

Review

Human variation databases

Jan Küntzer^{1,*}, Daniela Eggle², Stefan Klostermann¹ and Helmut Burtscher³

¹Pharma Research and Early Development, pRED Informatics, Roche Diagnostics GmbH, Penzberg, ²Pharma Research and Early Development, Translational Research Sciences, Roche Diagnostics GmbH, Penzberg and ³Pharma Research and Early Development, Discovery Oncology, Roche Diagnostics GmbH, Penzberg, Germany

*Corresponding author: Tel: +49 8856 60 5971; Fax: +49 8856 6079 5971; E-mail: jan.kuentzer@roche.com

Submitted 13 April 2010; Revised 11 June 2010; Accepted 29 June 2010

More than 100 000 human genetic variations have been described in various genes that are associated with a wide variety of diseases. Such data provides invaluable information for both clinical medicine and basic science. A number of locus-specific databases have been developed to exploit this huge amount of data. However, the scope, format and content of these databases differ strongly and as no standard for variation databases has yet been adopted, the way data is presented varies enormously. This review aims to give an overview of current resources for human variation data in public and commercial resources.

Background

Over the recent years the cloning of genes involved in complex diseases such as cancer as well as the development of new high throughput techniques like single nucleotide polymorphism (SNP) arrays has made enormous progress. This resulted in more than 100 000 human genetic variations which have been described in various genes associated with a wide variety of diseases (1–3). Somatic variations in cancer are used in clinical studies and molecular pathology to characterize tumor types, to improve the best suited treatment choice, and to predict response to treatment. Thus, mutation analysis can play an important role in drug discovery and drug development. Identification of genetic variants will yield new drug targets and biomarkers.

Cancer, as a disease of genome alterations, arises through the sporadic acquisition of multiple somatic variations (4). However, not all mutations contribute equally to the cancer type in which they are found. The proportion of mutations causally implicated in cancer is still unknown especially due to the high number of variations between different tumors (5–9) Although the number of unique variations for each cancer genome can be very high (10,11), only a few somatic variations will be critical for the development of the tumor. These causative variations, the so-called ‘drivers’, are emerging because of selective

pressure during tumorigenesis, whereas many mutations are only incidental or caused by genome instabilities, so-called ‘passengers’ (12). The differentiation of disease causing driver mutations from the passenger variations is a challenge for mutation analysis (13).

Usefulness of mutation analysis

Analysis of mutations is useful in many ways: the study of cancer-prone DNA repair diseases (Xeroderma pigmentosum, Ataxia telangiectasia, Fanconi’s anemia, Bloom’s syndrome and others) has given valuable insights in the type and function of genes responsible for maintaining DNA integrity (14–18). Mutation analysis can help to predict the risk for developing certain types of cancer, *BRCA1* and *BRCA2* (increased breast cancer risk) (19) and *APC* (increased risk for colon cancer) (20) being among the best known so far.

Mutations can also influence the response of patients to cancer drugs, e.g. the *KRAS* (21,22) or *BRAF* (23,24) mutations. The presence of certain mutations can also influence progression free or overall survival rates of patients (22,25).

Germline versus somatic mutations

In general, mutations can be grouped into two different categories: germline and somatic. *Germline mutations* are

variations found in all cells of an organism including germ line cells. They play an important role in evolution by giving every human its genetic individuality (see SNPs) but also give a rise to hereditary diseases like sickle-cell anemia or phenylketonuria. Germline mutations can also lead to increased risk for developing cancer, like *BRCA1* and *BRCA2* gene mutations which are associated with an increased risk for breast and ovarian cancer (26–28). Other examples of familial cancer syndromes include von Hippel–Lindau syndrome (caused by mutations in *VHL*) (29), Peutz–Jeghers syndrome (caused by mutations in *LKB1*) (30) and Li–Fraumeni syndrome (caused by mutations in *TP53*) (31).

Detection of germline mutations with current technologies is state of the art but time-consuming. Usually a large amount of genetic material of good quality can be extracted from blood cells. However, in addition to the mutation detection, the differentiation of disease causing and neutral germline mutations having no effect on the phenotype is an important but non-trivial task. Currently, no generally applicable solution for this problem exists and this question often remains unsolved.

Somatic mutations are not inherited but acquired during lifetime in somatic cells of an organism and might cause tissue specific tumors. An important problem with somatic mutations is the difficulty of their detection. Tumor samples can be very heterogeneous and are very often ‘contaminated’ with normal cells, such as stromal cells. However, since somatic mutations are identified through a comparison of a tumor sample with a normal sample of the same organism the identification of the mutation is unambiguous. Also for somatic mutations the differentiation between drivers and passengers is an important but still unsolved problem. In contrast to germline mutations however, all somatic mutations are tumor associated. Therefore, all non-silent somatic mutations are potential candidates for biomarker development.

Mutation types

Genome alterations are typically classified by the mutation type. The different databases characterize all variations first by the effect on the nucleotide sequence: *deletions*, *insertions* and *single nucleotide variations*. Mutations occurring in the coding region of a gene can also be classified by their effect on the amino acid sequence. A variation of the coding sequence without any change of the amino acid sequence of the protein is called *silent* mutation. Single nucleotide mutations causing the substitution of a different amino acid are called *missense* mutation. A *frameshift* mutation is an insertion or deletion in the coding sequence which changes the reading frame resulting in a different translation of the subsequent sequence.

Nonsense mutations generate a premature stop codon and often a non-functional truncated protein product.

SNPs versus germline mutations

Single nucleotide germline mutations and SNP are often used as synonyms, since both describe variations of single nucleotides, which are inherited and not tumor-associated per se. However, concerning the databases presented here these synonyms are used in two different meanings: SNPs as presented in public databases like dbSNP (32,33) or HapMap (34) are germline variations for which at most population frequencies are known. In literature it is usually assumed that the variation should be found in more than 1% of the population in order to be called a SNP. Such information is very useful for biomarker development since it describes the prevalence of the mutation in different populations. However, it is normally not possible to get additional information (like gender, age, or disease status) on the individuals having the SNP, only the population a person belongs to is given. Since it is not known if the information comes from a tumor or normal sample, a correlation between diseases and SNPs cannot be calculated.

In contrast, germline mutations presented in cancer or disease mutation databases like ‘The Cancer Genome Atlas (TCGA)’ (35) are usually connected with additional sample information like patient gender, age, histology or tissue. Germline mutations are found in the normal as well as the tumor sample. Hence, the sample information allows for further analyses of associations between germline mutations and diseases.

Standardization efforts

A standard problem occurring in every field where huge amounts of data are generated is standardization. Without standardization the task to identify and integrate the data is very complicated, laborious, error-prone and time-consuming. Although databases may have different scope and aims it is important to standardize content and annotation. The Human Genome Variation Society (HGVS) has proposed a recommendation for the nomenclature of genetic variations and content of mutation databases and scientific publications (36). This naming of mutations has now become widely accepted. Some journals (e.g. Human Mutation) already accept only publications with mutation notation following the HGVS proposed recommendations. If more publishers should follow this trend it would have a very positive effect on the usability of mutation databases including an increase of the quality and amount of their content.

HGVS and members have published number of recommendations e.g. for the collection of somatic mutations, sharing data, etc. There are also projects at European

Bioinformatics Institute (EBI) and National Center for Biotechnology Information (NCBI) to develop reference sequences, locus reference genomics (LRGs) (<http://www.lrg-sequence.org>) and RefSeqGenes (<http://www.ncbi.nlm.nih.gov/projects/RefSeq/RSG/>), respectively. In addition, the Gen2Phen (<http://www.gen2phen.org>) project works on data models and standards for a number of aspects related to variation data description, storage and integration in databases.

Except for the already widely accepted naming recommendations of mutations by the HGVS, a promising standardization effort for integrating all cancer genome data is still missing.

Structure and accessibility

Historically, mutations and variations in human have been reported only in the published literature. Mutation descriptions were often not precise, no standard notation existed, and the sequence of the reference gene under study was almost never indicated. To this end a sophisticated mutation analysis was mostly unfeasible. However, with the explosion of large-scale cancer genome sequencing (35,37–40) more and more information on genetic variations has been captured over the last years in publicly available databases that can be used by clinicians or scientists as a research tool. These databases are widespread and their scope, format and content can be very different. Current data related to somatic mutations is mostly buried in journals or scattered between several locus-specific databases (LSDBs) and general databases that have no or very limited connections between them.

Only a few large public resources exist that comprehensively compile data on somatic gene alterations in cancer: International Agency for Research on Cancer (IARC) TP53 Database (41), Catalog of Somatic Mutations in Cancer (COSMIC) (42), TCGA (35), Roche Cancer Genome Database (RCGDB) and Human Gene Mutation Database (HGMD[®]) (43).

The LSDBs often originate as loosely organized compilation of data. Since no standard system similar to the HGVS recommendation for mutation notation has yet been established, the presentation of the data in LSDBs varies enormously. The data is mostly presented in flat files, plain text databases, or Microsoft Excel spreadsheets making it easy to collect and store the information, but nearly impossible to search or retrieve specific data. More ambitious databases use open source database management software (DBMS)—like MySQL or PostgreSQL—whereas only a minority of curators use specialized software such as the UMD (44), the Mutation Storage and Retrieval Program (MuStar) (45), or the Leiden Open Source Variation Database (LOVD) (46). The use of such relational DBMSs

allows to specify complex queries and specific analyses of customized subsets of the database.

Cancer variation databases

Currently, the best-known publicly available primary database on somatic mutations in human cancer is the 'COSMIC'(42) hosted at the Wellcome Trust Sanger institute in Cambridge. The data is gathered from scientific publications and genome-wide screens from the Cancer Genome Project (CGP) at the same institute. The project has been continuously updated and improved for over 9 years and currently contains more than 108 773 mutations in >13 500 different genes observed in over 449 676 different tumor samples. The curation process in COSMIC is largely manual resulting in a very high quality of the data. For each mutation all details on the sample like patient age, gender, histology and tissue are available. COSMIC uses its own internal classification system to provide tissue and histology consistency within the database and to reduce redundancy. All tissue and histology information from scientific publications is translated using this classification system. In addition, for each study the project offers the information which genes were actually screened, since published studies often focus on mutation hot spots, for example *KRAS* (47), *BRAF* (48) or *TP53* (49). This information enables frequency data to be calculated for mutations in various genes and different cancer types. COSMIC offers also somatic mutations found in cell-lines including the NCI-60 (50). The website of COSMIC has a clear structure and is easy to use. The interface allows to browse by gene, or search by phenotype. Summary information on mutation counts and frequencies are presented graphically for a better understanding. In addition, all information can be downloaded as txt files, or as an Oracle dump file.

Another large mutation data source is 'TCGA' (35), a project at the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI). The main goal of TCGA is to understand the molecular basis of cancer through the application of genome analysis techniques, including large-scale genome sequencing and SNP analysis. For each patient a whole genome analysis of a normal, a tumor and control samples (a second normal and tumor sample as control) is performed enabling researchers to distinguish between somatic and germline mutations. All mutations found are publicly available in a special Mutation Analysis file Format (MAF) and can be downloaded via the TCGA Data Portal. This portal contains all TCGA data concerning to clinical information associated with tumors and human subjects, genomic characterization, and high-throughput sequencing analysis of the tumor genomes. However, no advanced search interface or graphical visualization of the mutation data is available. In the starting phase the project focused on only two cancer

types: brain cancer (glioblastoma multiforme) and ovarian serous adenocarcinoma. After the pilot phase, which was completed in 2009, TCGA matured to a full project and is now dealing with more than 20 types of cancer.

Another concept focusing on the integration of heterogeneous mutation data sources is pursued by the RCGDB (51), developed at Roche Pharma Research. The freely available warehouse system integrates somatic and germline mutations gathered by manual curation from scientific publications and public cancer mutation databases (COSMIC, TCGA, etc.). In addition, these mutations are enriched by SNP data from the HapMap (34) project. Updates are provided on a regular basis depending on the update frequency of the external data sources (approximately every 3 months). Access to the RCGDB is offered via a publicly available web interface. A major aspect in designing the user interface was that users should be able to search and view mutations in an intuitive and straight-forward manner, without having to understand the architectural details of the warehouse system. Therefore, the database offers a Google-like web interface to search for cancer genome information on a single gene, sample or cell-line, and on multiple genes, samples or cell-lines. As a special feature the search is supported by an auto-suggestion functionality allowing to search by NCBI GeneIDs, names, or synonyms.

The HGMD® (43) at the Institute of Medical Genetics in Cardiff is a commercial mutation database providing information on somatic and germline mutations. Furthermore, the database offers a less up-to-date public version which is freely available only to registered users from academic institutions or non-profit organizations, respectively. The data is gathered from scientific publications and from publicly available LSDBs. The project claims to include all mutations causing or associated with human inherited disease, plus disease-associated/functional polymorphisms reported in the literature. Currently, HGMD provides information on 96 631 mutations in 3611 genes under the professional license and 69 660 mutation in 2572 genes in the public version of the database. The website of HGMD allows to search by gene, publication or mutation id and presents the results in a table view. A downloadable version of HGMD is only available under the professional license.

In addition to multi-gene LSDBs, various single-gene LSDBs are publicly available. The largest and best-known single-gene LSDB is the *TP53 mutation database* from the IARC (41), with all TP53 gene variations identified in human populations and tumor samples since 1989. The database contains information on somatic as well as germline mutations of TP53 in patient samples, human cell-lines, and mouse models. This data is compiled from the peer-reviewed literature and from generalist databases. The website offers different sophisticated interfaces for searching and mining the database by multiple criteria.

Furthermore, all information can be downloaded in tab-delimited txt-files. A large number of other single gene databases exists like the *L1CAM mutation database* from the university of Groningen (52) containing single gene somatic mutations. Most of these LSDBs are small containing mostly <500 variants.

For a detailed list of cancer mutation databases see Table 1.

Disease variation databases

In addition to the Cancer variation database a large number of publicly available databases focuses on disease specific variations. An overview on such disease variation databases can be found in Table 2.

Prominent disease mutation databases are the public *IDbases* (53) maintained at the Institute of Medical Technology, University of Tampere. The IDbases are LSDBs for immunodeficiency-causing mutations. The project maintains 122 different IDbases containing altogether data for 5359 patients. In addition to gene mutations, IDbases provide information about clinical presentation. All information has been collected from the literature as well as directly from researchers. The databases do not provide any sophisticated search interface and allow to download the data as a txt-file.

Conclusion

All databases presented are good starting points to retrieve human variation data for certain use cases depending on the provided interfaces. However, as soon as a query gets more complicated, an integrative approach will be necessary. Unfortunately, the diversity of current mutation information systems and the underlying data models make it difficult to mine human variation databases in an integrative approach. Currently, researchers typically have to browse and search several databases to obtain the required information. No unified access to all different cancer genome related data sources exists resulting in a need for more efficient integrative systems. With COSMIC, which is currently integrating TCGA and IARC TP53 information, and the RCGDB, which already integrates most of in this review presented data sources, two promising integrative data resources are available. Nevertheless, the standardization and virtual consolidation of existing databases will be one major challenge for future developments. Although these problems have already been discussed in previous publications (54–56), the current situation concerning mutation databases and their heterogeneity is still an acute problem due to the exponential growth of data generated by genome sequencing. This review is meant to provide an overview on the current status of mutation data in public

Table 1. Cancer variation database: a list of available cancer variation databases including web links

| Database | URL | Gene(s) | Mutation type | Remark |
|--|---|--|-------------------|-------------------------------------|
| BLMbase | http://bioinf.uta.fi/BLMbase | BLM | Germline | |
| CASP10base | http://bioinf.uta.fi/CASP10base | CASP10 | Germline | |
| CASP8base | http://bioinf.uta.fi/CASP8base | CASP8 | Germline | |
| Catalog of Somatic Mutations in Cancer (COSMIC) | http://www.sanger.ac.uk/genetics/CGP/cosmic/ | > 13 500 | Somatic | |
| Fanconi Anemia Mutation Database | http://www.rockefeller.edu/fanconi/mutate/ | FANCA, FANCB, FANCC, FANCD1, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, FANCN | | |
| Genetic Alterations in Cancer DB (GAC) | http://www.niehs.nih.gov/research/resources/databases/gac/ | 32 cancer genes | Somatic, germline | |
| HNPCC database | http://www.insight-group.org/ | APC, EPCAM, MUTYH, MSH2, MSH6, MLH1, MLH3, PMS1, PMS2 | | |
| Human Genome Variation and Genotype/Phenotype Database (HGVBbaseG2P) | http://www.hgvbaseg2p.org/index | | Germline | |
| IARC TP53 database | http://www-p53.iarc.fr/ | TP53 | Somatic, germline | |
| International HapMap Project | http://www.hapmap.org/ | | SNP | |
| KinMutBase | http://bioinf.uta.fi/KinMutBase/ | Protein kinases | Germline | |
| LOVD-ATM | http://www.LOVD.nl/ATM | ATM | Germline | Uses Leiden Open Variation Database |
| LOVD-B3GALTL | http://www.LOVD.nl/B3GALTL | B3GALTL | Germline | Uses Leiden Open Variation Database |
| LOVD-BRCA2 | http://www.LOVD.nl/BRCA2 | BRCA2 | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCA | http://www.LOVD.nl/FANCA | FANCA | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCB | http://www.LOVD.nl/FANCB | FANCB | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCC | http://www.LOVD.nl/FANCC | FANCC | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCD2 | http://www.LOVD.nl/FANCD2 | FANCD2 | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCE | http://www.LOVD.nl/FANCE | FANCE | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCF | http://www.LOVD.nl/FANCF | FANCF | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCG | http://www.LOVD.nl/FANCG | FANCG | Germline | Uses Leiden Open Variation Database |
| LOVD-FANCL | http://www.LOVD.nl/FANCL | FANCL | Germline | Uses Leiden Open Variation Database |

(Continued)

Table 1. Continued

| Database | URL | Gene(s) | Mutation type | Remark |
|--|---|-----------|-------------------|--|
| LOVD-MUTYH | http://www.LOVD.nl/MUTYH | MUTYH | Germline | Uses Leiden Open Variation Database |
| LOVD-NOTCH3 | http://www.LOVD.nl/NOTCH3 | NOTCH3 | Germline | Uses Leiden Open Variation Database |
| LOVD-NROB1 | http://www.LOVD.nl/NROB1 | NROB1 | Germline | Uses Leiden Open Variation Database |
| LOVD-OTC | http://www.LOVD.nl/OTC | OTC | Germline | Uses Leiden Open Variation Database |
| LOVD-TSC1 | http://www.LOVD.nl/TSC1 | TSC1 | Germline | Uses Leiden Open Variation Database |
| LOVD-TSC2 | http://www.LOVD.nl/TSC2 | TSC2 | Germline | Uses Leiden Open Variation Database |
| MDL EGFR Mutation Database | http://www.egfr.org/ | EGFR | Somatic, germline | |
| Mismatch Repair Genes Variant Database | http://www.med.mun.ca/mmrvariants/ | | | |
| MSH2,MSH6,MLH1,PMS2 | Germline | | | |
| NCBI dbSNP | http://www.ncbi.nlm.nih.gov/projects/SNP/ | | SNP | |
| Online Mendelian Inheritance in Man (OMIM) | http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim | All genes | Germline | |
| PTCH Mutation Database | http://www.cybergene.se/PTCH/ | PTCH | Somatic, germline | |
| Roche Cancer Genome Database (RCGDB) | http://rcgdb.bioinf.uni-sb.de/MutomeWeb | > 10 000 | Somatic, SNP | |
| The Cancer Genome Atlas (TCGA) | http://cancergenome.nih.gov/ | >400 | Somatic, germline | Login for TCGA Data Portal necessary, more disease to come |
| The Human Gene Mutation Database (HGMD) | http://www.hgmd.cf.ac.uk/ac/index.php | 2572 | Somatic, germline | Commercial license: 3611 genes |
| The TP53 database | http://p53.free.fr/ | TP53 | Somatic, germline | |
| TSH Receptor mutation database | http://www.uni-leipzig.de/innere/tsh/ | TSHR | Somatic, germline | |
| UMD-APC | http://www.umd.be/APC/ | APC | Germline | |
| UMD-BRCA1 | http://www.umd.be/BRCA1/ | BRCA1 | Germline | Restricted access |
| UMD-BRCA2 | http://www.umd.be/BRCA2/ | BRCA2 | Germline | Restricted access |
| UMD-MEN1 | http://www.umd.be/MEN1/ | MEN1 | Germline | |
| UMD-VHL | http://www.umd.be/VHL/ | VHL | Germline | |
| UML-TGFBR2 | http://194.167.35.228/TGFBR2/ | TGFBR2 | Germline | |
| University of Groningen L1CAM Mutation DB | http://www.l1cammutationdatabase.info/ | L1CAM | Somatic | |
| WASbase | http://bioinf.uta.fi/WASbase | WASP | Germline | |
| Werner Syndrome Mutational Database | http://www.pathology.washington.edu/research/werner/database/ | WRN | Germline | |

For each database the type of mutations as well as the genes have been covered.

Table 2. Disease variation databases: a list of available disease variation databases including web links

| Database | URL | Gene(s) | Diseases | Remark |
|---|---|----------|--|--------|
| ADAbase | http://bioinf.uta.fi/ADAbase | ADA | Adenosine deaminase deficiency | |
| AICDAbase | http://bioinf.uta.fi/AICDAbase | AICDA | Non-X-linked hyper-IgM syndrome | |
| AIREbase | http://bioinf.uta.fi/AIREbase | AIRE | Autoimmune polyendocrinopathy with candidiasis and ectodermal dystrophy (APECED) | |
| Albinism Database (CHS) | http://albinismdb.med.umn.edu/chs1mut.html | LYST | Chediak–Higashi Syndrome | |
| Albinism Database (HPS2) | http://albinismdb.med.umn.edu/hps2mut.htm | AP3B1 | Hermansky–Pudlak syndrome 2 | |
| ALPSbase (II) | http://research.nhgri.nih.gov/ALPS/alpsII_mut.shtml | CASP10 | Autoimmune lymphoproliferative syndrome, type II | |
| ALPSbase (Ia) | http://research.nhgri.nih.gov/ALPS/alpsIa_mut.shtml | FAS | Autoimmune lymphoproliferative syndrome, type Ia | |
| AP3B1base | http://bioinf.uta.fi/AP3B1base | AP3B1 | Hermansky–Pudlak syndrome 2 | |
| BIRC4base | http://bioinf.uta.fi/BIRC4base | BIRC4 | X-linked lymphoproliferative syndrome | |
| BLMbase | http://bioinf.uta.fi/BLMbase | BLM | Bloom syndrome | |
| BLNKbase | http://bioinf.uta.fi/BLNKbase | BLNK | BLNK deficiency | |
| BTKbase | http://bioinf.uta.fi/BTKbase | BTK | X-linked agammaglobulinemia (XLA) | |
| C1QAbase | http://bioinf.uta.fi/C1QAbase | C1QA | C1q α polypeptide deficiency | |
| C1QBbase | http://bioinf.uta.fi/C1QBbase | C1QB | C1q β polypeptide deficiency | |
| C1QCbase | http://bioinf.uta.fi/C1QCbase | C1QC | C1q γ -polypeptide deficiency | |
| C1Sbase | http://bioinf.uta.fi/C1Sbase | C1 S | C1 s deficiency | |
| C2base | http://bioinf.uta.fi/C2base | C2 | C2 deficiency | |
| C3base | http://bioinf.uta.fi/C3base | C3 | C3 deficiency | |
| C5base | http://bioinf.uta.fi/C5base | C5 | C5 deficiency | |
| C6base | http://bioinf.uta.fi/C6base | C6 | C6 deficiency | |
| C7base | http://bioinf.uta.fi/C7base | C7 | C7 deficiency | |
| C8Bbase | http://bioinf.uta.fi/C8Bbase | C8B | C8B deficiency | |
| C9base | http://bioinf.uta.fi/C9base | C9 | C9 deficiency | |
| CA2base | http://bioinf.uta.fi/CA2base | CA2 | Osteopetrosis with renal tubular acidosis | |
| CASP10base | http://bioinf.uta.fi/CASP10base | CASP10 | Autoimmune lymphoproliferative syndrome, type II | |
| CASP8base | http://bioinf.uta.fi/CASP8base | CASP8 | Caspase 8 deficiency | |
| Catalogue of Somatic Mutations in Cancer (COSMIC) | http://www.sanger.ac.uk/genetics/CGP/cosmic/ | > 13 500 | multiple tissues and histologies | |
| CD19base | http://bioinf.uta.fi/CD19base | CD19 | CD19 deficiency | |
| CD247base | http://bioinf.uta.fi/CD247base | CD247 | CD3 ζ deficiency | |
| CD3Dbase | http://bioinf.uta.fi/CD3Dbase | CD3D | CD3 δ deficiency | |
| CD3Ebase | http://bioinf.uta.fi/CD3Ebase | CD3 E | CD3 ϵ deficiency | |
| CD3Gbase | http://bioinf.uta.fi/CD3Gbase | CD3 G | CD3 γ deficiency | |
| CD40base | http://bioinf.uta.fi/CD40base | CD40 | CD40 deficiency | |
| CD40Lbase | http://bioinf.uta.fi/CD40Lbase | CD40 L | X-linked Hyper-IgM syndrome (XHIM) | |

(Continued)

Table 2. Continued

| Database | URL | Gene(s) | Diseases | Remark |
|----------------------------------|---|---|---|--------|
| CD55base | http://bioinf.uta.fi/CD55base | CD55 | Decay-accelerating factor (CD55) deficiency | |
| CD59base | http://bioinf.uta.fi/CD59base | CD59 | CD59 deficiency | |
| CD79Abase | http://bioinf.uta.fi/CD79Abase | CD79 A | Ig α deficiency | |
| CD79Bbase | http://bioinf.uta.fi/CD79Bbase | CD79B | Ig β deficiency | |
| CD8Abase | http://bioinf.uta.fi/CD8Abase | CD8 A | CD8 α deficiency | |
| CEBPEbase | http://bioinf.uta.fi/CEBPEbase | CEBPE | Neutrophil-specific granule deficiency | |
| CFDbase | http://bioinf.uta.fi/CFDbase | CFD | Factor D deficiency | |
| CFHbase | http://bioinf.uta.fi/CFHbase | CFH | Factor H deficiency | |
| CFIbase | http://bioinf.uta.fi/CFIbase | CFI | Complement factor I deficiency | |
| CFPbase | http://bioinf.uta.fi/CFPbase | CFP | Properdin deficiency | |
| CIITAbase | http://bioinf.uta.fi/CIITAbase | CIITA | MHCII transactivating protein deficiency | |
| CLCN7base | http://bioinf.uta.fi/CLCN7base | CLCN7 | Autosomal dominant osteopetrosis, type 2 | |
| CTSCbase | http://bioinf.uta.fi/CTSCbase | CTSC | Papillon-Lefevre syndrome | |
| CXCR4base | http://bioinf.uta.fi/CXCR4base | CXCR4 | WHIM syndrome | |
| CYBAbase | http://bioinf.uta.fi/CYBAbase | CYBA | Autosomal recessive p22phox deficiency | |
| CYBBbase | http://bioinf.uta.fi/CYBBbase | CYBB | X-linked chronic granulomatous disease (XCGD) | |
| DCLRE1Cbase | http://bioinf.uta.fi/DCLRE1Cbase | DCLRE1 C | Artemis deficiency | |
| DKC1base | http://bioinf.uta.fi/DKC1base | DKC1 | Hoyeraal-Hreidarsson syndrome | |
| DNMT3Bbase | http://bioinf.uta.fi/DNMT3Bbase | DNMT3B | ICF syndrome | |
| ELA2base | http://bioinf.uta.fi/ELA2base | ELA2 | Cyclic neutropenia; Congenital neutropenia | |
| F12base | http://bioinf.uta.fi/F12base | F12 | Hereditary angioedema type III | |
| Fanconi Anemia Mutation Database | http://www.rockefeller.edu/fanconi/mutate/jumpa.html | FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCL | Fanconi anemia | |
| FASLGbase | http://bioinf.uta.fi/FASLGbase | FASLG | Autoimmune lymphoproliferative syndrome, type 1B (ALPS1B) | |
| FCGR1Abase | http://bioinf.uta.fi/FCGR1Abase | FCGR1 A | CD64 deficiency | |
| FCGR3Abase | http://bioinf.uta.fi/FCGR3Abase | FCGR3 A | Natural killer cell deficiency | |
| FH aHUS Mutation Database | http://www.fh-hus.org/ | CFH | Hemolytic uraemic syndrome (HUS) | |
| FOXN1base | http://bioinf.uta.fi/FOXN1base | FOXN1 | T-cell immunodeficiency, congenital alopecia, and nail dystrophy | |
| FOXP3base | http://bioinf.uta.fi/FOXP3base | FOXP3 | Immunodysregulation, polyendocrinopathy, and enteropathy, X-linked; IPEX | |
| GFI1base | http://bioinf.uta.fi/GFI1base | GFI1 | Severe congenital neutropenia (SCN); Nonimmune chronic idiopathic neutropenia of adults (NI-CINA) | |
| HAEdb | http://hae.enzim.hu/ | SERPING1 | Hereditary angioedema | |
| HAX1base | http://bioinf.uta.fi/HAX1base | HAX1 | Severe congenital neutropenia (Kostmann disease) | |

(Continued)

Table 2. Continued

| Database | URL | Gene(s) | Diseases | Remark |
|-------------|---|----------|--|--------|
| ICOSbase | http://bioinf.uta.fi/ICOSbase | ICOS | ICOS deficiency | |
| IFNGR1base | http://bioinf.uta.fi/IFNGR1base | IFNGR1 | IFN γ 1-receptor deficiency | |
| IFNGR2base | http://bioinf.uta.fi/IFNGR2base | IFNGR2 | IFN γ 2-receptor deficiency | |
| IGHG2base | http://bioinf.uta.fi/IGHG2base | IGHG2 | IgG2 deficiency | |
| IGHMbase | http://bioinf.uta.fi/IGHMbase | IGHM | μ heavy chain deficiency | |
| IGLL1base | http://bioinf.uta.fi/IGLL1base | IGLL1 | λ 5 surrogate light-chain deficiency | |
| IKBKGbase | http://bioinf.uta.fi/IKBKGbase | IKBKG | Nemo deficiency | |
| IL12Bbase | http://bioinf.uta.fi/IL12Bbase | IL12B | Interleukin-12 (IL12) p40 deficiency | |
| IL12RB1base | http://bioinf.uta.fi/IL12RB1base | IL12RB1 | Interleukin-12 receptor β 1 deficiency | |
| IL2RAbase | http://bioinf.uta.fi/IL2RAbase | IL2RA | Interleukin-2 receptor α deficiency | |
| IL2RGbase | http://research.nhgri.nih.gov/scid/ | IL2RG | X-linked SCID | |
| IL7Rbase | http://bioinf.uta.fi/IL7Rbase | IL7R | Interleukin-7 receptor α deficiency | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | LPIN2 | Majeed syndrome | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | MEFV | Familial Mediterranean fever | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | MVK | Hyper IgD Syndrome and periodic fever | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | NLRP3 | Familial cold autoinflammatory syndrome, Muckle-Wells syndrome and chronic infantile neurological cutaneous and articular syndrome | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | NLRP7 | Recurrent Hydatidiform moles and reproductive wastage | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | NOD2 | Blau syndrome, Crohn's disease, early onset sarcoidosis | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | PSTPIP1 | Pyogenic sterile arthritis, pyoderma gangrenosum and acne syndrome | |
| Infervers | http://fmf.igh.cnrs.fr/ISSAID/infervers/ | TNFRSF1A | Tumor necrosis factor receptor-associated periodic syndrome | |
| IRAK4base | http://bioinf.uta.fi/IRAK4base | IRAK4 | IRAK4 deficiency | |
| ITGB2base | http://bioinf.uta.fi/ITGB2base | ITGB2 | Leukocyte adhesion deficiency I (LAD-I) | |
| JAK3base | http://bioinf.uta.fi/JAK3base | JAK3 | Jak3 deficiency | |
| LIG1base | http://bioinf.uta.fi/LIG1base | LIG1 | DNA ligase I deficiency | |
| LIG4base | http://bioinf.uta.fi/LIG4base | LIG4 | LIG4 syndrome | |
| LRRC8Abase | http://bioinf.uta.fi/LRRC8Abase | LRRC8A | Non-Bruton type autosomal dominant agammaglobulinemia | |
| LYSTbase | http://bioinf.uta.fi/LYSTbase | LYST | Chediak-Higashi syndrome | |
| MAPBPIPbase | http://bioinf.uta.fi/MAPBPIPbase | MAPBPIP | Endosomal adaptor protein p14 deficiency | |
| MASP2base | http://bioinf.uta.fi/MASP2base | MASP2 | MASP2 deficiency | |
| MLPHbase | http://bioinf.uta.fi/MLPHbase | MLPH | Griscelli syndrome, type 3 (GS3) | |
| MPObase | http://bioinf.uta.fi/MPObase | MPO | Myeloperoxidase deficiency | |
| MRE11Abase | http://bioinf.uta.fi/MRE11Abase | MRE11A | Ataxia-telangiectasia-like disorder (ATLD) | |

(Continued)

Table 2. Continued

| Database | URL | Gene(s) | Diseases | Remark |
|---|---|----------|---|--------|
| Mutation Database - Papillon Lefevre Syndrome | http://www.genetics.pitt.edu/mutation/pls/ | CTSC | Papillon Lefevre syndrome | |
| MYO5Abase | http://bioinf.uta.fi/MYO5Abase | MYO5A | Griscelli syndrome, type 1 (GS1) | |
| NCF1base | http://bioinf.uta.fi/NCF1base | NCF1 | Autosomal recessive p47phox deficiency | |
| NCF2base | http://bioinf.uta.fi/NCF2base | NCF2 | Autosomal recessive p67phox deficiency | |
| NFKBIAbase | http://bioinf.uta.fi/NFKBIAbase | NFKBIA | Autosomal dominant anhidrotic ectodermal dysplasia and T-cell immunodeficiency | |
| NHEJ1base | http://bioinf.uta.fi/NHEJ1base | NHEJ1 | Combined immunodeficiency (CID) associated with microcephaly and increased cellular sensitivity to IR | |
| NPbase | http://bioinf.uta.fi/Npbase | NP | PNP deficiency | |
| NRASbase | http://bioinf.uta.fi/NRASbase | NRAS | Autoimmune lymphoproliferative syndrome type IV | |
| ORAI1base | http://bioinf.uta.fi/ORAI1base | ORAI1 | Severe combined immunodeficiency | |
| OSTM1base | http://bioinf.uta.fi/OSTM1base | OSTM1 | Autosomal recessive osteopetrosis | |
| PIK3R1base | http://bioinf.uta.fi/PIK3R1base | PIK3R1 | Pathogenic mutations in the p85 α SH2 domain | |
| PRF1base | http://bioinf.uta.fi/PRF1base | PRF1 | Familial haemophagocytic lymphohistiocytosis, type II (FHL2) | |
| PTPN11base | http://bioinf.uta.fi/PTPN11base | PTPN11 | Pathogenic mutations in the SHP-2 SH2 domain | |
| PTPRCbase | http://bioinf.uta.fi/PTPRCbase | PTPRC | CD45 deficiency | |
| RAB27Abase | http://bioinf.uta.fi/RAB27Abase | RAB27A | Griscelli syndrome, type 2 (GS2) | |
| RAC2base | http://bioinf.uta.fi/RAC2base | RAC2 | Neutrophil immunodeficiency syndrome | |
| RAG1base | http://bioinf.uta.fi/RAG1base | RAG1 | RAG1 deficiency | |
| RAG2base | http://bioinf.uta.fi/RAG2base | RAG2 | RAG2 deficiency | |
| RASA1base | http://bioinf.uta.fi/RASA1base | RASA1 | Pathogenic mutations in the RasGAP SH2 domain | |
| RASGRP2base | http://bioinf.uta.fi/RASGRP2base | RASGRP2 | Leukocyte adhesion deficiency III | |
| RFX5base | http://bioinf.uta.fi/RFX5base | RFX5 | MHCII promoter X box regulatory factor 5 deficiency | |
| RFXANKbase | http://bioinf.uta.fi/RFXANKbase | RFXANK | Ankyrin repeat containing regulatory factor X-associated protein deficiency | |
| RFXAPbase | http://bioinf.uta.fi/RFXAPbase | RFXAP | Regulatory factor X-associated protein deficiency | |
| Roche Cancer Genome Database (RCGDB) | http://rcgdb.bioinf.uni-sb.de/MutomeWeb | > 10 000 | multiple tissues and histologies | |
| SBDSbase | http://bioinf.uta.fi/SBDSbase | SBDS | Shwachman–Diamond syndrome | |
| SERPING1base | http://bioinf.uta.fi/SERPING1base | SERPING1 | Hereditary angioedema | |
| SH2base | http://bioinf.uta.fi/SH2base | SH2 | Pathogenic SH2 domain mutations | |

(Continued)

Table 2. Continued

| Database | URL | Gene(s) | Diseases | Remark |
|--------------------------------|---|---|--|--|
| SH2D1Abase | http://bioinf.uta.fi/SH2D1Abase | SH2D1 A | X-linked lymphoproliferative syndrome (XLP) | |
| SLC35C1base | http://bioinf.uta.fi/SLC35C1base | SLC35C1 | Leukocyte adhesion deficiency I I (LAD-II) | |
| SMARCAL1base | http://bioinf.uta.fi/SMARCAL1base | SMARCAL1 | Schimke immuno-osseous dysplasia | |
| SP110base | http://bioinf.uta.fi/SP110base | SP110 | Hepatic veno-occlusive disease with immunodeficiency syndrome (VODI) | |
| SPINK5base | http://bioinf.uta.fi/SPINK5base | SPINK5 | Netherton syndrome | |
| STAT1base | http://bioinf.uta.fi/STAT1base | STAT1 | STAT1 deficiency | |
| STAT3base | http://bioinf.uta.fi/STAT3base | STAT3 | Hyper-IgE syndrome | |
| STAT5Bbase | http://bioinf.uta.fi/STAT5Bbase | STAT5B | Growth hormone insensitivity with immunodeficiency | |
| STX11base | http://bioinf.uta.fi/STX11base | STX11 | Familial haemophagocytic lymphohistiocytosis 4 | |
| TAP1base | http://bioinf.uta.fi/TAP1base | TAP1 | TAP1 deficiency | |
| TAP2base | http://bioinf.uta.fi/TAP2base | TAP2 | TAP2 deficiency | |
| TAPBPbase | http://bioinf.uta.fi/TAPBPbase | TAPBP | Tapasin deficiency | |
| TAZbase | http://bioinf.uta.fi/TAZbase | TAZ | Barth syndrome | |
| TCIRG1base | http://bioinf.uta.fi/TCIRG1base | TCIRG1 | Autosomal recessive osteopetrosis (arOP) | |
| TCN2base | http://bioinf.uta.fi/TCN2base | TCN2 | Transcobalamin II deficiency | |
| The Cancer Genome Atlas (TCGA) | http://cancergenome.nih.gov/ | >400 | Brain (glioblastoma multiforme), ovarian (serous cystadenocarcinoma) | Login for TCGA Data Portal necessary, more disease to come |
| TLR3base | http://bioinf.uta.fi/TLR3base | TLR3 | Influenza-associated encephalopathy | |
| TMC6base | http://bioinf.uta.fi/TMC6base | TMC6 | Epidermodysplasia verruciformis | |
| TMC8base | http://bioinf.uta.fi/TMC8base | TMC8 | Epidermodysplasia verruciformis | |
| TNFRSF13Bbase | http://bioinf.uta.fi/TNFRSF13Bbase | TNFRSF13B | TACI deficiency | |
| TYK2base | http://bioinf.uta.fi/TYK2base | TYK2 | TYK2 deficiency | |
| UMD-ATP7B | http://www.umd.be/ATP7B/ | ATPase, Cu ⁺⁺ transporting, beta polypeptide | Wilson disease | |
| UMD-COL3A1 | http://www.umd.be/COL3A1/ | COL3A1 | COL3A1 deficiency | Restricted access |
| UMD-CSA | http://www.umd.be/CSA/ | ERCC8 | ERCC8 deficiency | |
| UMD-CSB | http://www.umd.be/CSB/ | ERCC6 | ERCC6 deficiency | |
| UMD-DFNB1-GJB2 | http://www.umd.be/DFNB1-GJB2/ | DFNB1, GJB2 | DFNB1 deficiency | Restricted access |
| UMD-DMD | http://www.umd.be/DMD/ | DMD | DMD deficiency | |
| UMD-DPYD | http://www.umd.be/DPYD/ | DPYD | Dihydropyrimidine dehydrogenase disease | Restricted access |
| UMD-EMD | http://www.umd.be/EMD/ | EMD | EMD deficiency | |
| UMD-FBN1 | http://www.umd.be/FBN1/ | FBN1 | Marfan syndrome and related disorders | |
| UMD-FBN2 | http://194.167.35.168/FBN2/ | FBN2 | Congenital contractural arachnodactyly | |

(Continued)

Table 2. Continued

| Database | URL | Gene(s) | Diseases | Remark |
|--------------|---|---------|--|-------------------|
| UMD-LDLR | http://www.umd.be/LDLR/ | LDLR | Familial hypercholesterolemia (FH) | |
| UMD-LMNA | http://www.umd.be/LMNA/ | LMNA | LMNA deficiency | |
| UMD-TGFBR1 | http://www.umd.be/LSDB.html | TGFBR1 | TGFBR1 deficiency | Restricted access |
| UMD-TGFBR2 | http://www.umd.be/TGFBR2/ | TGFBR2 | Marfan syndrome, Loey's–Dietz syndrome, Familial thoracic aortic aneurysms and dissections | |
| UMD-USHbases | http://www.umd.be/usher.html | | Usher syndrome | |
| UNC13Dbase | http://bioinf.uta.fi/UNC13Dbase | UNC13D | Familial hemophagocytic lymphohistiocytosis 3 | |
| UNC93B1base | http://bioinf.uta.fi/UNC93B1base | UNC93B1 | UNC93B deficiency (Herpes simplex encephalitis) | |
| UNGbase | http://bioinf.uta.fi/UNGbase | UNG | UNG deficiency | |
| WASbase | http://bioinf.uta.fi/WASbase | WAS | Wiskott–Aldrich syndrome (WAS) | |
| WASPbase | http://homepage.mac.com/kohsukeimai/wasp/WASPbase.html | WASP | Wiskott–Aldrich syndrome | |
| ZAP70base | http://bioinf.uta.fi/ZAP70base | ZAP70 | ZAP70 deficiency | |

For each database the disease as well as the genes have been covered.

resources to overcome the difficulties for users to know where to find the information they are looking for.

Funding

This work was supported by the Roche Postdoc Fellowship Program.

Conflict of interest. None declared.

References

- Vogelstein,B. and Kinzler,K. (2002) *The Genetic Basis of Human Cancer*. McGraw-Hill Professional.
- Thomas,R.K., Baker,A.C., Debiasi,R.M. et al. (2007) High-throughput oncogene mutation profiling in human cancer. *Nat. Genet.*, **39**, 347–351.
- Wood,L., Parsons,D., Jones,S. et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science*, **318**, 1108–1113.
- Pinkel,D. and Albertson,D.G. (2005) Comparative genomic hybridization. *Annu. Rev. Genomics Hum. Genet.*, **6**, 331–354.
- Wang,Z., Shen,D., Parsons,D.W. et al. (2004) Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science*, **304**, 1164–1166.
- Stephens,P., Edkins,S., Davies,H. et al. (2005) A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nat. Genet.*, **37**, 590–592.
- Davies,H., Hunter,C., Smith,R. et al. (2005) Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res.*, **65**, 7591–7595.
- Bignell,G., Smith,R., Hunter,C. et al. (2006) Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults. *Genes Chromosomes Cancer*, **45**, 42–46.
- Futreal,P.A., Wooster,R. and Stratton,M.R. (2005) Somatic mutations in human cancer: insights from resequencing the protein kinase gene family. *Cold Spring Harb. Symp. Quant. Biol.*, **70**, 43–49.
- Haber,D.A. and Settleman,J. (2007) Cancer: drivers and passengers. *Nature*, **446**, 145–146.
- Futreal,P., Coin,L., Marshall,M. et al. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
- Stratton,M.R., Campbell,P.J. and Futreal,P.A. (2009) The cancer genome. *Nature*, **458**, 719–724.
- Greenman,C., Stephens,P., Smith,R. et al. (2007) Patterns of somatic mutation in human cancer genomes. *Nature*, **446**, 153–158.
- Cleaver,J.E. (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat. Rev. Cancer*, **5**, 564–573.
- Hoeijmakers,J.H. (2009) DNA damage, aging, and cancer. *N. Engl. J. Med.*, **361**, 1475–1485.
- Shen,X., Do,H., Li,Y. et al. (2009) Recruitment of fanconi anemia and breast cancer proteins to DNA damage sites is differentially governed by replication. *Mol. Cell*, **35**, 716–723.
- Capell,B.C., Tlougan,B.E. and Orlow,S.J. (2009) From the rarest to the most common: insights from progeroid syndromes into skin cancer and aging. *J. Invest. Dermatol.*, **129**, 2340–2350.
- Wu,L. (2007) Role of the BLM helicase in replication fork management. *DNA Repair*, **6**, 936–944.
- Osorio,A., Milne,R.L., Pita,G. et al. (2009) Evaluation of a candidate breast cancer associated SNP in ERCC4 as a risk modifier in BRCA1 and BRCA2 mutation carriers. Results from the Consortium of Investigators of Modifiers of BRCA1/BRCA2 (CIMBA). *Br. J. Cancer*, **101**, 2048–2054.
- Kwong,L.N. and Dove,W.F. (2009) APC and its modifiers in colon cancer. *Adv. Exp. Med. Biol.*, **656**, 85–106.

21. Normanno,N., Tejpar,S., Morgillo,F. *et al.* (2009) Implications for KRAS status and EGFR-targeted therapies in metastatic CRC. *Nat. Rev. Clin. Oncol.*, **6**, 519–527.
22. Walther,A., Johnstone,E., Swanton,C. *et al.* (2009) Genetic prognostic and predictive markers in colorectal cancer. *Nat. Rev. Cancer*, **9**, 489–499.
23. Nucera,C., Goldfarb,M., Hodin,R. *et al.* (2009) Role of B-Raf(V600E) in differentiated thyroid cancer and preclinical validation of compounds against B-Raf(V600E). *Biochim. Biophys. Acta*, **1795**, 152–161.
24. Halilovic,E. and Solit,D.B. (2008) Therapeutic strategies for inhibiting oncogenic BRAF signaling. *Curr. Opin. Pharmacol.*, **8**, 419–426.
25. Loriot,Y., Mordant,P., Deutsch,E. *et al.* (2009) Are RAS mutations predictive markers of resistance to standard chemotherapy? *Nat. Rev. Clin. Oncol.*, **6**, 528–534.
26. Ford,D., Easton,D.F., Stratton,M. *et al.* (1998) Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.*, **62**, 676–689.
27. Kadouri,L., Hubert,A., Rotenberg,Y. *et al.* (2007) Cancer risks in carriers of the BRCA1/2 Ashkenazi founder mutations. *J. Med. Genet.*, **44**, 467–471.
28. Thompson,D. and Easton,D.F. (2002) Cancer Incidence in BRCA1 mutation carriers. *J. Natl Cancer Inst.*, **94**, 1358–1365.
29. Lonser,R., Glenn,G., Walther,M. *et al.* (2003) von Hippel-Lindau disease. *The Lancet*, **361**, 2059–2067.
30. Hastings,M.L., Resta,N., Traum,D. *et al.* (2005) An LKB1 AT-AC intron mutation causes Peutz-Jeghers syndrome via splicing at non-canonical cryptic splice sites. *Nat. Struct. Mol. Biol.*, **12**, 54–59.
31. Tinat,J., Bougeard,G., Baert-Desurmont,S. *et al.* (2009) 2009 version of the Chompret criteria for Li Fraumeni syndrome. *J. Clin. Oncol.*, **27**, e108–e109.
32. Sherry,S.T., Ward,M.H., Kholodov,M. *et al.* (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
33. Wheeler,D.L., Barrett,T., Benson,D.A. *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **36**, D13–D21.
34. Frazer,K.A., Ballinger,D.G., Cox,D.R. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
35. McLendon,R., Friedman,A., Bigner,D. *et al.* (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
36. den Dunnen,J.T. and Antonarakis,S.E. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum. Mutat.*, **15**, 7–12.
37. Greenman,C., Wooster,R., Futreal,P.A. *et al.* (2006) Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics*, **173**, 2187–2198.
38. Sjöblom,T., Jones,S., Wood,L.D. *et al.* (2006) The consensus coding sequences of human breast and colorectal cancers. *Science*, **314**, 268–274.
39. Jones,S., Zhang,X., Parsons,D.W. *et al.* (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*, **321**, 1801–1806.
40. Parsons,D.W., Jones,S., Zhang,X. *et al.* (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science*, **321**, 1807–1812.
41. Petitjean,A., Mathe,E., Kato,S. *et al.* (2007) Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. *Hum. Mutat.*, **28**, 622–629.
42. Forbes,S.A., Bhamra,G., Bamford,S. *et al.* (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr. Protoc. Hum. Genet.*, **10**.
43. Stenson,P.D., Ball,E.V., Mort,M. *et al.* (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.*, **21**, 577–581.
44. Beroud,C., Colod-Beroud,G., Boileau,C. *et al.* (2000) UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. *Hum. Mutat.*, **15**, 86–94.
45. Brown,A.F. and McKie,M.A. (2000) MuStaR and other software for locus-specific mutation databases. *Hum. Mutat.*, **15**, 76–85.
46. Fokkema,I.F., den Dunnen,J.T. and Taschner,P.E. (2005) LOVD: easy creation of a locus-specific sequence variation database using an 'LSDB-in-a-box' approach. *Hum. Mutat.*, **26**, 63–68.
47. Edkins,S., O'Meara,S., Parker,A. *et al.* (2006) Recurrent KRAS codon 146 mutations in human colorectal cancer. *Cancer Biol. Ther.*, **5**, 928–932.
48. Hostein,I., Faur,N., Primois,C. *et al.* (2010) BRAF mutation status in gastrointestinal stromal tumors. *Am J. Clin. Pathol.*, **133**, 141–148.
49. Cui,W., Kong,X., Cao,H.L. *et al.* (2008) Mutations of p53 gene in 41 cases of human brain gliomas. *Ai Zheng*, **27**, 8–11.
50. Shoemaker,R.H. (2006) The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev. Cancer*, **6**, 813–823.
51. Kuntzer,J., Eggle,D., Lenhof,H.P. *et al.* (2010) The Roche Cancer Genome database (RCGDB). *Hum. Mutat.*, **31**, 407–413.
52. Vos,Y.J., de Walle,H.E., Bos,K.K. *et al.* (2009) Genotype-phenotype correlations in L1 syndrome: a guide for genetic counselling and mutation analysis. *J. Med. Genet.*, **47**, 169–175.
53. Piirila,H., Valiaho,J. and Vihinen,M. (2006) Immunodeficiency mutation databases (IDbases). *Hum. Mutat.*, **27**, 1200–1208.
54. Claustres,M., Horaitis,O., Vanevski,M. *et al.* (2002) Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Res.*, **12**, 680–688.
55. Horaitis,O. and Cotton,R.G. (2004) The challenge of documenting mutation across the genome: the human genome variation society approach. *Hum. Mutat.*, **23**, 447–452.
56. Kaput,J., Cotton,R.G., Hardman,L. *et al.* (2009) Planning the human variome project: the Spain report. *Hum. Mutat.*, **30**, 496–510.