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: +98-21-6640-4880; E-mail: amasoudi@lbb.ut.ac.ir

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


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.
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 combinational 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 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://lbb.ut.ac.ir/project/Co-Occurrence/home.php. The database may be cited by referencing this article.

---

**Figure 3:** Pathway of co-occurrence mutations in human acute lymphoblastic cancer [11]. The dotted arrows indicate co-mutations.

References


