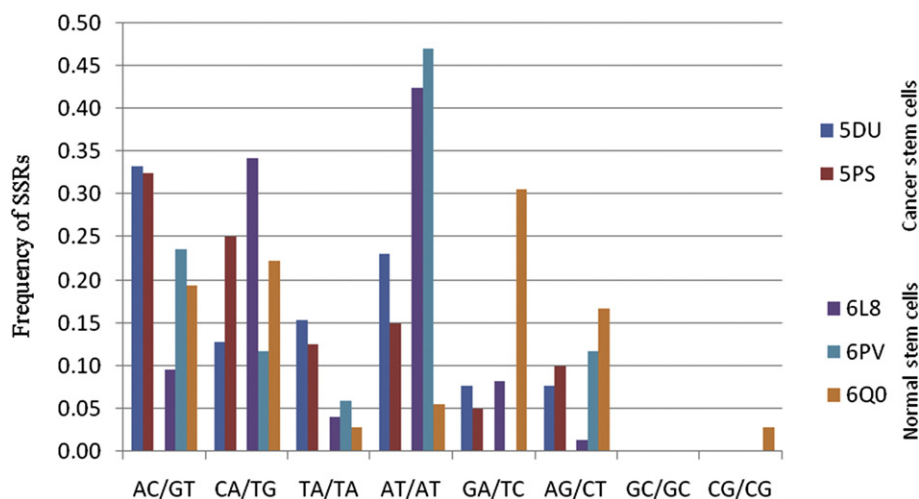


Fig. 3. The distribution of different sequences of expressed dinucleotide SSRs in normal and cancer stem cells. AC/GT and TA/TA are more frequent in cancer stem cells than in normal stem cells.

Table 3

Correlation between cancer and normal stem cells in generating different types of SSRs in each SSR class (the Pearson correlation test on percentage of different SSR types was used).

SSR level	Correlation between cancer and normal stem cells	p-Value*
Dinucleotides	+ 60.1%	NS
Trinucleotides	0.0%	NS
Tetranucleotides	− 3.0%	NS

Types of SSRs in each level are presented in Supplementary 2.

* NS: not significant.

2007; Zou et al., 2007). Activation of NF- κ B in AML cells is a hallmark of the presence of leukemic stem cells (Jordan, 2007).

It has also been reported that NF- κ B inhibits apoptosis and induces apoptosis resistance, leading to chemotherapeutic failures (Rahman et al., 2007; Singh et al., 2007; Van Antwerp et al., 1996). Interestingly, EST-SSR implicated negative regulation of cell death (Table 4, Supplementary 4).

Tat (Trans activator of transcription) protein binding was one of the significant GOs in our analysis. The Tat protein is encoded by lentiviruses such as human immunodeficiency virus type 1 (HIV-1). Tat is an extracellular protein released from infected cells and causes progression of the disease. Interestingly, there are various significant GOs involved in cell–cell communication including membrane bound organelle transport and RNA export and import. Together, such evidences derived from EST-SSR analysis implicate the presence of the factors penetrating from cancer stem cells to normal stem cells. The Tat protein interacts with its RNA target sequence (TAR) and in combination with cellular factors acts to increase human immunodeficiency virus type 1 (HIV-1) transcription (Arya et al., 1985; Dayton et al., 1986). The Tat protein increases binding of NF- κ B to its consensus sequence in the viral promoter. On the other hand, the Tat binding protein is also significant among GOs. The roles of Tat binding proteins in human cancer have been previously reported. Interestingly, human immunodeficiency virus Tat-binding protein-1 (TBP-1) is also involved in transcriptional regulation and with a supposed role in the control of cell proliferation (Pollice et al., 2004).

Another interesting gene identified by expressed SSRs is nucleophosmin (NPM). Approximately one third of acute myeloid leukemias (AMLs) are characterized by aberrant cytoplasmic localization of nucleophosmin, consequent to mutations in the NPM putative nuclear localization signal (Alcalay et al., 2005). Another differential expressed GO in cancer stem cells was associated with the regulation of phosphorylation of proteins in. Phosphorylation of NF- κ B proteins is necessary for optimal induction of NF- κ B target genes (Viatour et al., 2005).

Additionally, DNA topoisomerase, protein kinase 1, zinc finger homeobox protein 4, and transport protein homolog 4 were among the genes identified by EST-SSR. Based on the obtained results in this

Table 4

Gene Ontology (GO) analysis of the EST-SSR containing genes in cancer stem cells compared to normal stem cells. The complete list of GOs is presented in Supplementary 4.

Term	GO ID	p-Value
GO:2001141	Regulation of RNA biosynthetic process	9.65E−04
GO:0006355	Regulation of transcription, DNA-dependent	9.65E−04
GO:0043231	Intracellular membrane-bounded organelle	0.008331
GO:0045859	Regulation of protein kinase activity	0.009415
GO:0042325	Regulation of phosphorylation	0.009415
GO:0008285	Negative regulation of cell proliferation	0.015933
GO:0008284	Positive regulation of cell proliferation	0.015933
GO:0023051	Regulation of signaling	0.015933
GO:0051092	Positive regulation of NF-kappaB transcription factor activity	0.026554
GO:0030957	Tat protein binding	0.026554
GO:0005547	Phosphatidylinositol-3,4,5-trisphosphate binding	0.026554
GO:0043069	Negative regulation of programmed cell death	0.031278
GO:0042981	Regulation of apoptosis	0.031278

Table 5

Gene Ontology (GO) analysis of the identified genes with EST-SSRs in normal stem cells compared to cancer stem cells. The complete list of GOs is presented in Supplementary 4.

GO-ID	Term	p-Value
GO:0006412	Translation	2.03E−05
GO:0006414	Translational elongation	9.14E−05
GO:0022626	Cytosolic ribosome	0.004157
GO:0044445	Cytosolic part	0.004157
GO:0035270	Endocrine system development	0.004157
GO:0031016	Pancreas development	0.004157
GO:0034623	Cellular macromolecular complex disassembly	0.004157
GO:0043241	Protein complex disassembly	0.004157
GO:0071845	Cellular component disassembly at cellular level	0.004157
GO:0006415	Translational termination	0.004157
GO:0019538	Protein metabolic process	0.015493
GO:0030529	Ribonucleoprotein complex	0.015997
GO:0044444	Cytoplasmic part	0.017783
GO:0008152	Metabolic process	0.031278
GO:0042277	Peptide binding	0.034833
GO:0030942	Endoplasmic reticulum signal peptide binding	0.034833
GO:0045900	Negative regulation of translational elongation	0.034833
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	0.034833
GO:0017148	Negative regulation of translation	0.034833
GO:0008312	7S RNA binding	0.034833
GO:0006417	Regulation of translation	0.034833
GO:0051248	Negative regulation of protein metabolic process	0.034833
GO:0022625	Cytosolic large ribosomal subunit	0.037495
GO:0005829	Cytosol	0.03835

study, we suggest that these proteins are intercellularly active in cancer stem cells.

Comparison of normal stem cells with cancer stem cells highlighted different GO categories in normal stem cells, including translation, translational elongation, and the cytosolic part. Heat shock protein 90, ribosomal protein, transcription factor 3c and signal recognition particle were of some hit genes in normal stem cells (Supplementary 4). Generally most of the annotated GO groups in normal stem cells were related

Table 6

Comparison of the amino acid distributions on EST-SSRs containing trinucleotide tandem repeats between normal and cancer stem cells (the distribution was different by F-test and Levene's test at $p = 0.05$).

Amino acid type	Number of amino acid loci		Number of amino acid repeats		Percentage of different types of amino acid repeats to the total number of repeats	
	Cancer	Normal	Cancer	Normal	Cancer	Normal
Ala	7	7	49	47	8.25	11.96
Arg	2	14	12	84	2.02	21.37
Asn	2	3	12	20	2.02	5.09
Asp	34	0	204	0	34.34	0.00
Cys	0	1	0	8	0.00	2.04
Gln	0	2	0	11	0.00	2.80
Glu	2	2	12	13	2.02	3.31
Gly	3	0	19	0	3.20	0.00
His	0	1	0	6	0.00	1.53
Ile	18	0	110	0	18.52	0.00
Leu	14	10	114	98	19.19	24.94
Lys	1	6	6	33	1.01	8.40
Met	0	0	0	0	0.00	0.00
Phe	0	0	0	0	0.00	0.00
Pro	2	1	11	7	1.85	1.78
Ser	4	7	33	57	5.56	14.50
Thr	1	0	6	0	1.01	0.00
Trn	0	0	0	0	0.00	0.00
Tyr	0	1	0	9	0	2.29
Val	0	0	0	0	0	0
Total	90	55	588	393		

to translation, cytosolic and protein metabolic processes. In contrast, in cancer stem cells, GOs were mostly associated with transcription process and nuclear events. This finding is consistent with previous reports (Wilhelm et al., 2011). Extracellular factors are reported to be involved in human stem cell regulation, as shown for the hematopoietin receptor c-mpl and its ligand thrombopoietin and the cytokine tyrosine kinase receptor C-KIT like FMS-like tyrosine kinase 3 and its ligand (Buza-Vidas et al., 2009; Geissler and Russell, 1983; Kimura et al., 1998; Miller et al., 1997; Pulikkan et al., 2010; Qian et al., 2007; Thoren et al., 2008). Transcription factors have been discovered to be the key targets of mutation in acute myeloid leukemia (AML) (Rosenbauer and Tenen, 2007), such as the CCAAT enhancer binding protein alpha (C/EBP α) or E2F3 (Pulikkan et al., 2010). It should be noted that even a low transient expression level of transcription factors can alter the expression of many genes due to the ability of transcription factors in binding to promoter regions of many genes (Mahdi et al., 2013).

Chromatin modification and p53 class mediator were the other identified GO groups in cancer stem cells. MOZ associates with AML1 (RUNX1), PU.1, and p53, and cooperatively activates target gene transcription. MOZ (monocytic leukemia zinc finger protein) is a Myst (MOZ, Ybf2 (Sas3), Sas2, Tip60)-type histone acetyltransferase (HAT) that generates fusion genes, such as *MOZ-TIF2*, *MOZ-CBP* and *MOZ-p300*, in acute myeloid leukemia (AML) by chromosomal translocation (Katsumoto et al., 2008).

As it has been discussed recently, GO-based gene discovery offers a new informative approach in gene selection, leading to determine the key element of functional genomics (Fruzangohar et al., 2013, 2014; Seema, 2004). The promoter regions of overhit genes by SSR instability can be analyzed in future studies on a genome-wide scale to find the possible transcription factor binding sites, discover similar cancer affective genes and grasp an opportunity to find new candidates with similar promoter architecture (Hosseinpour et al., 2013; Kozłowski et al., 2010).

In coding regions, trinucleotide repeats are in fact amino acid runs. Because each amino acid is encoded by one or more synonymous codons, we analyzed how trinucleotide SSRs have contributed to the single amino acid runs. EST-SSRs were also translated into amino acids. Asp (coded by GAT) and Ile (coded by ATC) were the most abundant amino acids in cancer stem cells. In contrast, Arg was abundant in normal stem cells. The length of the amino acid tracts encoded by SSRs may affect the protein–protein interactions of transcription factors (Bacolla et al., 2008). Microsatellites are known to participate in both gene and protein functions (Bacolla et al., 2008). We (Ebrahimi, 2010) previously showed that the frequency of certain structural amino acid properties, such as Ile-Ile, can efficiently and precisely predict and discriminate malignant from benign breast-cancer cells. In another study (Bakhtiarzadeh et al., 2011), we demonstrated that Arg, Pro, Ser, Gly, Lys and Thr were the most abundant amino acids in lung-cancer tissues.

Here for the first time, comparative analysis of EST-SSR in cancer stem cells was carried out. The analysis identified a novel triplet tandem repeat which exactly hit the NF- κ B, a well-known marker of leukemia cancer stem cells. Such tagged ESTs demonstrated differential changes in amino acid and protein levels. This study introduces new clues for understanding the regulatory mechanism of cancer stem cells. Recently it has been described that genome-wide localized small molecules can provide new insights to therapeutic drugs. EST-SSRs can also be considered as potential hits for this new method of drug discovery through highlighting functional parts of genome which are associated with tumors or diseases (Anders et al., 2013). This study opens a new window on functional genetics of cancer stem cells due to their central function based on instable EST-SSRs which are of significant importance in the success of cancer therapy. The interaction between EST-SSRs and cancer over-representing transcription factors such as NF- κ B has the potential to develop a new layer of understanding cancer biology. Future laboratory experiments are needed to examine the relationship between NF- κ B, apoptosis, and differentially expressed SSRs in cancer stem cells.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.gene.2014.07.069>.

References

- Alcalay, M., Tiacci, E., Bergomas, R., Bigerna, B., Venturini, E., Minardi, S.P., Meani, N., Diverio, D., Bernard, L., Tizzoni, L., Volorio, S., Luzi, L., Colombo, E., Lo Coco, F., Mecucci, C., Falini, B., Pelicci, P.G., 2005. Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc + AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* 106, 899–902.
- Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., Clarke, M.F., 2003. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3983–3988.
- Alvarez, A., Lokeshwar, V.B., 2007. Bladder cancer biomarkers: current developments and future implementation. *Curr. Opin. Urol.* 17, 341–346.
- Anders, L., Guenther, M.G., Qi, J., Fan, Z.P., Marineau, J.J., Rahl, P.B., Loven, J., Sigova, A.A., Smith, W.B., Lee, T.I., Bradner, J.E., Young, R.A., 2013. Genome-wide localization of small molecules. *Nat. Biotechnol.* <http://dx.doi.org/10.1101/004911>.
- Annunziata, C.M., Davis, R.E., Demchenko, Y., Bellamy, W., Gabrea, A., Zhan, F., Lenz, G., Hanamura, I., Wright, G., Xiao, W., Dave, S., Hurt, E.M., Tan, B., Zhao, H., Stephens, O., Santra, M., Williams, D.R., Dang, L., Barlogie, B., Shaughnessy Jr., J.D., Kuehl, W.M., Staudt, L.M., 2007. Frequent engagement of the classical and alternative NF- κ B pathways by diverse genetic abnormalities in multiple myeloma. *Cancer Cell* 12, 115–130.
- Arya, S.K., Guo, C., Josephs, S.F., Wong-Staal, F., 1985. Trans-activator gene of human T-lymphotropic virus type III (HTLV-III). *Science* 229, 69–73.
- Bacolla, A., Larson, J.E., Collins, J.R., Li, J., Milosavljevic, A., Stenson, P.D., Cooper, D.N., Wells, R. D., 2008. Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. *Genome Res.* 18, 1545–1553.
- Bakhtiarzadeh, M.R., Ebrahimi, M., Ebrahimi, E., 2011. Discovery of EST-SSRs in lung cancer: tagged ESTs with SSRs lead to differential amino acid and protein expression patterns in cancerous tissues. *PLoS ONE* 6, e27118.
- Bakhtiarzadeh, M.R., Arefnejad, B., Ebrahimi, E., Ebrahimi, M., 2012. Application of functional genomic information to develop efficient EST-SSRs for the chicken (*Gallus gallus*). *Genet. Mol. Res.* 11, 1558–1574.
- Bharaj, B., Scorilas, A., Gai, M., Diamandis, E.P., 2000. TA repeat polymorphism of the Salpa-reductase gene and breast cancer. *Cancer Epidemiol. Biomarkers Prev.* 9, 387–393.
- Bhatia, R., Holtz, M., Niu, N., Gray, R., Snyder, D.S., Sawyers, C.L., Arber, D.A., Slovak, M.L., Forman, S.J., 2003. Persistence of malignant hematopoietic progenitors in chronic myelogenous leukemia patients in complete cytogenetic remission following imatinib mesylate treatment. *Blood* 101, 4701–4707.
- Bonnet, D., Dick, J.E., 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* 3, 730–737.
- Brouwer, J.R., Willemsen, R., Oostra, B.A., 2009. Microsatellite repeat instability and neurological disease. *Bioessays* 31, 71–83.
- Buza-Vidas, N., Cheng, M., Duarte, S., Charoudeh, H.N., Jacobsen, S.E., Sitnicka, E., 2009. FLT3 receptor and ligand are dispensable for maintenance and posttransplantation expansion of mouse hematopoietic stem cells. *Blood* 113, 3453–3460.
- Chua, H.L., Bhat-Nakshatri, P., Clare, S.E., Morimiya, A., Badve, S., Nakshatri, H., 2007. NF- κ B represses E-cadherin expression and enhances epithelial to mesenchymal transition of mammary epithelial cells: potential involvement of ZEB-1 and ZEB-2. *Oncogene* 26, 711–724.
- Davis, D.L., Russell, D.W., 1993. Unusual length polymorphism in human steroid 5 alpha-reductase type 2 gene (SRD5A2). *Hum. Mol. Genet.* 2, 820.
- Dayton, A.I., Sodroski, J.G., Rosen, C.A., Goh, W.C., Haseltine, W.A., 1986. The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* 44, 941–947.
- Deng, S.S., Ding, Z.M., Qian, J.H., Lu, W.G., Wan, X.Y., Xie, X., 2007. Study on parental origin of complete hydatidiform mole detected by polymorphisms of microsatellite. *Zhonghua Fu Chan Ke Za Zhi* 42, 468–471.
- Ebrahimi, M., Ebrahimi, E., Shamabadi, Narges, Ebrahimi, Mahdi, 2010. Are there any differences between features of proteins expressed in malignant and benign breast cancers? *J. Res. Med. Sci.* 15, 299–309.
- Fabre, C., Carvalho, G., Tasdemir, E., Braun, T., Ades, L., Grosjean, J., Boehrer, S., Metivier, D., Souquere, S., Pierron, G., Fenaux, P., Kroemer, G., 2007. NF- κ B inhibition sensitizes to starvation-induced cell death in high-risk myelodysplastic syndrome and acute myeloid leukemia. *Oncogene* 26, 4071–4083.
- Fruzangohar, M., Ebrahimi, E., Ogunniyi, A.D., Mahdi, L.K., Paton, J.C., Adelson, D.L., 2013. Comparative GO: a web application for comparative gene ontology and gene ontology-based gene selection in bacteria. *PLoS ONE* 8, e58759.
- Fruzangohar, M., Ebrahimi, E., Adelson, D.L., 2014. Application of global transcriptome data in gene ontology classification and construction of a gene ontology interaction network. *bioRxiv*. <http://dx.doi.org/10.1101/004911>.
- Geissler, E.N., Russell, E.S., 1983. Analysis of the hematopoietic effects of new dominant spotting (W) mutations of the mouse. I. Influence upon hematopoietic stem cells. *Exp. Hematol.* 11, 452–460.
- Grady, W.M., Carethers, J.M., 2008. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* 135, 1079–1099.
- Graham, S.M., Jorgensen, H.G., Allan, E., Pearson, C., Alcorn, M.J., Richmond, L., Holyoake, T.L., 2002. Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to ST1571 in vitro. *Blood* 99, 319–325.

- Guan, Y., Gerhard, B., Hogge, D.E., 2003. Detection, isolation, and stimulation of quiescent primitive leukemic progenitor cells from patients with acute myeloid leukemia (AML). *Blood* 101, 3142–3149.
- Haddad, L.A., Parra, F.C., Pena, S.D., 1998. Characterization and mapping of four novel human expressed polymorphic trinucleotide microsatellites. *Gene* 223, 369–374.
- Hemmati, H.D., Nakano, I., Lazareff, J.A., Masterman-Smith, M., Geschwind, D.H., Bronner-Fraser, M., Kornblum, H.I., 2003. Cancerous stem cells can arise from pediatric brain tumors. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15178–15183.
- Holyoake, T., Jiang, X., Eaves, C., Eaves, A., 1999. Isolation of a highly quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. *Blood* 94, 2056–2064.
- Hope, K.J., Jin, L., Dick, J.E., 2004. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat. Immunol.* 5, 738–743.
- Horwitz, M., Goode, E.L., Jarvik, G.P., 1996. Anticipation in familial leukemia. *Am. J. Hum. Genet.* 59, 990–998.
- Hosseinpour, B., Bakhtiarzadeh, M.R., Khosravi, P., Ebrahimi, E., 2013. Predicting distinct organization of transcription factor binding sites on the promoter regions: a new genome-based approach to expand human embryonic stem cell regulatory network. *Gene* 531, 212–219.
- Izzo, J.G., Malhotra, U., Wu, T.T., Luthra, R., Correa, A.M., Swisher, S.G., Hofstetter, W., Chao, K.S., Hung, M.C., Ajani, J.A., 2007. Clinical biology of esophageal adenocarcinoma after surgery is influenced by nuclear factor-kappaB expression. *Cancer Epidemiol. Biomarkers Prev.* 16, 1200–1205.
- Jaiswal, S., Traver, D., Miyamoto, T., Akashi, K., Lagasse, E., Weissman, I.L., 2003. Expression of BCR/ABL and BCL-2 in myeloid progenitors leads to myeloid leukemias. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10002–10007.
- Janatova, M., Pohlreich, P., 2004. Microsatellite markers in breast cancer studies. *Prague Med. Rep.* 105, 111–118.
- Jordan, C.T., 2007. The leukemic stem cell. *Best Pract. Res. Clin. Haematol.* 20, 13–18.
- Karin, M., Lawrence, T., Nizet, V., 2006. Innate immunity gone awry: linking microbial infections to chronic inflammation and cancer. *Cell* 124, 823–835.
- Katsumoto, T., Yoshida, N., Kitabayashi, I., 2008. Roles of the histone acetyltransferase monocytic leukemia zinc finger protein in normal and malignant hematopoiesis. *Cancer Sci.* 99, 1523–1527.
- Kim, C.F., Jackson, E.L., Woolfenden, A.E., Lawrence, S., Babar, I., Vogel, S., Crowley, D., Bronson, R.T., Jacks, T., 2005. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* 121, 823–835.
- Kimura, S., Roberts, A.W., Metcalf, D., Alexander, W.S., 1998. Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin. *Proc. Natl. Acad. Sci. U. S. A.* 95, 1195–1200.
- Kozlowski, P., de Mezer, M., Krzyzosiak, W.J., 2010. Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.* 38, 4027–4039.
- Lazaro, C., Gaona, A., Estivill, X., 1994. Two CA/GT repeat polymorphisms in intron 27 of the human neurofibromatosis (NF1) gene. *Hum. Genet.* 93, 351–352.
- Li, Y.C., Korol, A.B., Fahima, T., Nevo, E., 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21, 991–1007.
- Mahdi, L.K., Ebrahimi, E., Adelson, D.L., Paton, J.C., Oguniyi, A.D., 2013. A transcription factor contributes to pathogenesis and virulence in *Streptococcus pneumoniae*. *PLoS ONE* 8, e70862.
- Manly, B.F., 2004. *Multivariate Statistical Methods: A Primer*. CRC Press.
- Miller, L.D., Liu, E.T., 2007. Expression genomics in breast cancer research: microarrays at the crossroads of biology and medicine. *Breast Cancer Res.* 9, 206.
- Miller, C.L., Rebel, V.I., Helgason, C.D., Lansdorp, P.M., Eaves, C.J., 1997. Impaired steel factor responsiveness differentially affects the detection and long-term maintenance of fetal liver hematopoietic stem cells in vivo. *Blood* 89, 1214–1223.
- Missiaglia, E., Blaveri, E., Terris, B., Wang, Y.H., Costello, E., Neoptolemos, J.P., Crnogorac-Jurcic, T., Lemoine, N.R., 2004. Analysis of gene expression in cancer cell lines identifies candidate markers for pancreatic tumorigenesis and metastasis. *Int. J. Cancer* 112, 100–112.
- Muchardt, C., Yaniv, M., 1993. A human homologue of *Saccharomyces cerevisiae* SNF2/SWI2 and *Drosophila* brm genes potentiates transcriptional activation by the glucocorticoid receptor. *EMBO J.* 12, 4279–4290.
- Oda, S., Zhao, Y., Maehara, Y., 2005. Microsatellite instability in gastrointestinal tract cancers: a brief update. *Surg. Today* 35, 1005–1015.
- Panzer, S., Kuhl, D.P., Caskey, C.T., 1995. Unstable triplet repeat sequences: a source of cancer mutations? *Stem Cells* 13, 146–157.
- Pollice, A., Nasti, V., Ronca, R., Vivo, M., Lo Iacono, M., Calogero, R., Calabro, V., La Mantia, G., 2004. Functional and physical interaction of the human ARF tumor suppressor with Tat-binding protein-1. *J. Biol. Chem.* 279, 6345–6353.
- Pulikkan, J.A., Peramangalam, P.S., Dengler, V., Ho, P.A., Preudhomme, C., Meshinchi, S., Christopheit, M., Nibourel, O., Muller-Tidow, C., Bohlander, S.K., Tenen, D.G., Behre, G., 2010. C/EBPalpha regulated microRNA-34a targets E2F3 during granulopoiesis and is down-regulated in AML with CEBPA mutations. *Blood* 116, 5638–5649.
- Qian, H., Buza-Vidas, N., Hyland, C.D., Jensen, C.T., Antonchuk, J., Mansson, R., Thoren, L.A., Ekblom, M., Alexander, W.S., Jacobsen, S.E., 2007. Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell Stem Cell* 1, 671–684.
- Rae, C., Langa, S., Tucker, S.J., MacEwan, D.J., 2007. Elevated NF-kappaB responses and FLIP levels in leukemic but not normal lymphocytes: reduction by salicylate allows TNF-induced apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12790–12795.
- Rahman, K.M., Ali, S., Aboukameel, A., Sarkar, S.H., Wang, Z., Philip, P.A., Sakr, W.A., Raz, A., 2007. Inactivation of NF-kappaB by 3,3'-diindolylmethane contributes to increased apoptosis induced by chemotherapeutic agent in breast cancer cells. *Mol. Cancer Ther.* 6, 2757–2765.
- Romualdi, C., Bortoluzzi, S., d'Alessi, F., Danieli, G.A., 2003. IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. *Physiol. Genomics* 12, 159–162.
- Rosenbauer, F., Tenen, D.G., 2007. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat. Rev. Immunol.* 7, 105–117.
- Scartozzi, M., Bearzi, I., Pierantoni, C., Mandolesi, A., Loupakis, F., Zaniboni, A., Catalano, V., Quadri, A., Zorzi, F., Berardi, R., Biscotti, T., Labianca, R., Falcone, A., Cascinu, S., 2007. Nuclear factor-kB tumor expression predicts response and survival in irinotecan-refractory metastatic colorectal cancer treated with cetuximab-irinotecan therapy. *J. Clin. Oncol.* 25, 3930–3935.
- Seema, T., 2004. Microsatellites (SSRs): puzzles within puzzle. *Indian J. Biotechnol.* 3, 331–347.
- Seipel, K., Georgiev, O., Gerber, H.P., Schaffner, W., 1994. Basal components of the transcription apparatus (RNA polymerase II, TATA-binding protein) contain activation domains: is the repetitive C-terminal domain (CTD) of RNA polymerase II a "portable enhancer domain"? *Mol. Reprod. Dev.* 39, 215–225.
- Shamloo-Dashtpazgerdi, R., Razi, H., Lindlöf, A., Niazi, A., Dadkhodaie, A., Ebrahimi, E., 2013. Comparative analysis of expressed sequence tags (ESTs) from *Triticum monococcum* shoot apical meristem at vegetative and reproductive stages. *Genes Genom.* 35, 365–375.
- Sharma, R.K., Bhardwaj, P., Negi, R., Mohapatra, T., Ahuja, P.S., 2009. Identification, characterization and utilization of unigenes derived microsatellite markers in tea (*Camellia sinensis* L.). *BMC Plant Biol.* 9, 53.
- Singh, S.K., Clarke, I.D., Terasaki, M., Bonn, V.E., Hawkins, C., Squire, J., Dirks, P.B., 2003. Identification of a cancer stem cell in human brain tumors. *Cancer Res.* 63, 5821–5828.
- Singh, S., Shi, Q., Bailey, S.T., Palczewski, M.J., Pardee, A.B., Iglehart, J.D., Biswas, D.K., 2007. Nuclear factor-kappaB activation: a molecular therapeutic target for estrogen receptor-negative and epidermal growth factor receptor family receptor-positive human breast cancer. *Mol. Cancer Ther.* 6, 1973–1982.
- Slate, J., Hale, M.C., Birkhead, T.R., 2007. Simple sequence repeats in zebra finch (*Taeniopygia guttata*) expressed sequence tags: a new resource for evolutionary genetic studies of passerines. *BMC Genomics* 8, 52.
- Soltysova, A., Altanerova, V., Altaner, C., 2005. Cancer stem cells. *Neoplasma* 52, 435–440.
- Tew, G.W., Lorimer, E.L., Berg, T.J., Zhi, H., Li, R., Williams, C.L., 2008. SmgGDS regulates cell proliferation, migration, and NF-kappaB transcriptional activity in non-small cell lung carcinoma. *J. Biol. Chem.* 283, 963–976.
- Thoren, L.A., Liuba, K., Bryder, D., Nygren, J.M., Jensen, C.T., Qian, H., Antonchuk, J., Jacobsen, S.E., 2008. Kit regulates maintenance of quiescent hematopoietic stem cells. *J. Immunol.* 180, 2045–2053.
- Tinelli, A., Mezzolla, V., Leo, G., Pisano, M., Storelli, F., Alemanno, G., Malvasi, A., Tommasi, S., Ronzino, G., Lorusso, V., 2010. Microsatellite instability (MSI) as genomic markers in endometrial cancer: toward scientific evidences. *Mini-Rev. Med. Chem.* 10, 1356–1365.
- Toth, G., Gaspari, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10, 967–981.
- Van Antwerp, D.J., Martin, S.J., Kafri, T., Green, D.R., Verma, I.M., 1996. Suppression of TNF-alpha-induced apoptosis by NF-kappaB. *Science* 274, 787–789.
- Venkatesan, R.N., Bielas, J.H., Loeb, L.A., 2006. Generation of mutator mutants during carcinogenesis. *DNA Repair (Amst)* 5, 294–302.
- Vergara, D., Tinelli, A., Martignago, R., Malvasi, A., Chiuri, V.E., Leo, G., 2010. Biomolecular pathogenesis of borderline ovarian tumors: focusing target discovery through proteogenomics. *Curr. Cancer Drug Targets* 10, 107–116.
- Viatour, P., Merville, M.P., Bours, V., Chariot, A., 2005. Phosphorylation of NF-kappaB and IkappaB proteins: implications in cancer and inflammation. *Trends Biochem. Sci.* 30, 43–52.
- Vilimas, T., Mascarenhas, J., Palomero, T., Mandal, M., Buonamici, S., Meng, F., Thompson, B., Spaulding, C., Macaroun, S., Alegre, M.L., Kee, B.L., Ferrando, A., Miele, L., Aifantis, I., 2007. Targeting the NF-kappaB signaling pathway in Notch1-induced T-cell leukemia. *Nat. Med.* 13, 70–77.
- Weichert, W., Boehm, M., Gekeler, V., Bahra, M., Langrehr, J., Neuhaus, P., Denkert, C., Imre, G., Weller, C., Hofmann, H.P., Niesporek, S., Jacob, J., Dietel, M., Scheidert, C., Kristiansen, G., 2007. High expression of RelA/p65 is associated with activation of nuclear factor-kappaB-dependent signaling in pancreatic cancer and marks a patient population with poor prognosis. *Br. J. Cancer* 97, 523–530.
- Wilhelm, B.T., Briau, M., Austin, P., Faubert, A., Boucher, G., Chagnon, P., Hope, K., Girard, S., Mayotte, N., Landry, J.R., Hebert, J., Sauvageau, G., 2011. RNA-seq analysis of 2 closely related leukemia clones that differ in their self-renewal capacity. *Blood* 117, e27–e38.
- Wilson, J.D., Griffin, J.E., Russell, D.W., 1993. Steroid 5 alpha-reductase 2 deficiency. *Endocr. Rev.* 14, 577–593.
- Yang, J., Pan, W.H., Clawson, G.A., Richmond, A., 2007. Systemic targeting inhibitor of kappaB kinase inhibits melanoma tumor growth. *Cancer Res.* 67, 3127–3134.
- Zhang, B., Wang, Z., Li, T., Tsitsikov, E.N., Ding, H.F., 2007. NF-kappaB2 mutation targets TRAF1 to induce lymphomagenesis. *Blood* 110, 743–751.
- Zou, P., Kawada, J., Pesnick, L., Cohen, J.L., 2007. Bortezomib induces apoptosis of Epstein-Barr virus (EBV)-transformed B cells and prolongs survival of mice inoculated with EBV-transformed B cells. *J. Virol.* 81, 10029–10036.