D670–D675 Nucleic Acids Research, 2010, Vol. 38, Database issue Published online 11 November 2009
doi:10.1093/nar/gkp957

Network of Cancer Genes: a web resource to analyze duplicability, orthology and network properties of cancer genes

Adnan S. Syed, Matteo D'Antonio and Francesca D. Ciccarelli*

Department of Experimental Oncology, European Institute of Oncology, IFOM-IEO Campus, Via Adamello 16, 20139 Milan, Italy

Received August 12, 2009; Revised October 2, 2009; Accepted October 13, 2009

ABSTRACT

The Network of Cancer Genes (NCG) collects and integrates data on 736 human genes that are mutated in various types of cancer. For each gene, NCG provides information on duplicability, orthology, evolutionary appearance and topological properties of the encoded protein in a comprehensive version of the human protein-protein interaction network. NCG also stores information on all primary interactors of cancer proteins, thus providing a complete overview of 5357 proteins that constitute direct and indirect determinants of human cancer. With the constant delivery of results from the mutational screenings of cancer genomes, NCG represents a versatile resource for retrieving detailed information on particular cancer genes, as well as for identifying common properties of precompiled lists of cancer genes. NCG is freely available at: http://bio.ifom-ieo-campus.it/ncg.

INTRODUCTION

Cancer is a genetic disease caused by the accumulation of deleterious modifications within the genome of somatic cells (1). During tumorigenesis, genomic instability leads to the progressive acquisition of silent ('passenger') and selected ('driver') mutations (2). The latter provide cancer cells with selective growth advantages that initiate clonal expansion (3). The Cancer Genome Project (CGP) has the ambitious goal of identifying all genes that are implicated in the development of cancer (4). The Cancer Gene Census (CGC) is a part of CGP and collects information on more than 370 genes whose mutations are causally related to cancer (5). Recently, high-throughput mutational screenings of several cancer types have been promoted with the aim of identifying mutated genes, without any hypothesis-driven bias. So far, four of these high-throughput experiments have been delivered. Overall, they identified 380 Candidate Cancer Genes (CAN-genes) that are mutated in breast, colorectal, pancreatic cancers and glioblastoma (6–8). Furthermore, the pilot experiment from the Tumor Sequencing Project identified 26 genes (TSP-genes) mutated in lung adenocarcinoma (9). Altogether, these studies revealed that the number of cancer genes is surprisingly high and they are functionally more heterogeneous than previously thought. Despite this functional heterogeneity, cancer genes tend to share 'systems-level properties' (10), such as higher connectivity and lower duplicability when compared to the rest of human genes (11–13). The presence of shared properties, which are not strictly dependent on the gene function, indicates that cancer genes are fragile components of the human gene repertoire.

A number of databases have been set up over the years to collect and organize several types of information related to cancer, such as somatic mutations of cancer genes (14), experimental evidence for their involvement in cancer (15,16) or modifications in gene expression levels (17,18). Other databases are specialized on particular types of cancer (19,20), on single genes (21,22) or on specific genomic modifications (23,24). None of the available resources, however, focuses on properties of cancer genes that are not strictly dependent on their function, but that could help in interpreting cancer as a 'systems disease'. Here, we present the Network of Cancer Genes (NCG, http://bio.ifom-ieo-campus.it/ncg), a database that stores information on systems-level properties of a comprehensive dataset of more than 730 cancer genes. The collected features are duplicability, evolutionary appearance and topological properties in the human protein–protein interaction network. Protein interactions have been successfully used to infer functional links between proteins (25). In NCG, they are used to understand how the topological properties of the cancer proteins inside the

*To whom correspondence should be addressed. Tel: +39 02574303053; Fax: +39 0294375990; Email: francesca.ciccarelli@ifom-ieo-campus.it

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

protein–protein interaction network influence their role in cancer. NCG can be used to retrieve information on specific cancer genes, as well as to identify groups of cancer genes with identical properties, thus providing a flexible tool for investigating the complex landscape of cancer genetic determinants. In this paper, along with a general description of the features of NCG, we also provide a specific example of how NCG can be used by reporting the properties of *PTEN*, a tumor suppressor gene coding for a phosphatidylinositol phosphatase that is impaired in several cancer types.

MATERIALS AND METHODS

Dataset of human cancer genes

We define cancer genes as a collection of 736 genes that are mutated in different cancer types and derive from two different data sources. A total of 375 genes come from the Cancer Gene Census (CGC-genes, December 2008), a manually curated list of genes with at least two independent reports of mutations in primary tumors (5). The census provides information on the tumor type, as well as on the genetic effect of the mutation, i.e. whether the mutation is dominant or recessive (Figure 1A and B). The remaining 396 genes derive from high-throughput mutational screenings performed in glioblastoma (7), breast and colorectal (6), pancreatic (8) cancers (CAN-genes, Figure 1C) and lung adenocarcinoma (TSP-genes) (9). CAN- and TSP-genes result from the effort of massively sequencing the cancer gene repertoire (26), and provide the first unbiased mutational screenings in different cancer types. The lists of literature-curated and high-throughput derived cancer genes show poor overlap (Figure 1D), confirming the cancer-specificity of the mutational landscape (27). We gather the protein sequences associated to the 736 cancer genes from the RefSeq database [March 2009, (28)]. For eight genes no RefSeq is available and Ensembl protein sequence (29) is used instead.

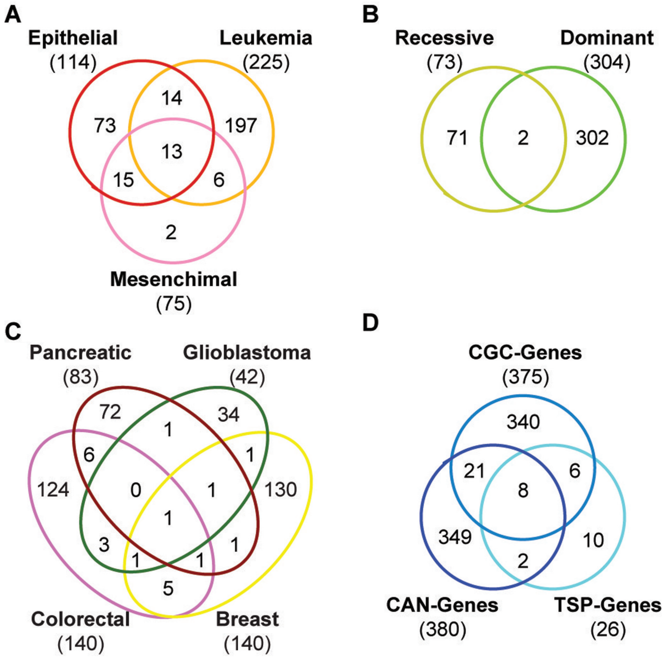


Figure 1. Cancer genes collected in NCG. Venn diagrams of the different lists of cancer genes stored in NCG. The Cancer Gene Census provides information on the cancer type (**A**) and on the phenotypic effect of the mutation (**B**). The CAN-genes reported so far refer to four cancer types (**C**). The overlap among the different data sources used in this study is overall very poor (**D**).

Gene duplicability

We define gene duplicability as in Rambaldi *et al*. (11). In brief, we first align the protein sequences of all human genes to the human genome reference assembly (hg18), using BLAT (30). We then retrieve the best hit of each gene, defined as the locus on the genome with the highest score in terms of coverage. By default, all genes with additional genomic matches that cover at least 60% of the query length are considered duplicable, while genes with no additional hits above this threshold are considered singleton (11). In addition to the results at the default threshold of 60%, we also provide the possibility of inspecting additional hits of the same gene covering higher or lower percentage of the original protein length. For each duplicated locus, we refer to the genome annotation provided by the UCSC Table Browser (31) to assess whether it corresponds to a known gene or instead to non-genic region.

Orthology assignment and evolutionary appearance

We derive the orthology relationships from the eggNOG database (32). Based on these relationships, we assign the evolutionary appearance of each cancer gene, defined as the deepest branch of the tree of life where an ortholog for that gene can be found. Overall, we divide the tree of life into seven main branches: Last Common Ancestor (LCA), which identifies the ancestral cellular organism, Eukaryotes, Opisthokonts, Metazoans, Vertebrates, Mammals and Primates. For example, a human gene whose orthologs are traceable in prokaryotes is considered to have appeared in the LCA, while a human gene with orthologs only in fungi and metazoans, but not in plants, is assumed to be born with Opisthokonts.

Depending on the number of paralogs of a given cancer gene at each branch, we also derive the corresponding orthology ratio, defined as the number of co-orthologs of that human gene in a given lineage. This ratio provides a useful indication of the number of intra-lineage duplications that the gene underwent during evolution. Orthology ratio can be 1 to 1 when no duplications occurred; 1 to *N*, indicating one-to-many relationship; *N* to 1, corresponding to many-to-one relationship; *N* to *N*, when multiple duplications occurred during evolution.

Protein interaction network

In order to gather the most complete representation of the human protein–protein interaction network, we integrate information from five resources: the Human Protein Reference Database (HPRD) (33), BioGRID (34), IntAct (35), the Molecular INTeraction database (MINT) (36)

Table 1. Integration of protein–protein interaction data

Database	Version	Proteins	Interactions	Independent reports
HPRD (33)	1 September 2007	8697	34 938	17 770
BioGRID (34)	1 February 2009	7163	23 588	8815
IntAct (35)	23 January 2009	7066	22 119	1374
MINT (36)	5 February 2009	5151	12 653	1210
DIP (37)	26 January 2009	1108	1326	739
NCG	21 June 2009	11 988	68 498	19 886

Data from five different sources are integrated in NCG. To derive a non-redundant version of the network, proteins are counted as number of non-redundant Entrez IDs. The number of interactions refers to non-redundant primary interactions; the independent reports refer to the number of published papers that define the interactions.

and the Database of Interacting Proteins (DIP) (37). We only consider primary data on interactions between human proteins, i.e. putative interactions inferred from orthology relationships are discarded. The resulting non-redundant network is composed of 68 498 interactions among 11 988 proteins, derived from 19 886 independent literature reports (Table 1). Overall, we find 4621 human proteins that interact with cancer proteins. To provide a complete view of the network of cancer proteins, NCG also allows retrieving information on the systems-level properties for all these primary interactors.

Given the poor overlap between the five data sets (Table 1), their integration allows a more complete coverage of the real interactions for each human protein. For example, the protein TP53 has a total of 408 interactions in NCG, 237 of which derive from HPRD, 214 from BioGRID, 159 from IntAct, 122 from MINT and 38 from DIP. The primary interaction network for each cancer protein is visualized using Medusa (38) and all interactors are provided with information on their duplicability, orthology, evolutionary appearance and possible involvement in cancer.

Database description

NCG is divided into four sections: (i) the gene summary table, which allows the conversion between different gene and protein identifiers, using the Entrez ID as primary key; (ii) the duplicability table, which includes all results of the BLAT alignments on the human genome; (iii) the orthology table, which stores the orthology relationships; and (iv) the network table, which includes the network properties for each protein. The data collected in NCG are stored in a MySQL database. The web interface to interrogate the database is built in Perl.

RESULTS AND DISCUSSION

Information retrieval

NCG allows retrieving information on cancer genes in three ways: (i) by using different types of identifiers, such as gene symbols, Entrez IDs (39), RefSeq (40) or Ensembl IDs (29), for specific genes or groups of genes of interest; (ii) by selecting precompiled lists of cancer genes; and (iii) by combining different criteria to analyze genes with similar duplicability, orthology and network properties. The primary output of the query is a summary table that provides links to several external databases, such as Entrez (www.ncbi.nlm.nih.gov/Entrez/), HPRD (http://www.hprd.org/), OMIM (http://www.ncbi.nlm.nih.gov/omim/) (41), RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/) and Ensembl (http://www.ensembl.org/), as well as to detailed reports on duplicability, orthology and network properties.

Duplicability of cancer genes

In accordance with our previous report (11), at 60% coverage we find 104 duplicable cancer genes (14.1% of the total), which are associated with 336 duplicated loci. According to the available genome annotation, 44% of these additional hits correspond to known genes, 15% to more than one gene and 41% to non-genic regions. Only 22% of duplicable cancer genes duplicate in loci with no evidence of transcription, indicating that, although our measure of duplicability is based on direct genome comparison, it mostly detects transcribed paralogs.

In the case of the tumor suppressor gene *PTEN*, we find an almost identical duplicate (97% coverage, 98% identity) corresponding to *PTENP1* (Figure 2A). While the activity of *PTEN* as repressor of the AKT pathway is well documented (42,43), *PTENP1* is known to transcribe a processed pseudogene (44,45) but the involvement in cancer has never been reported. At 10% coverage, an additional hit is found, which involves the last 50 amino acids of PTEN and matches to the intronic region between exons 3 and 4 of *ANKFN1* (ankyrin-repeat and fibronectin type III domain containing 1, Figure 2A).

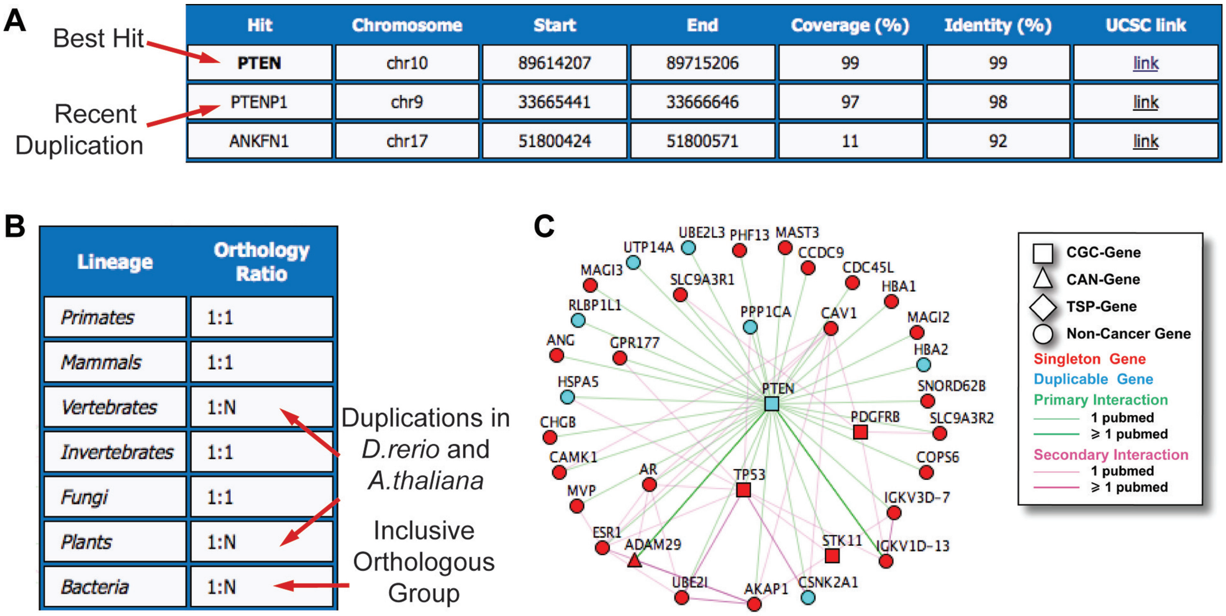
Orthologs and evolutionary appearance of cancer genes

NCG collects orthology information for 723 out of 736 cancer genes (98.2%), since 13 genes are not present in the eggNOG database. We find that 61% of cancer genes originated very early in evolution, because orthologs can be traced back either to LCA or to early Eukaryotes. As few as 2.5% of cancer genes appeared with Opisthokonts, 17.7% with Metazoans, 15% with Vertebrates and only the remaining 3.8% with Mammals and Primates. These results are consistent with previous reports, which assess that disease genes are overall depleted in recent genes (46).

As expected for an enzyme-coding gene, orthologs of *PTEN* are detectable in all branches of the tree of life, including prokaryotes, where they belong to the inclusive orthologous group of the tyrosine phosphatases. With the exception of *Danio rerio* and *Arabidopsis thaliana*, whose genomes underwent whole-genome duplications (47,48), orthologs of *PTEN* maintain a strict 1:1 relationship in all eukaryotic branches (Figure 2B). This suggests an early differentiation of *PTEN* and the maintenance of a strict singleton status during eukaryote evolution.

Network properties of cancer proteins

For each of the 579 cancer proteins with available network information (78.7% of the total), we calculate the degree,



Hit	Chromosome	Start	End	Coverage (%)	Identity (%)	UCSC link
PTEN	chr10	89614207	89715206	99	99	link
PTENP1	chr9	33665441	33666646	97	98	link
ANKFN1	chr17	51800424	51800571	11	92	link

B
Duplications in D.rerio and A.thaliana
Inclusive Orthologous Group

Lineage	Orthology Ratio
Primates	1:1
Mammals	1:1
Vertebrates	1:N
Invertebrates	1:1
Fungi	1:1
Plants	1:N
Bacteria	1:N

Figure 2. Duplicability, orthology and network properties of the tumor suppressor gene *PTEN*. (**A**) Using the PTEN protein sequence as a query, three hits are found on the human genome. The best hit corresponds to genomic locus of *PTEN*, while the two additional hits account for a recent duplication transcribing for the processed pseudogene PTENP1, and to a short region of identity lying in the intron of *ANKFN1*, respectively. (**B**) The orthology ratio reflects the co-orthology relationships of human PTEN at different branching points of the tree of life. The only inparalogs of PTEN in eukaryotes are found in *A. thaliana* and *D. rerio*, indicating that this gene maintained a strict singleton status during eukaryotic evolution. (**C**) PTEN interacts with 35 other human proteins, four of which are cancer proteins and 22 are hubs. This makes PTEN a central node of the human protein-protein interaction network.

i.e. the number of interactions, the clustering coefficient, i.e. the number of interactions between primary interactors, and the betweenness, i.e. the number of shortest paths crossing the protein. These parameters return a measure of connectivity, interconnectivity and centrality, thus providing a glance of the protein topology in the network. On the basis of the network degree, we discriminate between hubs and non-hubs, where the former are defined as the top 5% most connected proteins in the network. Likewise, we identify the central nodes of the network, defined as the proteins with top 5% values of betweenness.

Overall, we find 619 human hubs, 78 of which are cancer proteins (13.4% of the total set). This result is comparable to previous reports and confirms that cancer proteins are enriched in hubs when compared to the rest of human proteins (11,12). We also observe that cancer proteins have higher betweenness than the rest of human proteins (*P*-value <2.2*e*-16, Wilcoxon test), confirming that they occupy a central position in the network.

The PTEN has overall 35 interactors, 21 of which are hubs. This, together with a high betweenness value, makes PTEN a central node that acts as a bypass between several hubs inside the human protein–protein interaction network (Figure 2C). PTEN interacts with TP53 through phosphatase-dependent and -independent mechanisms (49); it is involved in the phosphorylation of ADAM29 (50); it attenuates the activity of the tyrosine kinase receptor PDGFRB (51); and, finally, it is phosphorylated by the serine/threonine kinase STK11 (52). This confirms the tendency of cancer proteins to interact with other cancer proteins, indicating that different components of the key biological processes can contribute to tumorigenesis (11).

FUTURE PROSPECTIVE

In the coming years, we will assist to a continuous delivery of data from the Cancer Genome Project as well as from other large-scale mutational screenings of cancer genes. This massive quantity of information will require *ad hoc* tools for data organization and mining.

NCG represents a first attempt in the direction of a systematic analysis of cancer genes, and it will be constantly updated and expanded with the delivery of new data.

ACKNOWLEDGEMENTS

We thank all members of the Ciccarelli lab for testing NCG and providing useful suggestion to improve it.

FUNDING

Associazione Italiana Ricerca sul Cancro; Fondazione Cariplo (to F.D.C.). Funding for open access charge: Associazione Italiana Ricerca sul Cancro Fondazione Cariplo.

Conflict of interest statement. None declared.

REFERENCES

1. Vogelstein,B. and Kinzler,K.W. (2004) Cancer genes and the pathways they control. *Nat. Med*, **10**, 789–799.
2. Greenman,C., Stephens,P., Smith,R., Dalgliesh,G.L., Hunter,C., Bignell,G., Davies,H., Teague,J., Butler,A., Stevens,C. *et al.* (2007) Patterns of somatic mutation in human cancer genomes. *Nature*, **446**, 153–158.
3. Hanahan,D. and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell*, **100**, 57–70.
4. Stratton,M.R., Campbell,P.J. and Futreal,P.A. (2009) The cancer genome. *Nature*, **458**, 719–724.
5. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
6. Wood,L.D., Parsons,D.W., Jones,S., Lin,J., Sjoblom,T., Leary,R.J., Shen,D., Boca,S.M., Barber,T., Ptak,J. *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science*, **318**, 1108–1113.
7. Parsons,D.W., Jones,S., Zhang,X., Lin,J.C., Leary,R.J., Angenendt,P., Mankoo,P., Carter,H., Siu,I.M., Gallia,G.L. *et al.* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science*, **321**, 1807–1812.
8. Jones,S., Zhang,X., Parsons,D.W., Lin,J.C., Leary,R.J., Angenendt,P., Mankoo,P., Carter,H., Kamiyama,H., Jimeno,A. *et al.* (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, **321**, 1801–1806.
9. Ding,L., Getz,G., Wheeler,D.A., Mardis,E.R., McLellan,M.D., Cibulskis,K., Sougnez,C., Greulich,H., Muzny,D.M., Morgan,M.B. *et al.* (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, **455**, 1069–1075.
10. Kitano,H. (2002) Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Curr. Genet.*, **41**, 1–10.
11. Rambaldi,D., Giorgi,F.M., Capuani,F., Ciliberto,A. and Ciccarelli,F.D. (2008) Low duplicability and network fragility of cancer genes. *Trends Genet.*, **24**, 427–430.
12. Jonsson,P.F. and Bates,P.A. (2006) Global topological features of cancer proteins in the human interactome. *Bioinformatics*, **22**, 2291–2297.
13. Hernandez,P., Huerta-Cepas,J., Montaner,D., Al-Shahrour,F., Valls,J., Gomez,L., Capella,G., Dopazo,J. and Pujana,M.A. (2007) Evidence for systems-level molecular mechanisms of tumorigenesis. *BMC Genomics*, **8**, 185.
14. Forbes,S.A., Bhamra,G., Bamford,S., Dawson,E., Kok,C., Clements,J., Menzies,A., Teague,J.W., Futreal,P.A. and Stratton,M.R. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.*, Chapter 10, Unit 10 11.
15. Higgins,M.E., Claremont,M., Major,J.E., Sander,C. and Lash,A.E. (2007) CancerGenes: a gene selection resource for cancer genome projects. *Nucleic Acids Res.*, **35**, D721–D726.
16. Huret,J.L., Dessen,P. and Bernheim,A. (2003) Atlas of genetics and cytogenetics in oncology and haematology, year 2003. *Nucleic Acids Res.*, **31**, 272–274.
17. Kato,K., Yamashita,R., Matoba,R., Monden,M., Noguchi,S., Takagi,T. and Nakai,K. (2005) Cancer gene expression database (CGED): a database for gene expression profiling with accompanying clinical information of human cancer tissues. *Nucleic Acids Res.*, **33**, D533–D536.
18. Elfilali,A., Lair,S., Verbeke,C., La Rosa,P., Radvanyi,F. and Barillot,E. (2006) ITTACA: a new database for integrated tumor transcriptome array and clinical data analysis. *Nucleic Acids Res.*, **34**, D613–D616.
19. Almeida,L.G., Sakabe,N.J., deOliveira,A.R., Silva,M.C., Mundstein,A.S., Cohen,T., Chen,Y.T., Chua,R., Gurung,S., Gnjatic,S. *et al.* (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res.*, **37**, D816–D819.
20. Kaur,M., Radovanovic,A., Essack,M., Schaefer,U., Maqungo,M., Kibler,T., Schmeier,S., Christoffels,A., Narasimhan,K., Choolani,M. *et al.* (2009) Database for exploration of functional context of genes implicated in ovarian cancer. *Nucleic Acids Res.*, **37**, D820–D823.
21. Sedlacek,Z., Kodet,R., Poustka. and Goetz,P. (1998) A database of germline p53 mutations in cancer-prone families. *Nucleic Acids Res.*, **26**, 214–215.
22. Cariello,N.F., Douglas,G.R., Gorelick,N.J., Hart,D.W., Wilson,J.D. and Soussi,T. (1998) Databases and software for the analysis of mutations in the human p53 gene, human hprt gene and both the lacI and lacZ gene in transgenic rodents. *Nucleic Acids Res.*, **26**, 198–199.
23. Knutsen,T., Gobu,V., Knaus,R., Padilla-Nash,H., Augustus,M., Strausberg,R.L., Kirsch,I.R., Sirotkin,K. and Ried,T. (2005) The interactive online SKY/M-FISH & CGH database and the Entrez cancer chromosomes search database: linkage of chromosomal aberrations with the genome sequence. *Genes Chromosomes Cancer*, **44**, 52–64.
24. He,X., Chang,S., Zhang,J., Zhao,Q., Xiang,H., Kusonmano,K., Yang,L., Sun,Z.S., Yang,H. and Wang,J. (2008) MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res.*, **36**, D836–D841.
25. Jensen,L.J., Kuhn,M., Stark,M., Chaffron,S., Creevey,C., Muller,J., Doerks,T., Julien,P., Roth,A., Simonovic,M. *et al.* (2009) STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.
26. Collins,F.S. and Barker,A.D. (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci. Am.*, **296**, 50–57.
27. Velculescu,V.E. (2008) Defining the blueprint of the cancer genome. *Carcinogenesis*, **29**, 1087–1091.
28. Pruitt,K.D., Tatusova,T., Klimke,W. and Maglott,D.R. (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.*, **37**, D32–D36.
29. Hubbard,T.J., Aken,B.L., Ayling,S., Ballester,B., Beal,K., Bragin,E., Brent,S., Chen,Y., Clapham,P., Clarke,L. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
30. Kent,W.J. (2002) BLAT–the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
31. Karolchik,D., Hinrichs,A.S., Furey,T.S., Roskin,K.M., Sugnet,C.W., Haussler,D. and Kent,W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.
32. Jensen,L.J., Julien,P., Kuhn,M., von Mering,C., Muller,J., Doerks,T. and Bork,P. (2008) eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.*, **36**, D250–D254.
33. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
34. Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bahler,J., Wood,V. *et al.* (2008) The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
35. Kerrien,S., Alam-Faruque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R. *et al.* (2007) IntAct–open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.
36. Cesareni,G., Chatr-aryamontri,A., Licata,L. and Ceol,A. (2008) Searching the MINT database for protein interaction information. *Curr. Protoc. Bioinform.*, Chapter 8, Unit 8 5.
37. Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
38. Hooper,S.D. and Bork,P. (2005) Medusa: a simple tool for interaction graph analysis. *Bioinformatics*, **21**, 4432–4433.
39. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
40. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
41. Amberger,J., Bocchini,C.A., Scott,A.F. and Hamosh,A. (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res.*, **37**, D793–D796.

42. Li,J., Yen,C., Liaw,D., Podsypanina,K., Bose,S., Wang,S.I., Puc,J., Miliaresis,C., Rodgers,L., McCombie,R. *et al*. (1997) PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*, **275**, 1943–1947.
43. Keniry,M. and Parsons,R. (2008) The role of PTEN signaling perturbations in cancer and in targeted therapy. *Oncogene*, **27**, 5477–5485.
44. Dahia,P.L., FitzGerald,M.G., Zhang,X., Marsh,D.J., Zheng,Z., Pietsch,T., von Deimling,A., Haluska,F.G., Haber,D.A. and Eng,C. (1998) A highly conserved processed PTEN pseudogene is located on chromosome band 9p21. *Oncogene*, **16**, 2403–2406.
45. Kim,S.K., Su,L.K., Oh,Y., Kemp,B.L., Hong,W.K. and Mao,L. (1998) Alterations of PTEN/MMAC1, a candidate tumor suppressor gene, and its homologue, PTH2, in small cell lung cancer cell lines. *Oncogene*, **16**, 89–93.
46. Domazet-Loso,T. and Tautz,D. (2008) An ancient evolutionary origin of genes associated with human genetic diseases. *Mol. Biol. Evol*., **25**, 2699–2707.
47. Postlethwait,J.H., Woods,I.G., Ngo-Hazelett,P., Yan,Y.L., Kelly,P.D., Chu,F., Huang,H., Hill-Force,A. and Talbot,W.S. (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res*., **10**, 1890–1902.
48. Ermolaeva,M.D., Wu,M., Eisen,J.A. and Salzberg,S.L. (2003) The age of the Arabidopsis thaliana genome duplication. *Plant Mol. Biol*., **51**, 859–866.
49. Freeman,D.J., Li,A.G., Wei,G., Li,H.H., Kertesz,N., Lesche,R., Whale,A.D., Martinez-Diaz,H., Rozengurt,N., Cardiff,R.D. *et al*. (2003) PTEN tumor suppressor regulates p53 protein levels and activity through phosphatase-dependent and -independent mechanisms. *Cancer Cell*, **3**, 117–130.
50. Miller,S.J., Lou,D.Y., Seldin,D.C., Lane,W.S. and Neel,B.G. (2002) Direct identification of PTEN phosphorylation sites. *FEBS Lett*., **528**, 145–153.
51. Takahashi,Y., Morales,F.C., Kreimann,E.L. and Georgescu,M.M. (2006) PTEN tumor suppressor associates with NHERF proteins to attenuate PDGF receptor signaling. *EMBO J*., **25**, 910–920.
52. Mehenni,H., Lin-Marq,N., Buchet-Poyau,K., Reymond,A., Collart,M.A., Picard,D. and Antonarakis,S.E. (2005) LKB1 interacts with and phosphorylates PTEN: a functional link between two proteins involved in cancer predisposing syndromes. *Hum. Mol. Genet*., **14**, 2209–2219.