

Novel candidate genes may be possible predisposing factors revealed by whole exome sequencing in familial esophageal squamous cell carcinoma

Narjes Forouzanfar¹, Ancha Baranova², Saman Milanizadeh¹,
Alireza Heravi-Moussavi³, Amir Jebelli¹ and
Mohammad Reza Abbaszadegan^{1,4}

Abstract

Esophageal squamous cell carcinoma is one of the deadliest of all the cancers. Its metastatic properties portend poor prognosis and high rate of recurrence. A more advanced method to identify new molecular biomarkers predicting disease prognosis can be whole exome sequencing. Here, we report the most effective genetic variants of the Notch signaling pathway in esophageal squamous cell carcinoma susceptibility by whole exome sequencing. We analyzed nine probands in unrelated familial esophageal squamous cell carcinoma pedigrees to identify candidate genes. Genomic DNA was extracted and whole exome sequencing performed to generate information about genetic variants in the coding regions. Bioinformatics software applications were utilized to exploit statistical algorithms to demonstrate protein structure and variants conservation. Polymorphic regions were excluded by false-positive investigations. Gene–gene interactions were analyzed for Notch signaling pathway candidates. We identified novel and damaging variants of the Notch signaling pathway through extensive pathway-oriented filtering and functional predictions, which led to the study of 27 candidate novel mutations in all nine patients. Detection of the trinucleotide repeat containing 6B gene mutation (a slice site alteration) in five of the nine probands, but not in any of the healthy samples, suggested that it may be a susceptibility factor for familial esophageal squamous cell carcinoma. Noticeably, 8 of 27 novel candidate gene mutations (e.g. epidermal growth factor, signal transducer and activator of transcription 3, MET) act in a cascade leading to cell survival and proliferation. Our results suggest that the trinucleotide repeat containing 6B mutation may be a candidate predisposing gene in esophageal squamous cell carcinoma. In addition, some of the Notch signaling pathway genetic mutations may act as key contributors to esophageal squamous cell carcinoma.

Keywords

Esophageal squamous cell carcinoma, Notch signaling pathway, whole exome sequencing, predisposing factor

Date received: 7 August 2016; accepted: 24 December 2016

Introduction

Esophageal squamous cell carcinoma (ESCC) is the eighth most common cancer, with more than 480,000 new cases and 400,000 deaths each year worldwide.¹ One of the world's highest rates of ESCC is found in northeastern Iran, which lies in the esophageal cancer belt that stretches from China to northeastern Iran.² The reasons for such high rates of ESCC occurrence in this area are not fully clear. ESCC is the second leading cause of cancer-related

¹Medical Genetics Research Center, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

²School of System Biology, George Mason University, Fairfax, VA, USA

³Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada

⁴Division of Human Genetics, Immunology Research Center, Avicenna Research Institute, Mashhad University of Medical Sciences, Mashhad, Iran

Corresponding author:

Mohammad Reza Abbaszadegan, Medical Genetics Research Center, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad 9177948564, Iran.

Email: Abbaszadeganmr@mums.ac.ir



deaths in northeastern Iran. Statistical data have demonstrated that its rates with age standardized in males and females are $(5.6-10.30)/10^5$ and $(5.7-19.8)/10^5$, respectively.^{2,3} Despite advances in multimodality therapy, the aggressiveness of human squamous cell carcinomas (SCCs) caused poor prognosis with average 5-year survival rates <10%. SCCs are usually diagnosed in the advanced stages and respond poorly to chemotherapy; thus, novel therapeutic strategies in the treatment of this disease are required.^{4,5} ESCC is a multifactorial disease; in addition to environmental factors, genetic variations contribute to its susceptibility risk. Better understanding of the molecular mechanisms underlying esophageal carcinogenesis will lead to the development of new diagnostic technologies, treatment modalities, and preventive approaches.⁵

For the past several decades, the genes underlying Mendelian disorders have been identified through physical mapping, positional cloning, candidate gene sequencing, and a process of meiotic mapping.⁶ Recently, the utilization of next-generation sequencing (NGS) technologies has facilitated interrogation of the genomes of human cancers.⁷ Whole exome sequencing (WES) allows a targeted enrichment and resequencing of nearly all exons of protein-coding genes. Accordingly, causative mutations in Mendelian disorders have been provided by WES using next generation technologies.⁸

The Notch signaling pathway is crucial for the development and homeostasis of most tissues and is evolutionarily conserved. It plays critical roles in balancing cell survival, proliferation, differentiation, and apoptosis, as well as maintenance of progenitor stem cell populations. Aberrant activation of Notch signaling has been revealed during the carcinogenesis of a variety of human cancers.^{9,10}

Clinical trials of Notch signaling pathway inhibitors have been reported in patients with solid tumors, including breast cancer and T-cell acute lymphoblastic leukemia (T-ALL),¹⁰ and preclinical evaluation of several approaches is under consideration.¹¹ However, the roles of the Notch signaling pathway in esophageal tumor biology remain elusive.¹² Because of its main role in triggering T-ALL, Notch is often considered as a model hematopoietic proto-oncogene.¹³

It was shown that the risk of developing esophageal cancer is greater in individuals with first degree relatives with ESCC than in the general population. To be more specific, offspring of two parents with ESCC has an eight-fold greater likelihood of developing ESCC than offspring of parents without ESCC.¹⁴

Currently, genetic analyses of diseases, such as cancer, have focused on the detection of rare variants that may contribute to individual susceptibility to common complex diseases.¹⁵ It is a matter of debate whether several DNA sequence variants with high penetrance are the major

contributors to disease. Based on the common disease–rare variant (CDRV) hypothesis, we aimed to find novel mutations in ESCC that can be introduced as biomarkers to be used for early diagnosis, prognosis, and possible novel therapeutics.¹⁶

Materials and methods

Patients

Nine ESCC probands from nine unrelated families participated in this study. All samples came from individuals residing in Khorasan or Golestan provinces in northeastern Iran, located in the ESCC belt. The patients' age ranged from 30–73 years. They were histologically confirmed to have ESCC at Omid or Emam Reza Hospitals between 2009 and 2011. All the patients had at least two consecutive generations of familial cancer or two or more affected family members in a generation. After enrollment, patients' clinical information was collected and genomic DNA was extracted from peripheral blood, followed by exome sequencing.

Ethical approval. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent. Informed consent was obtained from all individual participants included in this study.

Exome sequencing

DNA sample concentrations were 50–300 ng/μL (measured by NanoDrop, A260/280 = 1.7–2), and 3–5 μg of DNA was utilized per reaction (Thermo Fisher Scientific, Wilmington, USA). Covaris S2 equipment was applied to shear the DNA, resulting in an optimal range of fragment-size distribution. Library size and concentration were assessed by capillary electrophoresis (Bioanalyzer; Agilent, Santa Clara, CA). The SureSelect Target Enrichment System was utilized for post-capture pooling of enriched sequencing-ready libraries. The whole exome was characterized using the Illumina Genome Analyzer II (Illumina, San Diego, CA). Mapping generated reads against the human reference genome University of California, Santa Cruz (UCSC) hg19 were obtained with the Burrows-Wheeler Alignment (BWA) tool alignment algorithm, which efficiently aligned short sequence reads against a large reference sequence (rs). Reads were processed by SAMtools to detect single-nucleotide polymorphisms (SNPs) and insertions/deletions (indels) and to remove polymerase chain reaction (PCR) duplicates (<http://samtools.sourceforge.net>).

Data analysis

Real-Time Analysis (RTA) software (RTA V.1.12-4) was started on-instrument, which is integrated with the Genome Analyzer Sequencing Control software (SCS). It is followed by off-instrument data analysis with Illumina Consensus Assessment of Sequence and Variation (CASAVA) software (V1.8.2). Genes associated with the Notch signaling pathway were selected according to GeneCards (<http://www.genecards.org/>) software. Variants with total depth (>15) and those with allelic frequencies $\geq 0.01\%$ were filtered based on the data available in the 1000 Genomes Project (<http://www.1000genomes.org>) and dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>). Only variants in exonic region revealing indels, non-synonymous, and stop-gain changes were selected. Deleterious amino acid changes were selected with PolyPhen (score >0.85 ; <http://genetics.bwh.harvard.edu/pph2>) and Sorting Intolerant from Tolerant (SIFT; score <0.05 ; <http://sift.jcvi.org/>). Only non-synonymous amino acid changes were selected. Indels, which illustrate damaging deletions, were identified with MutationTaster (score >0.5 ; <http://www.mutationtaster.org>) and Protein Variation Effect Analyzer (PROVEAN; <http://provean.jcvi.org/index.php>). Moreover, the functional impact of amino acid substitutions in proteins was predicted with MutationTaster. Terms and previous bibliographies of variants according to National Center for Biotechnology Information (NCBI) Gene and Reactome (<http://www.reactome.org/PathwayBrowser/>) were evaluated for biological functions and pathways of the gene products.

The initial results may have contained false-positive reports, which were omitted using the method by Fuentes Fajardo et al.¹⁷ and Esteban-Jurado et al.¹⁸ Results were compared with the data of 51 healthy individuals (100 alleles) to eliminate possible polymorphisms.

To generate exact data, novel mutations from SIFT and 1000 Genomes Project were analyzed with the Exome Aggregation Consortium (ExAC) database, and mutations with allelic frequencies under 1×10^{-4} were selected (<http://exac.broadinstitute.org/>). GNCPro (<http://gncpro.sabiosciences.com/>), STRING (<http://string-db.org/>), and GeneMANIA (<http://www.genemania.org/>) software were used to demonstrate gene–gene interactions and explain the effects of other genes on the selected candidate mutations.

The miRBase database is utilized to search published micro RNA (miRNA) sequences and annotations to predict hairpins in the miRNA transcripts with information on the location and sequence of the mature miRNA sequence (www.mirbase.org).

After detection of a candidate gene in family 93 (CD3D), molecular analyses was performed to determine its familial variation. Primers for mutation detection with amplification-refractory mutation system–PCR (ARMS-PCR) were designed using Primer3web (<http://primer3.ut.ee/>), and checked with AlleleID and Gene Runner software applications. To confirm mutation detection and

WES results, samples were sequenced using Sanger technique. Molecular analysis of CD3D in family 93 was used to evaluate genetic segregation.

Results

Whole exomes from nine probands with familial ESCC from nine unrelated families were sequenced. Clinical features of the ESCC patients are shown in Table 1. Pedigrees of six of the ESCC patients are shown in Figure 2. Of note, patients 68-1, 149-1, and 150-1 consumed hot tea, which is a known risk factor for ESCC.

After exome sequencing, about 600,000 variants in all cases were detected where only exonic mutations were filtered in the following steps. All levels of filtering are depicted in Figure 1. Notch signaling pathway genes were selected from GeneCards and compared with our data to extract non-synonymous, indels, and stop-gain variants. Moreover, subjects were filtered a second time to remove variants with total depth greater than 15 and allelic frequency less than 0.01, which found 232 mutations in the nine probands. Variants were detected as damaging if reported in PolyPhen and SIFT/PROVEAN for point mutations and PROVEAN and MutationTaster for indels.

Manual inspection of 68 variants resulted in no false positives among our data. We found 45 mutations that did not have rs numbers in 1000 Genomes Project or were novel in SIFT. After all filtration procedures, final results were reached by analyzing with ExAC, which eliminated another 18 more mutations, leaving 27 mutations as exonic, novel, low allelic frequency, damaging, total depth >15 , and without rs numbers. Comparing these results with 51 normal genome sequences further confirmed that these 27 mutations are perhaps novel mutations and not polymorphisms. The final prioritized variants are shown in Supplementary 1.

All the mutations reported by WES were checked on the Notch signaling pathway to demonstrate the fundamental genes, with relevance scores of 10–14 in GeneCards (<http://www.genecards.org/>), in this pathway. Among the main genes in this category four of nine mutations were revealed as damaging, indicating a need for further analysis of Notch signaling pathway genes.

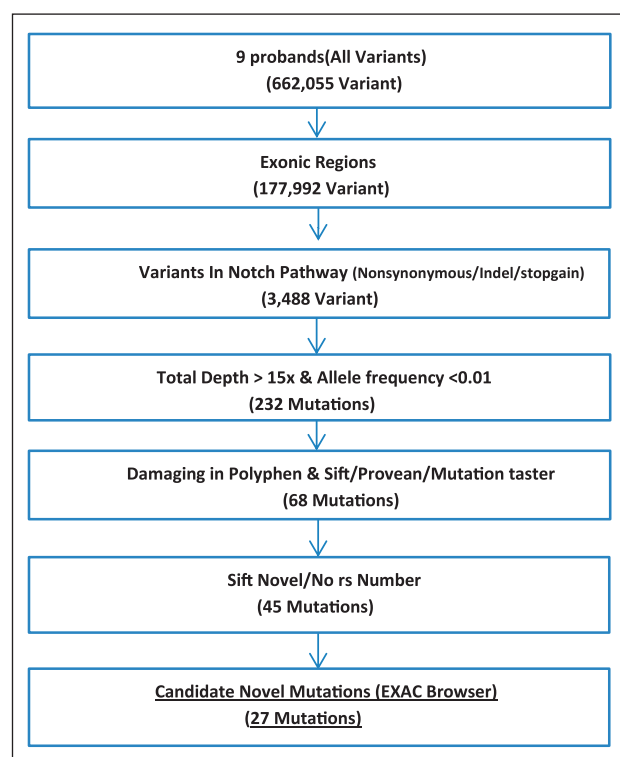
Gene–gene interaction

Four candidate gene products, epidermal growth factor (EGF), MET, AKT1, and growth differentiation factor 15 (GDF15) act through a cascade that downregulates GDF15 (growth factor) gene expression. All these genes are proto-oncogenes, and their genetic variations may cause cancer susceptibility. Detailed data on the distribution of mutated genes in each family are shown in Figure 2. Familial ESCC criteria were confirmed in all pedigrees.

Table 1. Clinical features and cancer history of patients.

Patient number	Age	Gender	Tumor grade	Smoking state	Family history
68-I	50	Female	III	No	Father and older sister with ESCC.
90-I	53	Female	III	No	Mother with unknown cancer, uncle with ESCC.
91-I	43	Female	III	No	Sister with ESCC at the age of 49.
93-I	54	Female	III	No	Sister with ESCC under the age of 47 and proband encountered breast cancer 18 months after diagnosis.
94-I	54	Male	III	Yes	Mother and older brother with ESCC. Father and son of patient with leukemia.
95-I	47	Female	III	No	Grandfather and two of aunts with ESCC.
105-I	73	Male	III	Yes	Sister with ESCC.
149-I	50	Male	III	No	Father and both the uncles with ESCC.
150-I	60	Male	III	Yes	Mother and brother with ESCC.
Characteristics	53.77	55.5% (male) 44.5% (female)	100%	0.33% (yes)	

ESCC: esophageal squamous cell carcinoma.

**Figure 1.** Schematic data analysis after whole exome sequencing. Nine patients from different families were sequenced for familial ESCC.

Although these variants appeared in different pedigrees, only EGF and AKT1 were altered in the same proband (149-I). Moreover, signal transducer and activator of transcription 3 (STAT3) mutated in proband 91 is upregulated by EGF (mutated in proband 149) and downregulated by MET (mutated in proband 150) (<http://gncpro.sabiosciences.com/>). Detailed interactions among candidate genes are

depicted in Figure 3. In conclusion, in this study the end results of all mutations found in various genes in the Notch pathway resulted in downregulation of GDF15 gene. Three other co-expressed candidate genes were CD3D in family 93, CD3E in family 94, and LCK in family 95.

Segregation study

Among these variants, the CD3D gene was analyzed in two individuals in family 93. The observed mutation resulted in a serine to cysteine substitution in cytoplasmic domain of the protein, which is involved in T-cell development and signal transduction. It was demonstrated that this mutation segregates through the generations (pedigree 93).

This mutation was transferred to three of the five healthy children of the proband. Neither the proband's affected sister nor did her children inherited this mutation (Figure 4). The proband developed a primary breast tumor 2 years after she was first diagnosed with ESCC.

Trinucleotide repeat containing 6B

After various filtration procedures, trinucleotide repeat containing 6B (TNRC6B) gene variation was detected in five of nine probands. This mutation has a rs number; however, it was not found on the ExAC database. The mutation was a deleterious CAG repeat deletion found in probands 68-I, 93-I, 105-I, 149-I, and 94-I. Of cases with mutations, only sample 94-I had six-nucleotide deletion, two repeats of CAG; however, others had three-nucleotide deletions, only one repeat unit. This variant results in a polyglutamine and alters the length of the amino acid repeat. Amino acid sequence analysis of mutated TNRC6B demonstrated that protein features might be affected by alterations in the splice site. These alterations change the donor sequence of the splice site in this region and increase the rate of splicing (MutationTaster data: wild type = 0.33/mutant = 1).

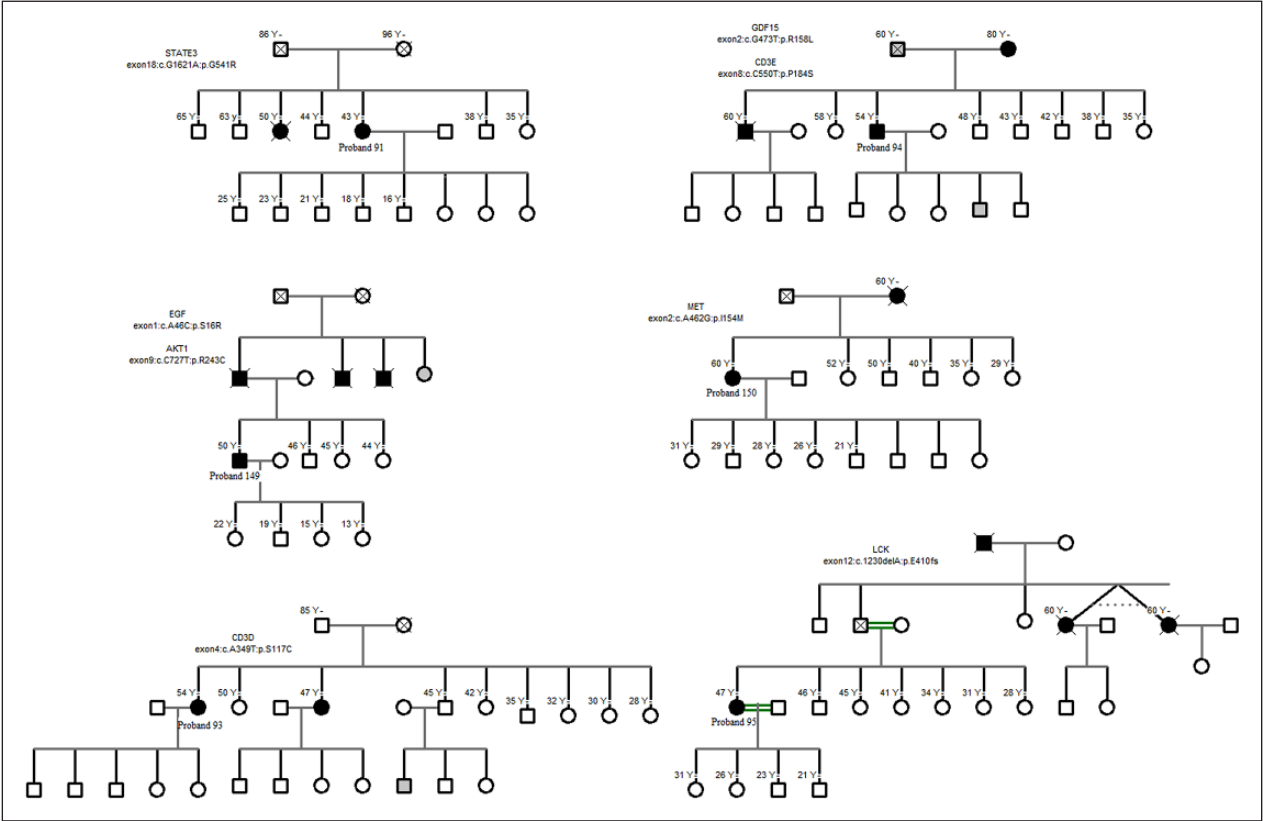


Figure 2. Pedigrees from families 91, 94, 149, 150, 93, and 95 showing the key Notch signaling pathway mutations in probands of each pedigree.

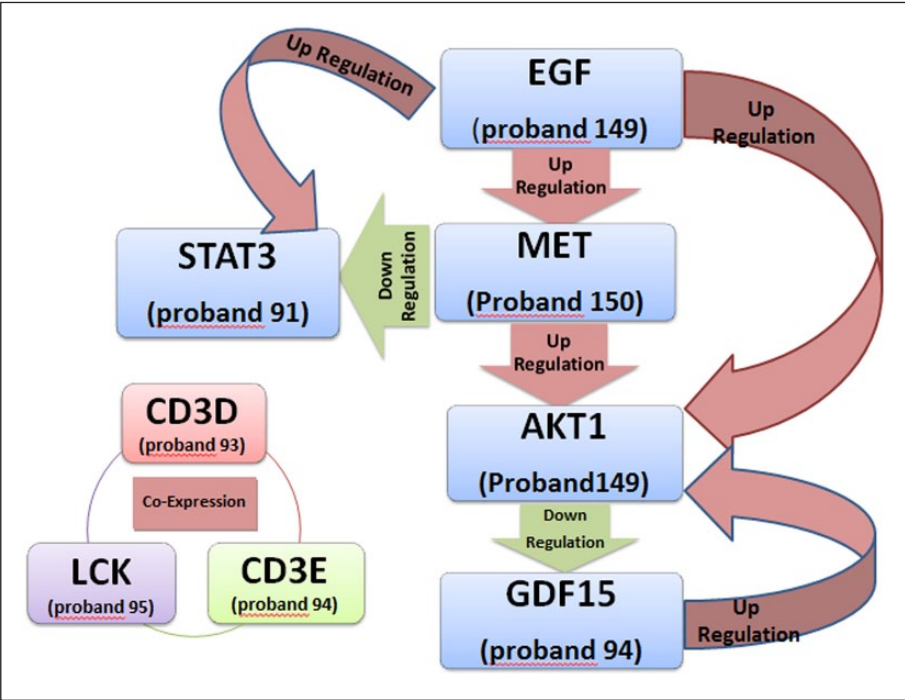


Figure 3. Gene-gene interactions among ESCC candidate genes.

messenger RNA (mRNA); so, it does not undergo miRNA-mediated decay and causes extensive cell proliferation due to its overexpression (www.mirbase.org). Further studies are needed to confirm this phenomenon. It is predicted that this deletion may cause changes in histone modification sites, open chromatin, or transcription factor binding sites such as the histone H3 lysine 36 tri-methylation and Pax5 transcription factor binding site.

The CAG deletion mutation in TNRC6B exists in the RNA recognition motif in the silencing domain and interacts with the CNOT1 and PAN3 regions of the protein, which may inhibit protein interaction with the miRNAs (UniProtKB: Q9UPQ9 (TNRC6B_HUMAN)).

Previous studies revealed that the expression of the immune-related gene, CD3D, was associated with overall survival in breast cancer patients.³ The CD3D genetic variant as an appropriate ESCC candidate was studied in family 93 based on the online databases. It demonstrated genetic segregation among the family.

This observation suggests that the CD3D mutation may be an important genetic factor for new primary breast tumors in ESCC patients and confirms the prognostic significance of CD3D gene expression in breast cancer.³¹ CD3D was shown to form a complex with many genes, including the cancer-related genes IKZF1 and TRA@, and activates modification of complexes in the cancer-related genes interleukin 2 (IL2), tyrosine-protein phosphatase non-receptor type 11 (PTPN11), and casitas B-lineage lymphoma (CBL).³²

In conclusion, exome sequencing strategies should be carefully selected and modified depending on what the data set allows.³³ Candidates for implication in ESCC predisposition have been identified after rigorous scrutiny and filtration from the final ESCC variant list. This study identified several genes that may be associated with susceptibility to ESCC, including TNRC6B, CD3D, CD3E, LCK, STAT3, MET, EGF, AKT1, and GDF15. Furthermore, to enhance the efficiency of whole exome studies, improved methods, such as utilizing a provisional gene exclusion list, must be implemented.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Zhou K, Zhang S-S, Yan Y, et al. Overexpression of transient receptor potential vanilloid 2 is associated with poor prognosis in patients with esophageal squamous cell carcinoma. *Med Oncol* 2014; 31(7): 17.
2. Arnold M, Soerjomataram I, Ferlay J, et al. Global incidence of oesophageal cancer by histological subtype in 2012. *Gut* 2015; 64: 381–387.
3. Rouhollahi MR, Mohagheghi MA, Mohammadrezaei N, et al. Situation analysis of the National Comprehensive Cancer Control Program in the I.R. of Iran; assessment and recommendations based on the IAEA imPACT mission. *Arch Iran Med* 2014; 17(4): 222–231.
4. Gholamin M, Moaven O, Memar B, et al. Overexpression and interactions of interleukin-10, transforming growth factor β , and vascular endothelial growth factor in esophageal squamous cell carcinoma. *World J Surg* 2009; 33(7): 1439–1445.
5. Hu YC, Lam KY, Law S, et al. Profiling of differentially expressed cancer-related genes in esophageal squamous cell carcinoma (ESCC) using human cancer cDNA arrays overexpression of oncogene MET correlates with tumor differentiation in ESCC. *Clin Cancer Res* 2001; 7(11): 3519–3525.
6. Biasecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet* 2010; 42(1): 13–14.
7. Streppel M, Lata S, Delabastide M, et al. Next-generation sequencing of endoscopic biopsies identifies ARID1A as a tumor-suppressor gene in Barrett's esophagus. *Oncogene* 2014; 33(3): 347–357.
8. Diaz-Horta O, Duman D, Foster J II, et al. Whole-exome sequencing efficiently detects rare mutations in autosomal recessive nonsyndromic hearing loss. *PLoS One* 2012; 7: e50628.
9. Forghanifard MM, Taleb S and Abbaszadegan MR. Notch signaling target genes are directly correlated to esophageal squamous cell carcinoma tumorigenesis. *Pathol Oncol Res* 2015; 21(2): 463–467.
10. Lobry C, Oh P, Mansour MR, et al. Notch signaling: switching an oncogene to a tumor suppressor. *Blood* 2014; 123(16): 2451–2459.
11. Andersson ER and Lendahl U. Therapeutic modulation of Notch signalling—are we there yet? *Nat Rev Drug Discov* 2014; 13(5): 357–378.
12. Natsuizaka M, Ohashi S, Naganuma S, et al. Notch regulates squamous differentiation, cell plasticity and tumor heterogeneity in esophageal carcinoma. *Cancer Res* 2011; 71(Suppl. 8): 5194.
13. Lobry C, Oh P and Aifantis I. Oncogenic and tumor suppressor functions of Notch in cancer: it's NOTCH what you think. *J Exp Med* 2011; 208(10): 1931–1935.
14. Chen T, Cheng H, Chen X, et al. Family history of esophageal cancer increases the risk of esophageal squamous cell carcinoma. *Sci Rep* 2015; 5: 16038.
15. Friedberg EC, Walker GC, Siede W, et al. *DNA repair and mutagenesis*. Washington, DC: American Society for Microbiology Press, 2005.
16. Schork NJ, Murray SS, Frazer KA, et al. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 2009; 19(3): 212–219.
17. Fuentes Fajardo KV, Adams D, Mason CE, et al. Detecting false-positive signals in exome sequencing. *Hum Mutat* 2012; 33(4): 609–613.
18. Esteban-Jurado C, Vila-Casadesús M, Garre P, et al. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet Med* 2015; 17(2): 131–142.

19. Hicks C, Pannuti A and Miele L. Associating GWAS information with the Notch signaling pathway using transcription profiling. *Cancer Inform* 2011; 10: 93–108.
20. Agrawal N, Jiao Y, Bettegowda C, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov* 2012; 2(10): 899–905.
21. Mendelsohn J and Baselga J. The EGF receptor family as targets for cancer therapy. *Oncogene* 2000; 19(56): 6550–6565.
22. Kamran MZ, Patil P and Gude RP. Role of STAT3 in cancer metastasis and translational advances. *Biomed Res Int* 2013; 2013: 421821.
23. Shiojima I and Walsh K. Role of Akt signaling in vascular homeostasis and angiogenesis. *Circ Res* 2002; 90(12): 1243–1250.
24. Tan HY, Chen AG, Chen Q, et al. Epistatic interactions of AKT1 on human medial temporal lobe biology and pharmacogenetic implications. *Mol Psychiatry* 2012; 17(10): 1007–1016.
25. Baillat D and Shiekhhattar R. Functional dissection of the human TNRC6 (GW182-related) family of proteins. *Mol Cell Biol* 2009; 29(15): 4144–4155.
26. Fernandez-Mercado M, Pellagatti A, Di Genua C, et al. Mutations in SETBP1 are recurrent in myelodysplastic syndromes and often coexist with cytogenetic markers associated with disease progression. *Br J Haematol* 2013; 163(2): 235–239.
27. Sun J, Zheng SL, Wiklund F, et al. Sequence variants at 22q13 are associated with prostate cancer risk. *Cancer Res* 2009; 69(1): 10–15.
28. Yang Y, Li X, Yang Q, et al. The role of microRNA in human lung squamous cell carcinoma. *Cancer Genet Cytogenet* 2010; 200(2): 127–133.
29. Chiosea S, Jelezcova E, Chandran U, et al. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. *Cancer Res* 2007; 67(5): 2345–2350.
30. Murakami Y, Tamori A, Itami S, et al. The expression level of miR-18b in hepatocellular carcinoma is associated with the grade of malignancy and prognosis. *BMC Cancer* 2013; 13(1): 99.
31. Buus R, Klintman M, Sheri A, et al. Biomarkers in residual disease after neoadjuvant chemotherapy of breast cancer predict long-term outcome. *Cancer Res* 2015; 75(Suppl. 15): 5294.
32. Ahn J, Yoon Y, Park C, et al. Integrative gene network construction for predicting a set of complementary prostate cancer genes. *Bioinformatics* 2011; 27(13): 1846–1853.
33. Ma D, Whitehead P, Menold M, et al. Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism. *Am J Hum Genet* 2005; 77(3): 377–388.