

Co-occurrence: A Gene Reference Resource for Coincidental Patterns of Gene Mutations in Human Cancers

Isar Nassiri, Esmaeel Azadian, Roozbeh Sharafi and Ali Masoudi-Nejad*

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

Abstract

Cancer is a complex process in which the abnormalities of many genes appear to be involved. The synchronized patterns of gene mutations may reveal the functional relations between genes and pathways in tumor-genesis as well as identify targets for treatment. Co-occurrence database represents a comprehensive core collection of data on published coincidental mutations in nuclear genes underlying human cancers. By August 2013, the database contained 126 different coincidental lesions in 43 different genes in 25 different types of human cancers, and 8 cellular signaling pathways. In next step, this model for knowledge representation about coincidental mutations in human cancers can be extended to other synchronized patterns of genes alterations in human diseases. Although, co-occurrence database originally established for the scientific study of coincidental mutational mechanisms in human cancer genes, it is also applicable for physicians and genetic counselors. The database is freely available from <http://lbb.ut.ac.ir/project/Co-Occurrence/home.php>

Keywords: Gene Mutations; Human Cancers; Co-occurrence database; Database design; Family-specific mutational databases

Introduction

Cancers arise due to the accumulation of mutations in proto-oncogenes and tumor suppressors genes [1]. Mutations have variety effects on products of genes, including truncations, frame shifts, reduction of transcripts level, and disruption of the three dimensional protein structures [2]. These harmful alterations confer a selective advantage on the cell and its progeny for tumor genesis [3]. Knowledge of mutations is necessary to understand biology of cancer and development of molecular targeted therapies [1]. Tumor-genesis is a multi-step process and several mutations in different genes are necessary to transform normal cells into malignant [4]. There are rare examples of cancer evolution by a single gene mutation [5]. Although a single mutation can initiate the processes of tumor-genesis, in fact, multiple preliminary perturbations are required for transformation of normal cells to solid tumors including cell cycle deregulation, evasion of apoptosis, limitless replication, acquisition of genomic instability, metastasis and angiogenesis [6].

The co-occurrence database, maintained at the Laboratory of Systems Biology and Bioinformatics (LBB) in Iran, represents a comprehensive core collection of data on coincidental mutations underlying human cancers. Co-occurrence database comprises published single base-pair substitutions such as duplications, insertions, deletions, repeat expansions and indels in coding, regulatory, and splicing-relevant regions of human cancer nuclear genes. Mitochondrial genome mutations are not included. By bringing together co-occurrence database includes different recurrent combinatorial mutational patterns in 25 different human cancers.

Co-occurrence database aims to facilitate combinational mutation analysis, targeted chemotherapy, study of functional relations of genes in major cancer pathways and novel patterns in tumor-genesis. Although co-occurrence database, originally, established for the scientific study of mutational mechanisms in human cancer genes, this database can also provide information of practical importance to researchers in molecular oncology and human molecular genetics, genetic counselors, and physicians interested in a particular inherited condition in a given patient or a family. In view of its potential

usefulness, the curators of co-occurrence database made the database publicly available through the World Wide Web in March 2011.

Building the Co-Occurrence Database

Database design

Co-occurrence database was implemented as MySQL database on a Linux server. We used PHP scripts for all data retrieval and output. Its modular design is compatible with future expansion, and contains the following subsections: EntrezGene, RefSeq [7], Ensembl [8], Protein (Swiss-Prot/UniProt), Gene Ontology [9], Mutation Frequency in 25 cancer type [10,11], cancer data and protein interactions (ROCK) [3,11]. We retrieved gene sequences including genomic DNA, mRNA, and protein from UCSC browser [12]. We downloaded interaction datasets from [11]. The process annotations were done with information from *CancerGenes* [13]. Figure 1 present the diagram of entities and relationships in co-occurrence database.

Data coverage and structure

By August 2013, co-occurrence database contained 126 different recurrent combinatorial mutational patterns in 43 different genes in 25 types of human cancers (Table 1). The content of co-occurrence DB originates from published scientific literature [11]. In that article, Yeang et al. [11] retrieved the co-occurrence gene mutations in cancer by a combination of manual and computerized search procedures and statistical analysis [14]. They compiled collections of samples initially from an exhaustive search of the scientific literature from PubMed

*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@lbb.ut.ac.ir

Received July 24, 2013; Accepted September 24, 2013; Published September 27, 2013

Citation: Nassiri I, Azadian E, Sharafi R, Masoudi-Nejad A (2013) Co-occurrence: A Gene Reference Resource for Coincidental Patterns of Gene Mutations in Human Cancers. J Proteomics Bioinform 6: 197-201. doi:10.4172/jpb.1000281

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.

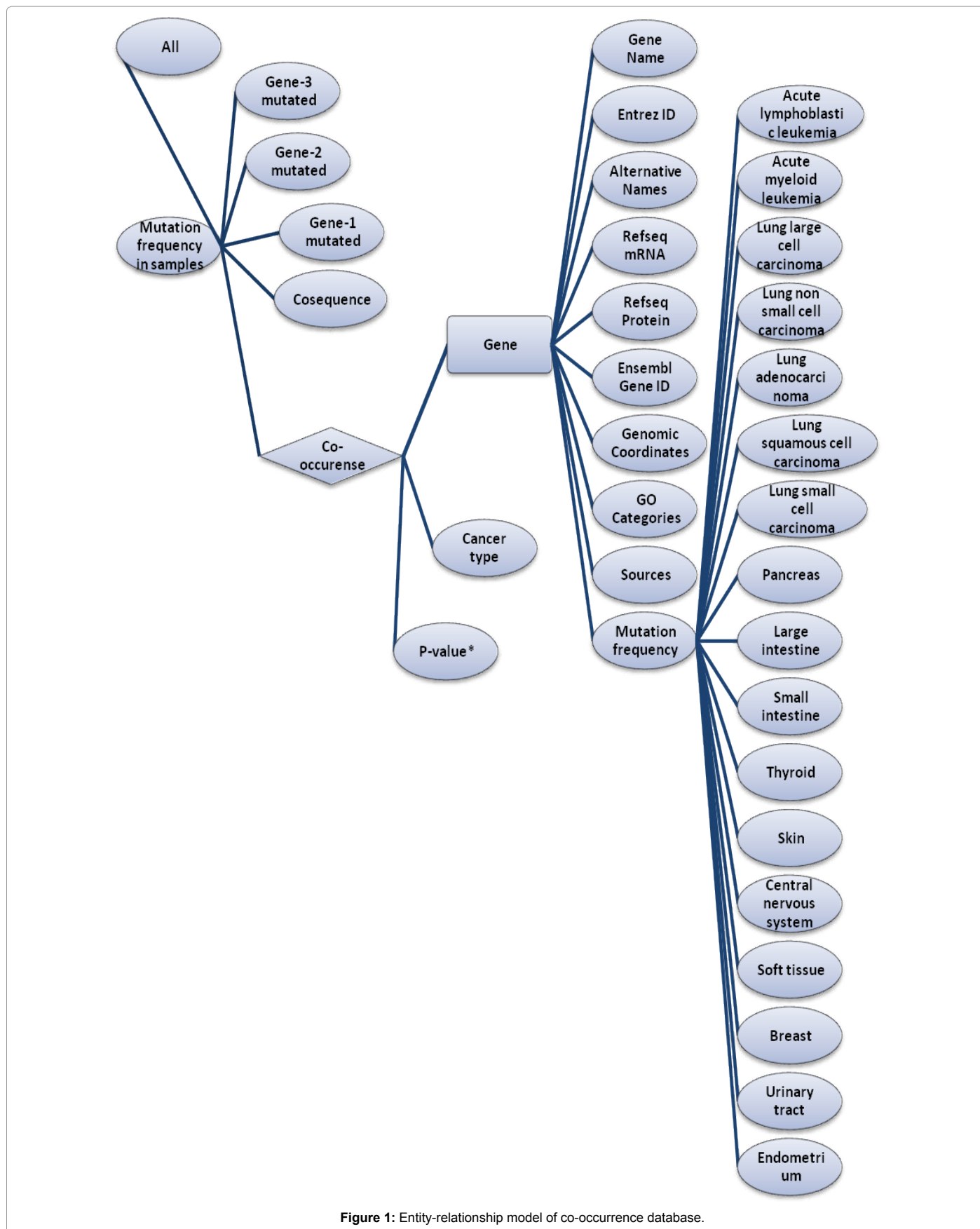


Figure 1: Entity-relationship model of co-occurrence database.

Row	Cancer Tissue	Frequency in DB	Percent
1	Acute lymphoblastic leukemia	11	8.7
2	Acute myeloid leukae	2	1.6
3	Acute myeloid leukemia	6	4.8
4	B cell lymphoma	1	0.8
5	Biliary tract	2	1.6
6	Breast	3	2.4
7	Central nervous system	8	6.3
8	Endometrium	2	1.6
9	Esophagus	1	0.8
10	Hematopoietic and lymphoid others	1	0.8
11	Kidney	2	1.6
12	Large intestine	20	15.9
13	Lung adenocarcinoma	11	8.7
14	Lung large cell carcinoma	1	0.8
15	Lung small cell carcinoma	4	3.2
16	Lung squamous cell carcinoma	1	0.8
17	Pancreas	10	7.9
18	Prostate	10	7.9
19	Skin	16	12.7
20	Soft tissue	2	1.6
21	T cell lymphoma	1	0.8
22	T cell lymphoma acute lymphoblastic	1	0.8
23	Thyroid	3	2.4
24	Upper aerodigestive tract	2	1.6
25	Urinary tract	5	4
Total		126	100

Table 1: List of types and frequencies of cancers in the co-occurrence database.

database at NCBI (<http://www.ncbi.nih.gov/pubmed>) and COSMIC database. They considered gene mutations in tumor samples which were screened at least for two genes. Finally, they classified candidate genes based on the type of cancers and mutations [11].

We can use Catalog of Somatic Mutations in Cancer (COSMIC) database to extend the co-occurrence database in next steps [15]. The COSMIC is a large database of cancer somatic gene mutations that extracted from the literatures [15]. Currently, COSMIC collects only small mutations in protein-coding regions, including missense mutations [16]. Figure 2 shows an example of co-occurrence database pages, includes the related annotations of a gene with combinatorial patterns of mutations. The list can be sorted alphabetically, according to the number of samples, number of combinatorial mutations and P-values. Gene names hyperlink to gene-annotation pages and cancer types are hyperlinked to pathways of co-occurrence mutations (Figure 2).

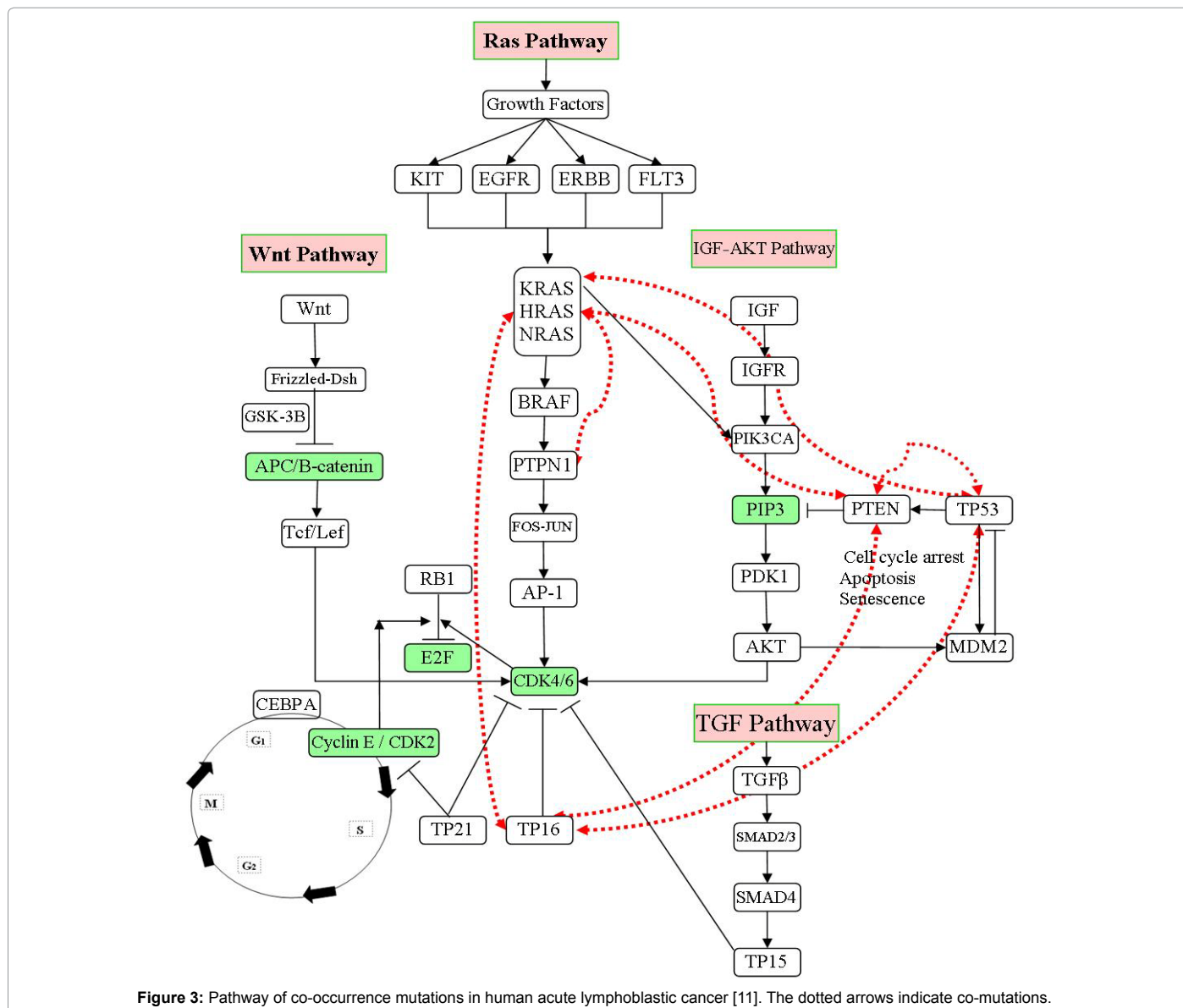
Initial entries in homepage of database are gene chromosomal location, cancer tissue type, or the 'search' box, where any text can be typed for a database search. Links from results page direct user to more details about mutations, publications, and summaries. In addition, hypertext links exist from each gene page to mutation frequencies in human cancers and list of co-occurrence mutations. All entries composed gene name, alternative names, different gene IDs, chromosomal location, GO category, references to the first literature report of mutations, and mutations frequency. User can search co-occurrence database by gene accession numbers, gene symbols, or HUGO-approved gene names.

Co-Occurrence Mutation

Co-Occurrence mutations of **KRAS** in human cancers.

Gene Name	Entrez Id	Cancer Type	Co Sequence With <i>KRAS</i>	<i>KRAS</i> Mutated	Gene-2 Mutated	Both	P-Value
NRAS	4893	Acute lymphoblastic leukemia	856	80	109	5	0.0246
NRAS	4893	Acute myeloid leukemia	578	58	159	10	0.0043
NRAS	4893	Lung adenocarcinoma	194	127	7	0	0.000473
NRAS	4893	Skin	604	20	182	4	0.0374
NRAS	4893	Prostate	109	35	8	0	0.0397
PIK3CA	5290	Large intestine	70	35	14	10	0.00335
PTEN	5728	Endometrium	59	12	24	7	0.00997
SMAD4	4089	Pancreas	157	87	37	24	5.41e-05
SMAD4	4089	Pancreas	17	16	6	3	0.048
TP53	7157	Lung adenocarcinoma	44	16	29	9	0.00415
TP53	7157	Pancreas	17	16	12	11	0.000444
TP53	7157	Large intestine	70	33	20	10	0.00335

Figure 2: List of genes with synchronized patterns of mutations with KRAS. Table includes related annotations with each synchronized pattern, including number of company-sequenced samples (co-sequenced), sample with mutations in KRAS (KRAS mutated), samples with mutations in gene-2 (gene-2 mutated), number of samples with mutation in both genes (both), and P-values.



The impact of the mutation on the cell signaling pathways and proteins interactions is a key in determining the impact of combinatorial patterns of gene mutations [17]. Towards this goal, general relations between co-mutated genes and cell signaling were provided for acute lymphoblastic, acute myeloid leukemia, central nervous system, large intestine, lung adenocarcinoma, lung small cell carcinoma, pancreas, and skin cancer (Figure 3) [11].

Discussion

Co-occurrence database has established as the database of coincidental gene mutations in human cancers to link current knowledge on co-mutated genes to human cancers and their relevance for the tumor-genesis [1,18]. For each relation, candidate genes are selected with referred examined samples and panoramic view of functional relations [2,11]. Our major focus was to provide a filtered and annotated reference set of co-occurrence of mutations in cancer linked genes. Co-occurrence database can be used in the areas of functional relations of genes with combinational mutations, gain

insight for identifying mutated driver pathways, and date and party hubs in network of cancer [17,19,20]. The co-occurrence database combines automated and expert annotation of coincidental mutations towards the specific needs of the cancer research community.

Future Developments

We designed co-occurrence database with a user-friendly interface to assess which genes malfunction were causal of most human cancers, and which co-occurrence mutations were cancer drivers. In next step, we will extend the database to include more genes, mutations, and pathway information. In our view, co-occurrence database annotation model can be applied to annotate co-mutated gene entries in other human inherited genetic diseases from family-specific mutational databases such as SH2base [21] and KinMutBase [22].

Availability

The database is freely available via the World Wide Web from <http://bb.ut.ac.ir/project/Co-Occurrence/home.php>. The database may be cited by referencing this article.

References

1. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, et al. (2007) High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 39: 347-351.
2. Reva B, Antipin Y, Sander C (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 39: e118.
3. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, et al. (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 446: 153-158.
4. Burnworth B, Arendt S, Muffler S, Steinkraus V, Bröcker EB, et al. (2007) The multi-step process of human skin carcinogenesis: a role for p53, cyclin D1, hTERT, p16, and TSP-1. *Eur J Cell Biol* 86: 763-780.
5. Daley GQ, Van Etten RA, Baltimore D (1990) Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* 247: 824-830.
6. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646-674.
7. NCBI Resource Coordinators (2013) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 41: D8-8D20.
8. Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, et al. (2009) Ensembl 2009. *Nucleic Acids Res* 37: D690-697.
9. Gene Ontology Consortium (2008) The Gene Ontology project in 2008. *Nucleic Acids Res* 36: D440-444.
10. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, et al. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* Chapter 10: Unit 10.
11. Yeang CH, McCormick F, Levine A (2008) Combinatorial patterns of somatic gene mutations in cancer. *FASEB J* 22: 2605-2622.
12. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, et al. (2013) The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* 41: D64-D69.
13. Higgins ME, Claremont M, Major JE, Sander C, Lash AE (2007) CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res* 35: D721-726.
14. Greenman C, Wooster R, Futreal PA, Stratton MR, Easton DF (2006) Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* 173: 2187-2198.
15. Gonzalez-Perez A, Mustonen V, Reva B, Ritchie GR, Creixell P, et al. (2013) Computational approaches to identify functional genetic variants in cancer genomes. *Nat Methods* 10: 723-729.
16. 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: D945-950.
17. Cui Q (2010) A network of cancer genes with co-occurring and anti-co-occurring mutations. *PLoS One* 5.
18. Jing Z, Xiaopei S, Yang Z, Zheng G, Hui X, et al. (2009) Identifying Candidate Cancer Genes Based on Their Somatic Mutations Co-Occurring with Cancer Genes in Cancer Genome Profiling. *BMEI '09. 2nd International Conference*.
19. Zhao J, Zhang S, Wu LY, Zhang XS (2012) Efficient methods for identifying mutated driver pathways in cancer. *Bioinformatics* 28: 2940-2947.
20. Muller H, Acquati F (2008) Topological properties of co-occurrence networks in published gene expression signatures. *Bioinform Biol Insights* 2: 203-213.
21. Lappalainen I, Thusberg J, Shen B, Vihinen M (2008) Genome wide analysis of pathogenic SH2 domain mutations. *Proteins* 72: 779-792.
22. Ortutay C, Väliaho J, Stenberg K, Vihinen M (2005) KinMutBase: a registry of disease-causing mutations in protein kinase domains. *Hum Mutat* 25: 435-442.