Atlas of Genetics and Cytogenetics in Oncology and Haematology

OPEN ACCESS JOURNAL INIST-CNRS

Deep Insight Section

General resources in Genetics and/or Oncology

Etienne De Braekeleer, Jean Loup Huret, Hossain Mossafa, Katriina Hautaviita, Philippe Dessen

Haematological Cancer Genetics & Stem Cell Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, United Kingdom; Medical Genetics, Dept Medical Information, University Hospital, F-86021 Poitiers, France; Laboratoire CERBA, 95310 Saint Ouen l'Aumone, France; (Mouse genomics, Wellcome Trust Sanger Institute); UMR 1170 INSERM, Gustave Roussy, 114 rue Edouard Vaillant, F-94805 Villejuif, France.

Published in Atlas Database: April 2016
Online updated version : http://AtlasGeneticsOncology.org/Deep/General_ResourcesID20144.html
Printable original version : http://documents.irevues.inist.fr/bitstream/handle/2042/62779/04-2015-General_ResourcesID20144.pdf
DOI: 10.4267/2042/62779

Abstract

Abstract

This "Deep Insight" is a detailed subchapter of a general review article and summary on Internet databases for cytogeneticists: Internet databases and resources for cytogenetics and cytogenomics.

I- Bibliography

PubMed (http://www.ncbi.nlm.nih.gov/pubmed/)
PubMed is a widely used and free search engine and database of biomedical citations and abstracts, based essentially on the MEDLINE database of references and abstracts on life sciences and biomedical topics. The database is maintained by the National Center for Biotechnology Information (NCBI), at the U.S. National Library of Medicine (NLM), located at the National Institutes of Health (NIH), as part of the Entrez system of information retrieval. From 1971 to 1997, the online version of MEDLINE through computerized database MEDLARS was mainly accessed through institutions, such as university libraries. In 1996, PubMed was launched but only as late as 1997 gave free access of MEDLINE to private home and office computers. PubMed Advanced Search Builder
http://www.ncbi.nlm.nih.gov/pubmed/advanced
uses keywords such as: Affiliation, All Fields, Author, Author First, Author Last, Journal, MeSH Major Topic, Title, Title/Abstract (Figure 1). It uses Booleans (AND, OR, NOT). You can query "(KMT2A[Title]) AND ((Acute myeloid leukemia) OR (Acute lymphoid leukemia))", you will get: Search results : Items 5. This only shows that the official name KMT2A remains totally ignored by scientists. If you replace KMT2A with MLL: "(MLL[Title]) AND ((Acute Myeloid Leukemia) OR (Acute lymphoid leukemia))", you get: Search results : Items 885. Which is what you were looking for. On the other hand, if you misuse the brackets in your query (e.g. "((KMT2A[Title]) AND Acute myeloid leukemia) OR Acute lymphoid leukemia", you will have a huge amount of background noise! Search results: Items 36,244. PubMed comprises of more than 25 million citations of biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites. As a broader research engine, PubMed also runs in several other databases like MEDLINE and Index Medicus, providing older references of the print versions as well as some journals not yet cited, like Science.
The research engine also accesses entries for an article before it gets indexed by the Medical Subject Headings (MeSH) and added to MEDLINE.

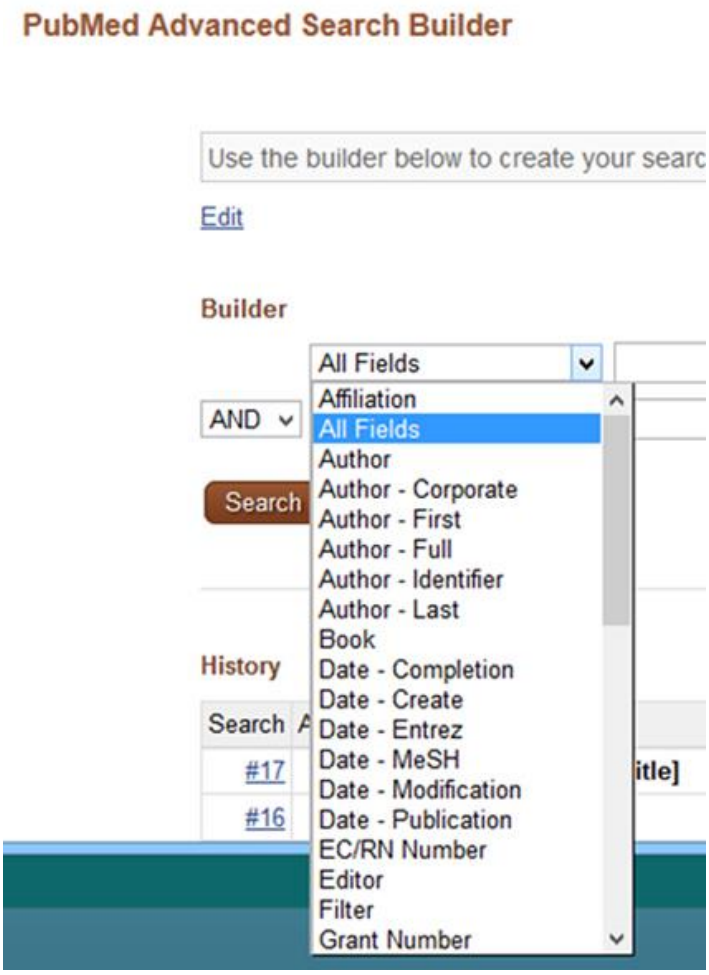


Figure 1: PubMed Advanced Search Builder: choice of fields for a query. (http://www.ncbi.nlm.nih.gov/pubmed/advanced)

Collections of full-text available books and other subsets of NLM records are available (https://www.nlm.nih.gov/pubs/factsheets/pubmed.html). The references catalogued in PubMed often contain links to the full text articles, some of them are free of access and more often in PubMed Central (http://www.ncbi.nlm.nih.gov/pmc/) and local mirrors like UK PubMed Central (http://www.jisc-content.ac.uk/node/52) or Europe PMC (https://europepmc.org/). NLM catalogue contains all the necessary information about the journals that are indexed in PubMed (http://www.ncbi.nlm.nih.gov/nlmcatalog). PubMed records back to 1966, selectively to the year 1865, and very selectively to 1809; about 500,000 new records are added each year. As of this date, 14,026,022 records are listed with their abstracts. Only journals achieving PubMed's scientific standards are indexed which, on the one hand, provides a way to control the quality of scientific publishing. PubMed, free of use, is an immense gift to the medical and scientific community. However, from the scientific editor's viewpoint, this quasi-monopoly position has an adverse aspect: to be referenced by PubMed is a terrifying verdict, in terms of recognition. This is all the more concerning, as the Literature Selection Technical Review Committee's decisions have been known to create controversy among scientific editor's and publisher's communities.

PubMed Central

(http://www.ncbi.nlm.nih.gov/pmc/)

PubMed Central (PMC) is a free archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM) developed by the National Center for Biotechnology Information (NCBI). PubMed Central should not be confused with PubMed. As an archive, PMC is designed to provide permanent access to all of its content. Articles are deposited by participating journals, as well as for author manuscripts that have been submitted in compliance with the public access policies of participating research funding agencies.

Medline

Medline is the U.S. National Library of Medicine (NLM) premier bibliographic database that contains more than 22 million references to journal articles in life sciences with emphasis on biomedicine. A distinctive feature of MEDLINE is that the records

are indexed with NLM Medical Subject Headings (MeSH). MEDLINE is the primary component of PubMed.

Scopus (http://www.scopus.com/)

Scopus is a large abstract and citation database of peer-reviewed literature: scientific journals (more than 60 million records in Scopus, which includes over 21,500 peer-reviewed journals), books (more than 113,000 books) and conference proceedings. It is owned by Elsevier and it is available online by subscription.

II- Nomenclatures

Gene Nomenclature: HGNC (http://www.genenames.org/)

The HUGO Gene Nomenclature Committee (HGNC) is the worldwide authority assigning standardised nomenclature to human genes. HGNC approves unique names and symbols for human loci, including protein coding genes, ncRNA genes and pseudogenes, gene families and associated resources including links to genomic, proteomic and phenotypic information, to allow more unambiguousity to scientific communication. This database contains 39,000 approved symbols (Gray KA et al., 2015).

Nomenclature for the description of sequence variations (http://www.hgvs.org/mutnomen/)

The nomenclature for the description of sequence variations is maintained by the Human Genome Variation Society (HGVS). When describing a variation, first, i) Indicates the reference sequence (e.g. coding DNA: "c."; RNA: "r."; Protein: "p."), followed by ii) the type of variation/mutation (e.g. base substitution: ">"; deletion: "del"). The following codes therefore, have the following meanings: "c.123A>G": on cDNA, A is replaced by G in 123; "p.P252R": on protein, at 252, proline (P) is replaced by arginine (R); "c.546delT": deletion of T in 546; "c.586591del": for six bases deleted, from 586; "p.F508del": deletion of phenylalanine (F) in 508. A summary is proposed at: http://atlasgeneticsoncology.org/Educ/NomMutID30067ES.html.

International System for Human Cytogenetic Nomenclature (ISCN)

ISCN is a language describing abnormal karyotypes. In logic and in mathematics languages with specific grammars have been invented with specific grammars (see https://en.wikipedia.org/wiki/Portal:Logic). The ISCN follows this model. It uses operands and, to act on them, unary and binary operators (e.g. "r" (ring) is an unary operator because it acts on one operand (one chromosome), and "t" (translocation) is a binary operator, because it acts on two operands, (the 2 chromosomes involved in the translocation). ISCN originates at the Denver conference, in 1960 (proposed nomenclature, 1960). By periodic revisions and updates ISCN has now become ever more complicated (ISCN (2013). A new version will be released by the end of 2016: McGowan-Jordan J, Simons A and Schmid M. (eds) (2016) ISCN 2016 An International System for Cytogenomic Nomenclature. Reprint of Cytogenetic and Genome Research 149(1-2), but will not be freely accessible on the web.

International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) (http://www.who.int/classifications/icd/adaptations/oncology/en/)

For reasons of interoperability between different databases it is essential that a common language is found.

The WHO/OMS has established the ICD-O code for International Classification of Diseases - Oncology, first published in 1976. The third edition of ICD-O (ICD-O3) contains an ICD-O3-TOPO, which provides a topographical identifier for different organs (e.g. C220: Liver; C339: Trachea), and an ICD-O3-MORPH, which provides basic and detailed description of pathology (e.g. respectively: 801: Carcinoma, NOS (not otherwise specified); 8013/3: Large cell neuroendocrine carcinoma; 922: Chondrosarcoma, NOS; 9221/3: Juxtacortical chondrosarcoma). A "/0" means: benign tumor (e.g.: 9220/0: Chondroma): "/1" means: borderline malignancy (e.g. 9751/1: Langerhans cell histiocytosis); "/2" means: malignant tumor in situ (e.g. 8500/2: Intraductal carcinoma, noninfiltrating, NOS); and "/3" means full malignancy.

Nosology, thesaurus and census, with phylum of solid tumors and hematological malignancies can be found in the Atlas at: http://atlasgeneticsoncology.org/Tumors/Solid_Nosology.html and: http://atlasgeneticsoncology.org/Anomalies/ICD-O_Hematology.html. This classification is not used by all databases (e.g. the Mitelman database and the COSMIC database use different classifications, with no apparent matching). This makes any integration of data by new resources complicated.

III- Nucleic acid, genes and protein databases

III-1 Nucleic acid databases

The first database for DNA sequencing was The Los Alamos Sequence Database in 1979, which was consequently replaced by public GenBank ((http://www.ncbi.nlm.nih.gov/genbank/) (Burks C et al., 1985) in 1982. The database was funded by the National Institutes of Health, the National Science Foundation, the Department of Energy, and the Department of Defense. Los Alamos National Laboratories (LANL) collaborated with several firms like Bolt, Beranek, and Newman to increase the size of the database. By the end of 1983 more than 2,000 sequences were stored in it.

Mid 1980s, the Intelligenetics bioinformatics company from Stanford University collaborated with LANL to manage the GenBank project (Burks C et al., 1991). Since it was one of the earliest bioinformatics community projects on the Internet, BIOSCI/Bionet news groups was created to promote open access communications among bioscientists. From 1989 to 1992, the GenBank project transitioned to the newly created National Center for Biotechnology Information (Benton D, 1990).

From 1982 to present day, the number of bases in GenBank has doubled roughly every 1,5 years (Benson DA et al., 2015). As of February 2016, GenBank version 212.0 contains 190,250,235 loci, 207,018,196,067 bases, from 190,250,235 reported sequences (http://www.ncbi.nlm.nih.gov/genbank/statistics/).

The GenBank database includes additional data sets that are constructed mechanically from the main sequence data collection, and therefore are excluded from this count. In parallel, the EMBL database was created in 1981 and since this date there is an International Nucleotide Sequence Database Collaboration (INSDC) which is a long-standing foundational initiative that operates between DDBJ, EMBL-EBI and NCBI. INSDC covers the spectrum of data raw reads, though alignments and assemblies to functional annotation, enriched with contextual information relating to samples and experimental configurations. In particular there are numerous evolutions with the development of massive sequencing with creation of more integrated structures as ENA (European Nucleotide Archive at EBI, http://www.ebi.ac.uk/services/dna-rna) or SRA (Sequence read archive at NCBI, http://www.ncbi.nlm.nih.gov/sra/) (Cook CE et al., 2016).

In parallel with the genome projects, the need for the best representation of genomic and transcript sequences for diverse species has been the driver for creating consensus databases (as RefSeq, UCSC, Ensembl) with several methods of optimisation.

III-2 Genes and Functions

Genomic sequences and transcripts

As mentioned in the general resources, several consensus nucleic sequence databases provide detailed structures of genes and isoforms. All the information can easily be visualized using different browsers (UCSC, Ensembl) or described in detail on the Entrez Gene (see above) page at NCBI. RefSeq (http://www.ncbi.nlm.nih.gov/refseq/) maintains and curates a database recording annotated genomic, transcript, and protein sequences. RefSeq release 71 provides sequences from over 55,000 organisms (more than 4,800 viruses, 40,000 prokaryotes and 10,000 eukaryotes) (O'Leary NA et al., 2016). Ensembl (http://www.ensembl.org/) is a joint project between EMBL-EBI and the Wellcome Trust Sanger Institute to develop a software which develops and maintains automatic annotation of selected eukaryotic genomes (Gray KA et al., 2015).

The UCSC Genome Browser database is a large collection of 160 genome assemblies representing 91 species (Rosenbloom KR et al., 2015) (Figures 2 and 3: PAX5 at UCSC and at the Atlas site respectively).

Some standardisation within CCDS and GenCode http://www.gencodegenes.org/ gives an up-to-date information on them.

The nature of isoforms, expressed differently in normal tissues and in tumors, due to splicing variety, leads to protein product with different amino acid sequences.

This reflects the variations in the structure in domains and in the 3D structure are the basis of the activity. On the other hand, the level of expression of transcript in different tissues can be obtained from SOURCE, GEO (Clough and Barett, 2016), Expression Atlas (Petryszak et al., 2016), Gene expression viewer (Firebrowse), BioGPS (http://biogps.org/#goto=welcome) (Wu C et al., 2016) (Figure 4).

III-3 Protein sequence databases

In parallel with the nucleic databases, the first protein database was established by M. Dayhoff as NBRF protein database in 1983, in continuity of the first comprehensive collection of macromolecular sequences in the Atlas of Protein Sequence and Structure, published from 1965-1978.

This was followed by the development of SwissProt, a curated dataset, by Amos Bairoch in 1986 (http://www.isb-sib.ch/sp30/the-history-of-swiss-prot).

With collaboration between the Swiss Institute of Bioinformatics and the EBI to lead in 2002 (in association with the PIR database) the SwissProt was extended to UniProt Knowledgebase (UniProtKB) in 1998, consisting in the curated UniProtKB/Swiss-Prot databank, its automatically annotated supplement TrEMBL, and the PIR protein database.

Today, UniProtKB represents the world's most comprehensive catalogue of information on proteins. In the space of 30 years, the number of proteins entered in UniProtKB/Swiss-Prot has increased from 4,000 to 550,000 : 550,960 entries for the SwissProt part and 63,686,057 entries for the non-reviewed part for TrEMBL (Pundir S et al., 2015).

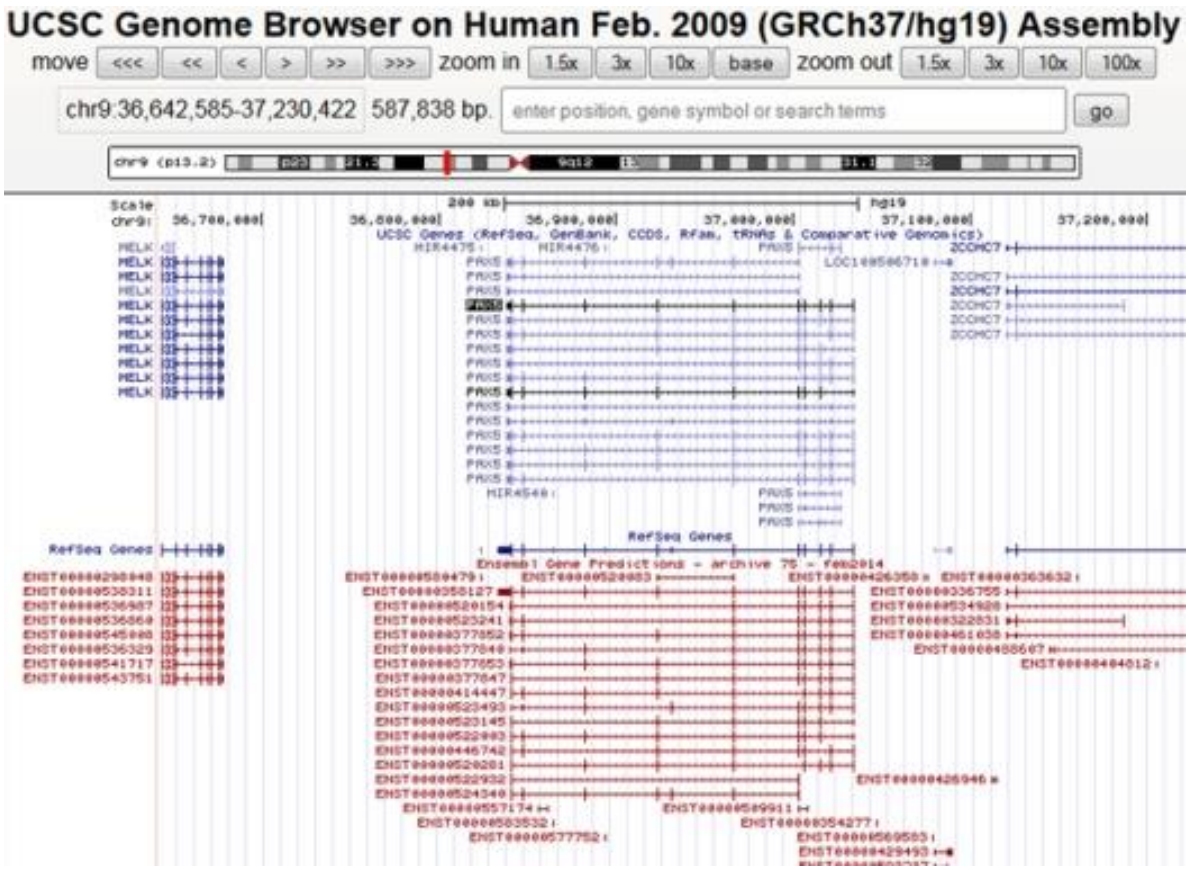


Figure 2: PAX5 gene with isoforms at UCSC (http://genome-euro.ucsc.edu/cgi-bin/hgGateway), Select Species: "Human"; Human Assembly: "Dec. 2013 (GRCh38/hg38)"; Position/Search Term: write "PAX5"; go!)

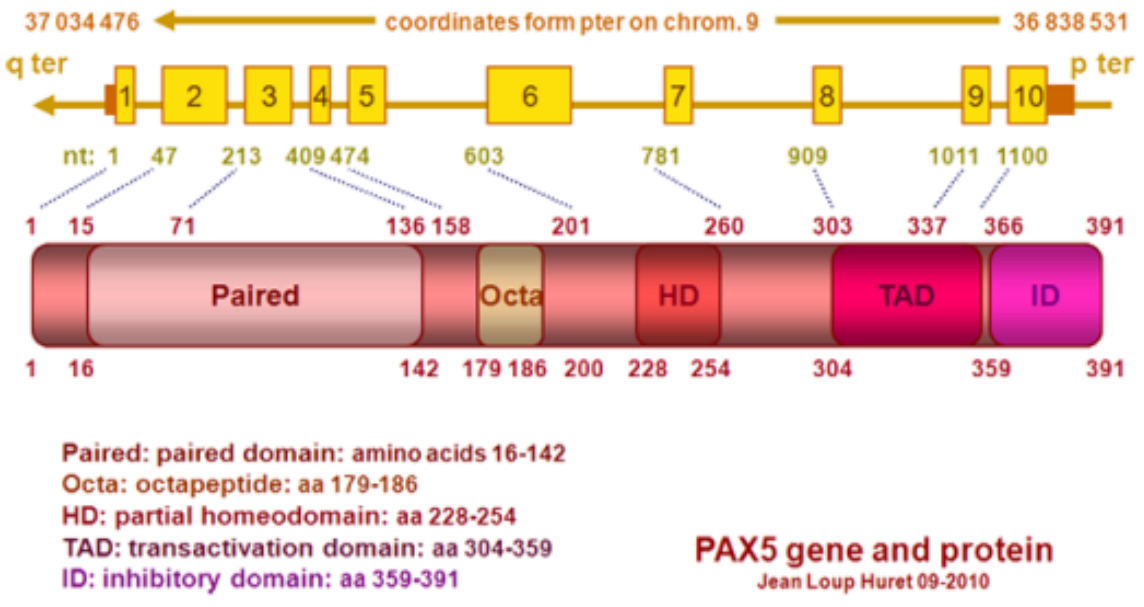


Figure 3: PAX5 gene and protein in the Atlas (http://atlasgeneticsoncology.org//Genes/PAX5ID62.html)

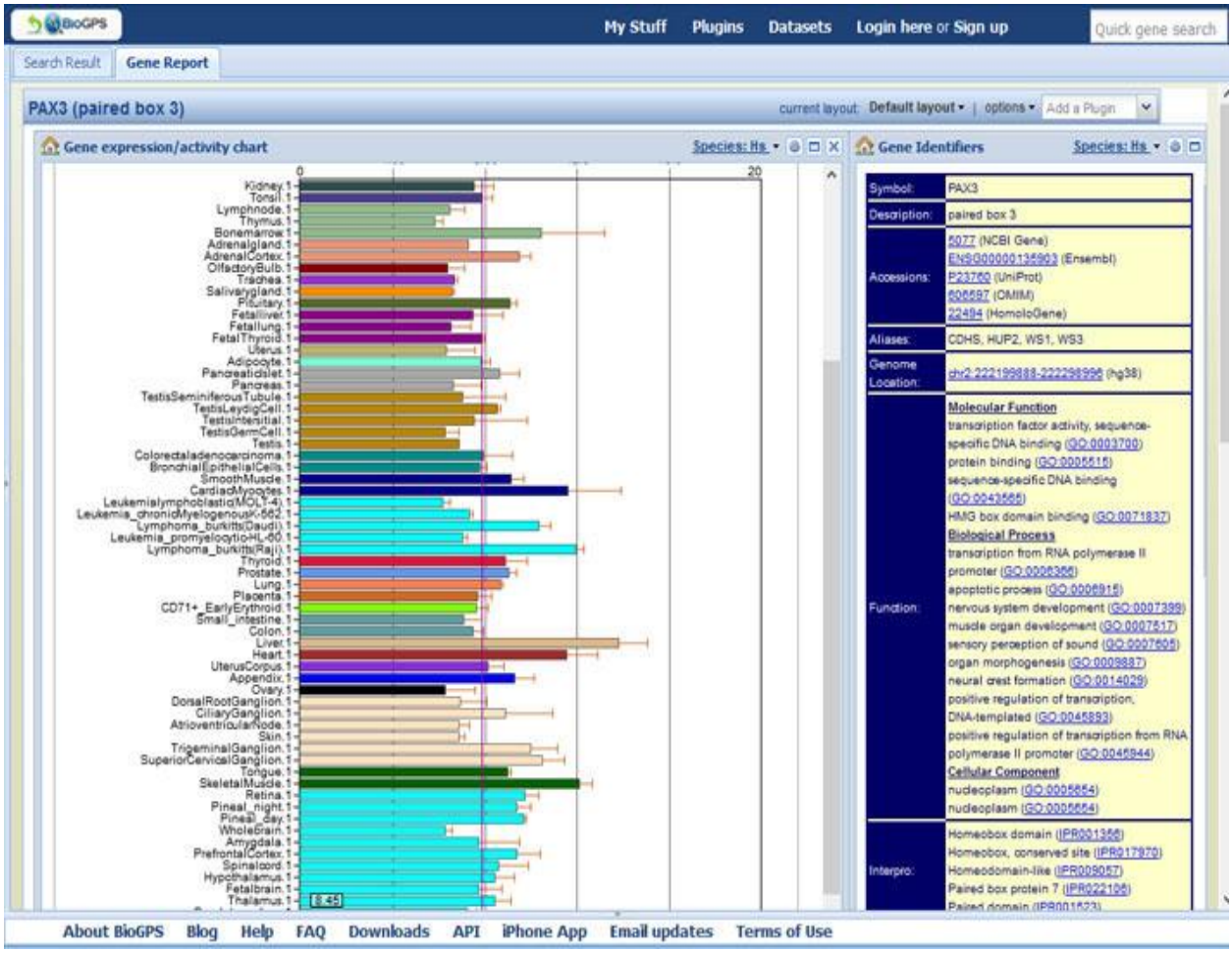


Figure 4: Expression of PAX3 in various tissues at BioGPS (http://biogps.org/#goto=genereport&id=5077)

UniProt (http://www.uniprot.org/)
The UniProt Knowledgebase (UniProt Consortium, 2015) (UniProtKB, http://www.uniprot.org/uniprot/) is a hub for the collection of information on proteins with annotation. In addition to amino acid sequences, protein names and domain descriptions, taxonomic data and citation information, it also provides brief annotation information (Figure 5). UniProtKB consists of two sections: computationally analyzed "TrEMBL" and manually annotated "Swiss-Prot", with information extracted from curator-evaluated computational analysis and literature. UniProt is a collaboration between the European Bioinformatics Institute (EMBL-EBI, http://www.ebi.ac.uk/), the SIB Swiss Institute of Bioinformatics (http://www.isb-sib.ch/) and the Protein Information Resource (PIR, http://pir.georgetown.edu/.)
It is comprised of two seperate tools: the Basic Local Alignment Search Tool (BLAST, http://www.uniprot.org/blast/), to find a region of local similarity between amino acids sequences used in identifying members of a gene family, and Align (http://www.uniprot.org/align/) to align two or more protein sequences.

neXtProt (http://www.nextprot.org/db/)neXtProt (Gaudet P et al., 2015) is a resource for human proteins, including information on the exons, proteins sequences, function, subcellular localisation, expression, interactions and role in diseases (Figure 6). The major part of the information in neXtProt is obtained from the UniProt Swiss-Prot database but is gradually being complemented by original data. neXtProt contains 20,055 protein entries, and is maintained by Amos Bairoch at the Swiss Institute of Bioinformatics and GeneBio.

Figure 5: PAX5 at UniProtKB (http://www.uniprot.org/uniprot/Q02548)

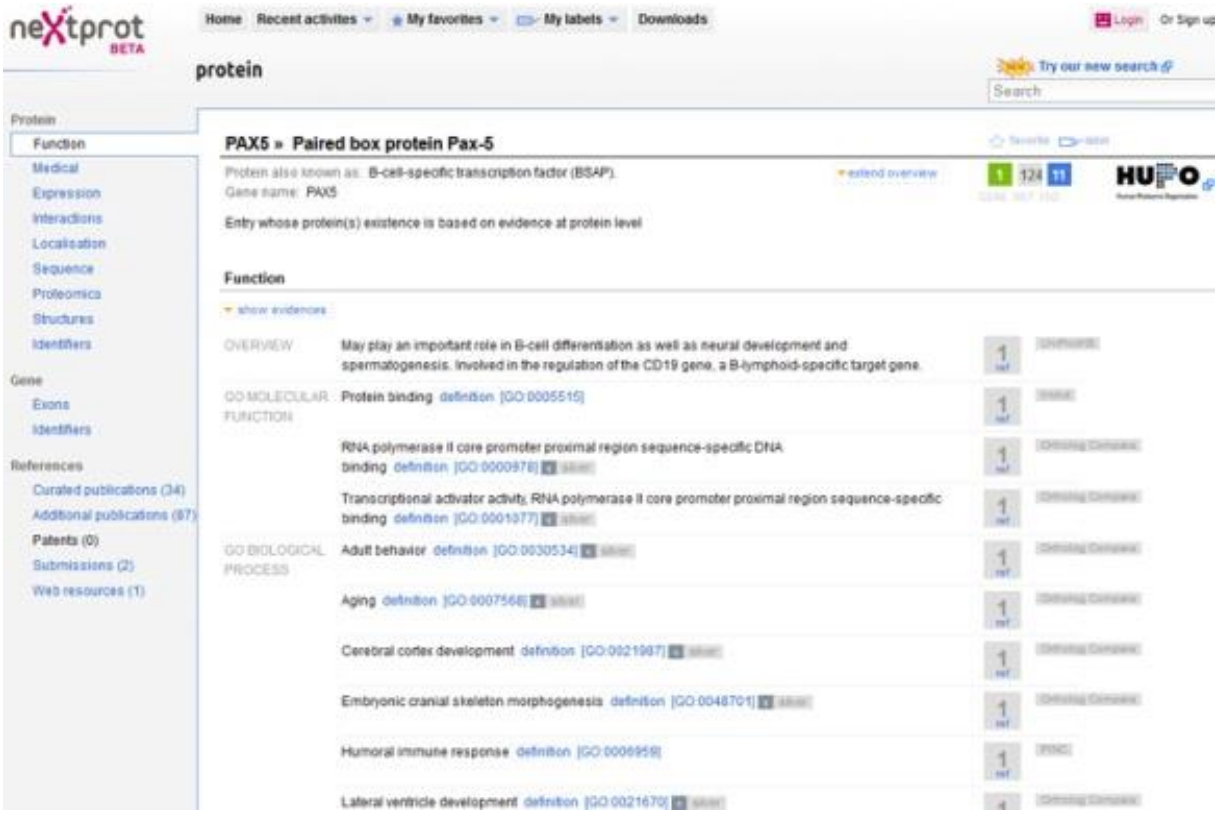


Figure 6: PAX5 at neXtProt, tab "Function" (see on the left) (http://www.nextprot.org/db/entry/NX_Q02548)

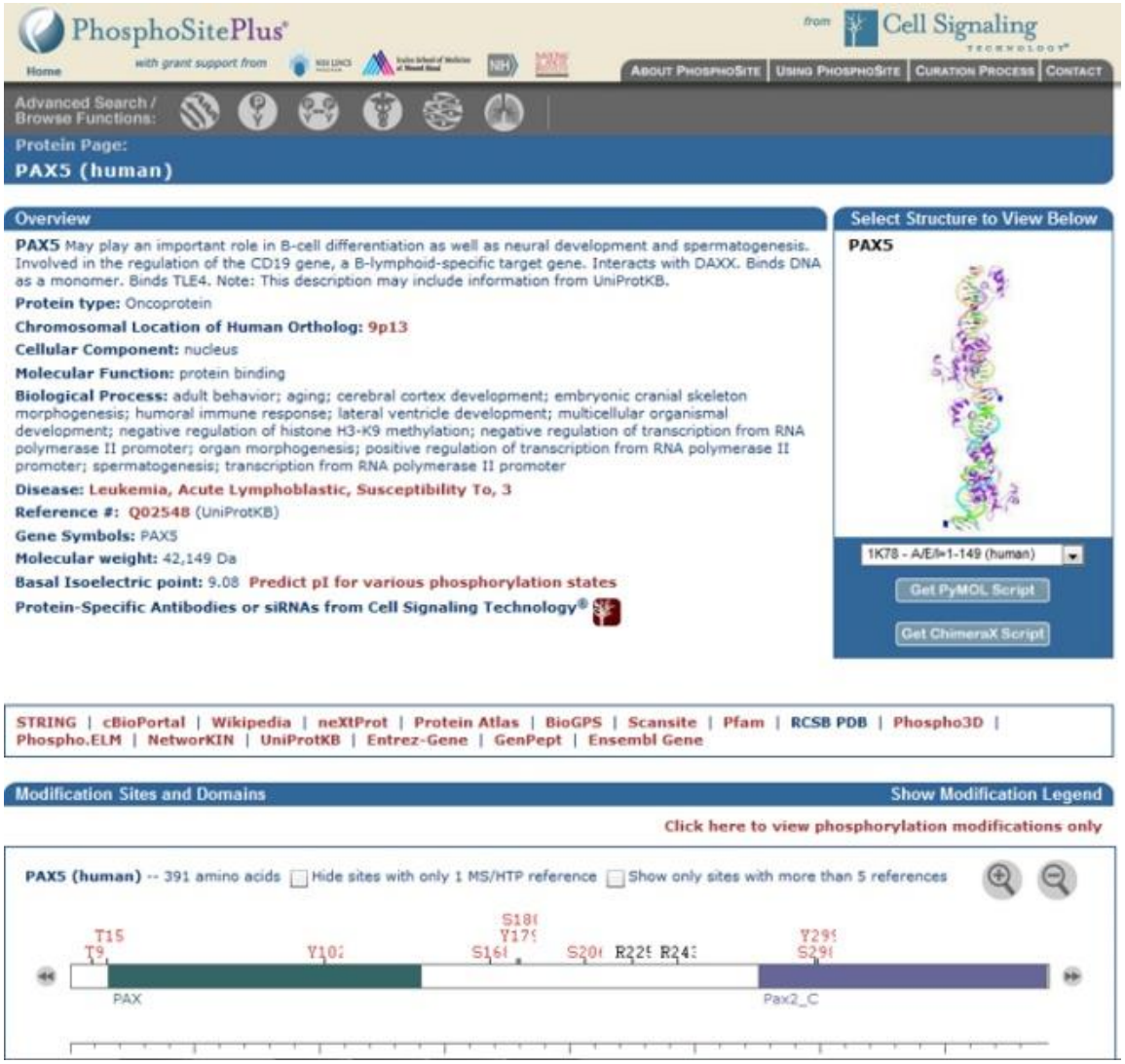


Figure 7: PAX5 at PhosphoSitePlus (http://www.phosphosite.org/proteinAction.action?id=19058&showAllSites=true)

PhosphoSitePlus
(http://www.phosphosite.org/homeAction.action)
PhosphoSitePlus (Hornbeck PV et al., 2015) is an excellent resource providing comprehensive information and tools for the study of protein post-translational modifications (PTMs) including phosphorylation, ubiquitination, acetylation and methylation (Figure 7). PhosphoSitePlus contains curated data on 53,219 human, mouse and to a lesser extent rat proteins, with protein name, protein type, domain, cellular component, and molecular weight. It is an excellent website. PhosphoSitePlus is based at Cell Signaling Technology, Danvers, Massachusetts.

PROSITE (http://prosite.expasy.org/)
PROSITE (Sigrist CJ et al., 2013) is one of the oldest catalogs of protein signatures, consisting of documentation entries describing protein domains, families and functional sites, via a specific pattern of conserved residues (manually defined). PROSITE contains 1756 documentation entries.

Pfam (http://pfam.xfam.org/)
Pfam (Finn RD et al., 2016) is a collection of multiple sequence alignments and hidden Markov models covering many common protein domains. The identification of domains that occur within proteins can provide insights into their function. Pfam contains 16295 entries. InterPro and Pfam are based at EMBL-EBI.

InterPro (http://www.ebi.ac.uk/interpro/)
InterPro integrates PROSITE, Pfam and certain other resources in order to provide functional analysis of proteins by classifying them into families and predicting domains (with signatures) and important sites; InterProScan is the software package that allows sequences to be scanned against InterPro's signatures (Mitchell A et al., 2015).

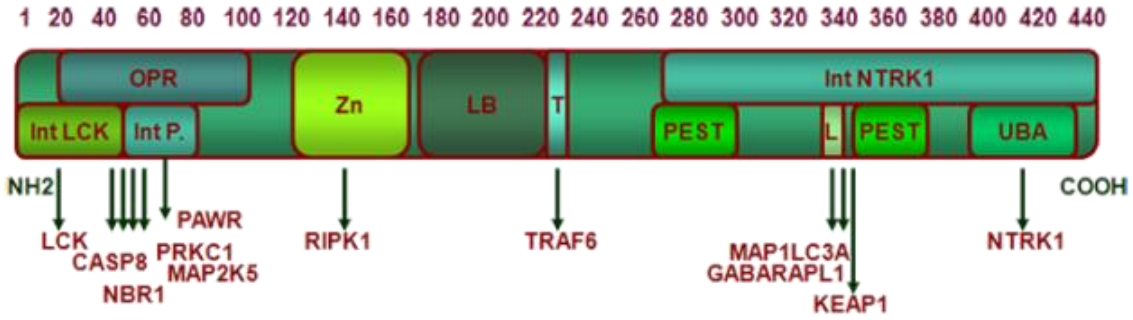
1132 amino acids (aa)
Interaction with cytokine/interferon/growth hormone receptors region: aa 1-239
FERM domain: aa 37-380
SH2: aa 401-482
Protein kinase 1 domain: aa 545-809
Protein kinase 2 Domain:aa 849-1124
A: ATP Nucleotide binding: aa 855-863
L: loop structure: aa 1056-1078 (JAK2 kinase insertion loop)

JAK2 (9p24.1)
Jean Loup Huret 2014
© AtlasGeneticsOncology

JAK Homology domains: JH7 aa 25-137; JH6: aa 144-284; JH5: aa 288-309;
JH4: aa 322-440; JH3: aa 451-538; JH2: aa 543-824; JH1: 836-1123

◆ Phosphotyrosine: aa 119, 372, 373, 523, 813, 868, 966, 972, 1007, 1008

According to Harpur et al., 1992, PMID:1620548; Saltzman et al.,1998, PMID:9618263;
Lucet et al., 2006, PMID:16174768, figure herein for JH7, and Swiss-Prot



SQSTM1 (Sequestosome-1) / p62 (440 amino acids)

Interaction with LCK (aa 1-50)
Int P. : Interaction with PAWR (aa 50-80)
OPR (PB1) (aa 20-102)
Zinc finger (ZZ-type) (aa 122-167)
LB ; LIM-binding (aa 170-220)
T : TRAF6-binding (aa 228-233)
PEST (aa 266-294)
Interaction with NTRK1 (aa 269-440)
L : LIR motif (SGGDDDWTHLSS) (aa 332-343)
PEST (aa 345-377)
UBA (aa 389-434)

SQSTM1 (5q35.3)
Jean Loup Huret 2012,
Adapted from Geetha and Wooten, 2002;
Moscat and Diaz-Meco, 2009; Moscat et al., 2009;
Ichimura and Komatsu, 2010;
amino acids are numbered as in Swiss-Prot.

Figure 8: and 9: JAK2 and SQSTM1 at Atlas: protein domains (http://atlasgeneticsoncology.org//Genes/JAKID98.html and http://atlasgeneticsoncology.org//Genes/GC_SQSTM1.html) There are also data and iconography on pathways (Figure 10).

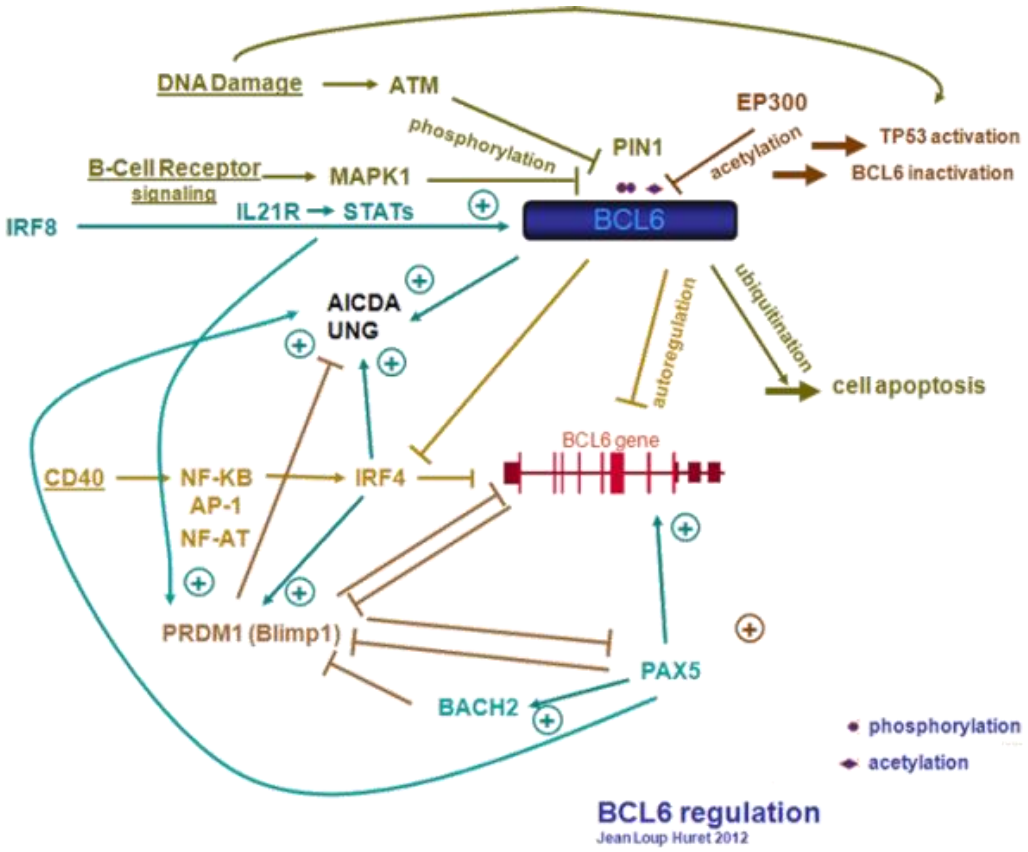


Figure 10: BCL6 regulation (involving PAX5) in the Atlas (http://atlasgeneticsoncology.org//Genes/BCL6ID20.html)

Atlas of Genetics and Cytogenetics in Oncology and Haematology

The Atlas presents highly curated paragraphs with the description of the protein listing domains and iconography, expression and localisation, function, homologs, and uniquely, a wide angle on cancers and other medical conditions where a gene or a protein is implicated (Figure 8 and 9).

IV- Cards

IV-1 Entrez Gene (http://www.ncbi.nlm.nih.gov/gene/)

Entrez is NCBI's primary text search and retrieval system integrating the PubMed database of biomedical literature with 39 other literature and molecular databases including DNA and protein sequences, structures, genes, genomes, genetic variation and gene expression. Entrez Gene, dedicated to gene information, integrates data from a wide range of species. A record can include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype- and locus-specific resources worldwide. Entrez Gene catalogs 59,941 human genes. Entrez Gene can be queried as a free text but also via a syntax with specific fields or filters (e.g. BRCA1[sym] ; 2[chr] AND adh*[sym] ;..) with output in different formats. Once a result is obtained as a list of gene symbols, it is possible to link it to related data in another part of the Entrez database (e.g. list of publication in PubMed from a selected list of genes symbols) (NCBI Resource Coordinators, 2016).

IV-2 Genecards (http://www.genecards.org/)

Genecards is an integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. It automatically integrates data from roughly 125 web sources and includes genomic, transcriptomic, proteomic, genetic, clinical and functional information.

There are some affiliated databases as MalaCards "The human disease database" (http://www.malacards.org/) which is an integrated database of human diseases and their annotations, modeled on the architecture and richness of the GeneCards database of human genes (Fishilevich S et al., 2016).

V- Genome cartography

The cartography of genes on a genome has been the favoured mean to represent genomic information. With the human Genome Project, several types of viewers have been developed. To date, two sites are of first interest for human genetics:

V-1 UCSC (http://genome.ucsc.edu/) and UCSC-Cancer (https://genome-cancer.ucsc.edu/.)

The UCSC Genome Browser contains a reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to ENCODE data at UCSC (2003 to 2012).

The Genome Browser zooms and scrolls over chromosomes, presenting the work of annotators worldwide. The "Gene Sorter" shows expression, homology and other information on groups of genes that can be related in many ways (with a chosen set of tracks). "Blat" maps sequences to the genome quickly. The Table Browser provides convenient access to the underlying database. "VisiGene" lets you browse through a large collection of in situ mouse and frog images to examine gene expression patterns. "Genome Graphs" allows you to upload and display genome-wide data sets. The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the UC Santa Cruz Genomics Institute at the University of California Santa Cruz (UCSC).

A parallel browser has been developed for visualizing and analysing cancer data. The UCSC Cancer Browser https://genome-cancer.ucsc.edu/proj/site/help/) allows researchers to explore cancer genomics data and its associated clinical information in an interactive manner. Data can be viewed in several different ways, including by value, chromosome location, clinical features, biological pathways or genes of interest. It is also possible to quickly perform and easily view statistical analysis on subsets of the data. The data heatmap displays genome-wide data from copy number, transcriptome, protein, epigenetic, mutation, sh/siRNA, and PARADIGM pathway analysis studies as well as associated clinical information. The left column shows datasets that are currently in view along with a button to add more. Today the system has 720 datasets for an exploration (Goldman M et al., 2015).

V-2 Ensembl (http://www.ensembl.org)

Ensembl produces genomic datasets through a system that is designed to analyse, store and distribute data, and which enables interpretation through open data release. As a hub of reference and baseline data similar to UCSC Genome Browser and RefSeq, Ensembl also distributes created datasets and promotes standards and interoperability between genomic resources. In addition, Ensembl collaborates with and often plays active leadership roles in projects such as ENCODE, the "Genome Reference Consortium" (GRC), the "Global Alliance for Genomics and Health" (GA4GH) and GENCODE. Ensembl is updated 4-5 times annually with each release representing a data and software freeze. Ensembl provides two sets of human data based on the hg19 genome build (http://grch37.ensembl.org/Homosapiens/Info/Index) which has been updated by the data set based on the December 2013 Homo sapiens high coverage assembly GRCh38 from the Genome Reference Consortium. This assembly is used by UCSC to create their hg38 database. The data set consists of gene models built from the alignments (for comparison) of the human proteome as well as from alignments of human cDNAs. This release of the assembly has the following properties: assembly length with a total of 3.4 Gb, chromosome length total 3.1 Gb (excluding haplotypes). It also includes 261 alternate loci scaffolds, mainly in the LRC/KIR complex on chromosome 19 (35 alternate sequence representations) and the MHC region on chromosome 6 (7 alternate sequence representations) (Yates A et al., 2016).

VI- Structural variation databases

Since the mid 2000's, there were several studies of copy number variation of DNA sequences to construct CNV map of the human genome through different populations using SNP genotypes and CGH (Iafrate AJ et al., 2004; Redon R et al., 2006). It is becoming clear that genomic structural variation (variation ranging from tens to millions of base pairs in size, and including insertions, deletions, inversions, translocations and locus copy number changes) accounts for individual differences at the DNA sequence level in humans and can play a major role in diseases. Many databases have integrated data produced in the literature.

VI-1 dbVar (http://www.ncbi.nlm.nih.gov/dbvar/)

dbVar is the NCBI's database of genomic structural variation. It contains data of insertions, deletions, duplications, inversions, multi-nucleotide substitutions, mobile element insertions, translocations, and complex chromosomal rearrangements (NCBI Resource Coordinators, 2016).

VI-2 DGV - Genomic Variants (http://dgv.tcag.ca/dgv/app/home)

DGV is a database with an objective to provide a comprehensive summary of structural variation in the human genome. Structural variation is defined as genomic alterations that involve segments of

DNA that are larger than 1kb. It also annotates InDels in 100bp-1kb range. The content of the database is only representing structural variations identified in healthy control samples (MacDonald JR et al., 2014).

VI-3 DECIPHER (https://decipher.sanger.ac.uk/)

DECIPHER (DatabasE of Genomic variants and Phenotype in Humans using Ensembl Resources) is an interactive web-based database which incorporates a series of tools designed to aid the interpretation of genomic variants. DECIPHER enhances clinical diagnosis by retrieving information from a variety of bioinformatics resources relevant to the variant found in a patient. The patient's variant is displayed in the context of both normal variation and pathogenic variation reported at that locus, thereby facilitating interpretation (Firth HV et al., 2009).

VI-4 1000 Genomes (http://www.1000genomes.org/)

The 1000 Genomes Project beneficiated from the progress in sequencing technology, which sharply reduced the cost of sequencing. It was the first project to sequence the genomes of a large number of people, to provide a comprehensive resource on human genetic variation. Data from the 1000 Genomes Project was quickly made available to the worldwide scientific community through freely accessible public databases.

In continuation of the 1000 Genome project (sequencing 1000 human genome as exomes or whole genomes), the International Genome Sample Resource (IGSR) aims to expand information to new populations, a better coverage for presenting a uniform analysis set. Data corresponds to both single nucleotide and structural variants (1000 Genomes Project Consortium et al., 2015).

The 1000 Genomes Project operated between 2008 and 2015, creating the largest public catalogue of human variation and genotype data. As the project ended, the Data Coordination Centre at EMBL-EBI received continuous funding from the Wellcome Trust to maintain and expand the resource. The International Genome Sample Resource (IGSR) is maintaining and extending the 1000 Genomes Project data.

VII- Polymorphism databases

It is important to distinguish between polymorphisms due to a change in a single nucleotide (SNP) as the variability within a population and mutations acquired in a neoplastic process. The determination of variants was previously obtained by SNP arrays, but is nowadays performed by massive parallel sequencing. As a result, a huge quantity of polymorphisms and mutations in tumors are compared to controls. The landscape of the majority of recurrent mutations is now known and can be used for diagnosis.

VII-1 dbSNP (http://www.ncbi.nlm.nih.gov/SNP/overview.html)

dbSNP is the main repository of Single Nucleotide Polymorphisms: A key aspect of research in genetics is associating sequence variations with heritable phenotypes. The most common variations are single nucleotide polymorphisms (SNPs), which occur approximately once every 100 to 300 bases. Because SNPs are expected to facilitate large-scale association genetics studies, there has recently been great interest in SNP discovery and detection. The database contains 164,986,514 SNPs for several species (NCBI Resource Coordinators, 2016).

VII- 2 HAPMAP (http://hapmap.ncbi.nlm.nih.gov/index.html.en)

The International HapMap Project was a collaboration of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom and the United States who wanted to develop a public resource that helps researchers find genes associated with human disease and consequently give response to pharmaceuticals. The goal of the project was to compare genetic sequences of different individuals in order to identify chromosomal regions where genetic variants are shared. An interface (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap28B36/) permits to query all the data collected in phases 1, 2 and 3 of the project (International HapMap 3 Consortium et al., 2010).

VII-3 1000 Genomes Project (see above)

As the Phase3, 1000 Genomes variants are in the process of being archived at dbSNP and DGVa and a version of the Ensembl databases has been created, containing the phase3 autosomal variants. This is presented alongside the v80 GRCh37 Ensembl core and regulatory databases. This release represents more than 80M short variants with genotypes for 2,504 individuals across 26 populations. The latest major update was realeased to the 1000 Genomes Website in February 2016 (1000 Genomes Project Consortium et al., 2015).

VII- 4 Exome Variant server (EVS) (http://evs.gs.washington.edu/EVS/)

The goal of the NHLBI GO Exome Sequencing Project (ESP) is to discover novel genes and mechanisms contributing to heart, lung, and blood

disorders by pioneering the application of next-generation sequencing of the protein coding regions of the human genome across diverse, richly-phenotyped populations and to share these datasets and findings with the scientific community to extend and enrich the diagnosis, management and treatment of the aforementioned disorders. Two categories of populations are considered: European-American and African-American. Some criteria or impact scores of the variation on the gene function are also presented (Tennessen JA et al., 2012).

VIII- Portals/Working consortium

VIII-1 TCGA (http://cancergenome.nih.gov/)

Since 2005 TCGA (The Cancer Genome Atlas) has indexed genetic mutations responsible for cancer, using genome sequencing and bioinformatics. TCGA applies high-throughput genome analysis to progress our ability to diagnose, treat, and prevent cancer.

TCGA is administered by the National Cancer Institute's Center for Cancer Genomics and the National Human Genome Research Institute funded by the US government. A pilot project, initiated in 2006, focused on analysing three types of human cancers: Glioblastoma multiforme, lung cancer, and Ovarian cancer (Cancer Genome Atlas Research Network, 2011).

In 2009, a second phase started, 20-25 different tumor types were included to complete the genomic characterization and sequence analysis (Figure 11). TCGA surpassed that goal, characterizing 33 different cancer types including 10 rare cancers (http://cancergenome.nih.gov/abouttcga/overview).

Funding is split between genome characterization centers (GCCs), which perform the sequencing, and genome data analysis centers (GDACs), which perform the bioinformatic analyses.

The project scheduled 500 patient samples using several analysing techniques: Gene expression profiling, copy number variation profiling, SNP genotyping, genome wide DNA methylation profiling, microRNA profiling, and exon sequencing of 1,200 or more genes (Figure 12). TCGA is sequencing some tumors, including at least 6,000 candidate genes and microRNA sequences.

This targeted sequencing is being performed by all three sequencing centers using hybrid-capture technology. In phase II, TCGA is performing whole exon sequencing on 80% of the cases and whole genome sequencing on 80% of the cases used in the project.

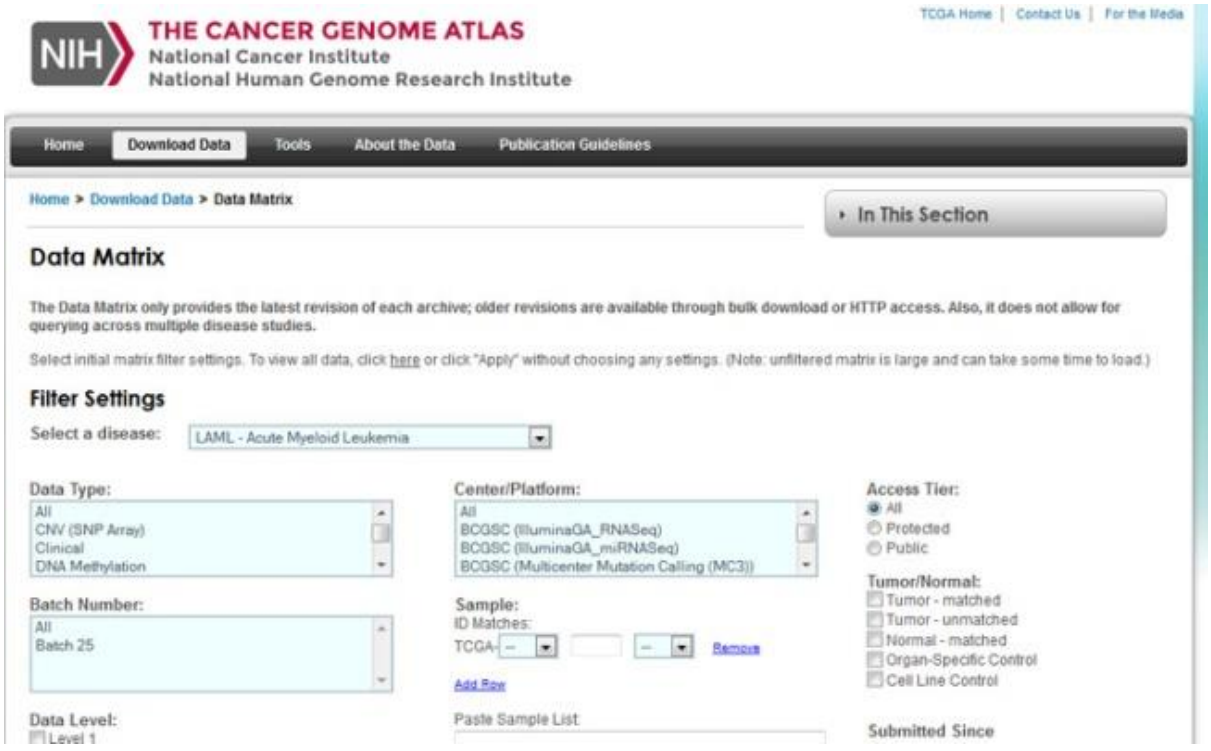


Figure 11: Acute Myeloïd Leukemia query in TCGA datasets with the Data Matrix option (https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm?mode=ApplyFilter)

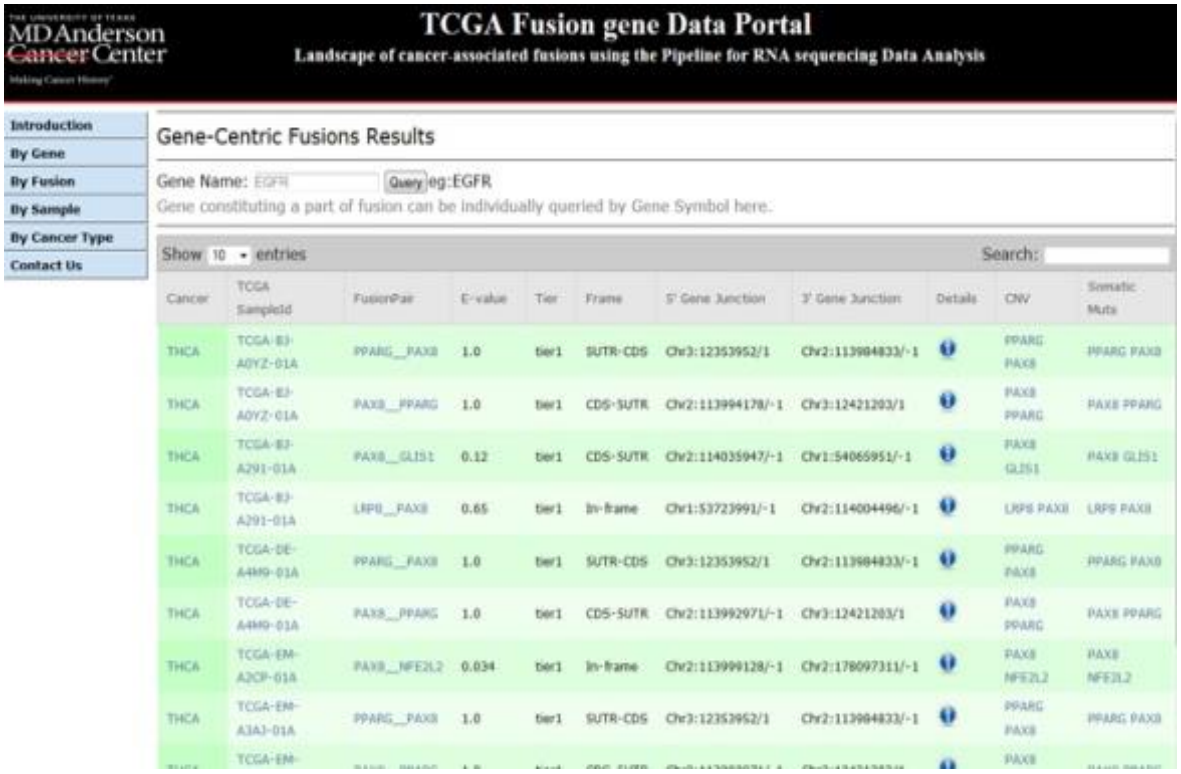


Figure 12: PAX8 gene fusions in TCGA (http://54.84.12.177/PanCanFusV2/)

VIII-2 ICGC (https://icgc.org/)

ICGC (The International Cancer Genome Consortium) was organized to launch and coordinate a large number of research projects with the common aim of comprehensively elucidating the genomic changes present in many forms of cancers.

Funding and Research members proposing a project must agree to the ICGC's policies (Figure 13). ICGC's primary objectives are to generate comprehensive catalogues of genomic abnormalities (somatic mutations, abnormal expression of genes, epigenetic modifications) in tumors representing 50 different cancer types and/or subtypes which are of clinical and societal importance across the globe and make the data available to the entire research community.

Each of the 50 projects will generate the genomic analyses on approximately 500 cancer samples of each class.

This will cover the various types and subtypes but cannot exhaustively cover the full spectrum of cancer types. The ICGC facilitates communication among the members and provides a forum for coordination with the objective of maximizing efficiency among the scientists working to understand, treat, and prevent these diseases.

ICGC data release 20 (November 2015) comprises data from 14,767 cancer genomes in total.

The ICGC Data Portal (https://dcc.icgc.org/) is developed by the Ontario Institute for Cancer research (OICR) (Zhang J et al., 2011).

VIII-3 OASIS (http://www.oasis-genomics.org/)

OASIS, which was created by Pfizer Oncology Research Computational Biology in collaboration with Research Business Technology (RBT), is an open-access web portal that provides the possibility to run exploratory and integrative analyses of somatic mutations, copy number variation (CNV) and gene expression data (Figure14).

This data originates from thousands of different tissues of tumour samples, normal tissues and cell lines thus representing a broad spectrum of malignancies. This portal contains 30 datasets, mainly from TCGA, with access to mutations, copy number variation, expression (microarrays) and expression (RNA-Seq).

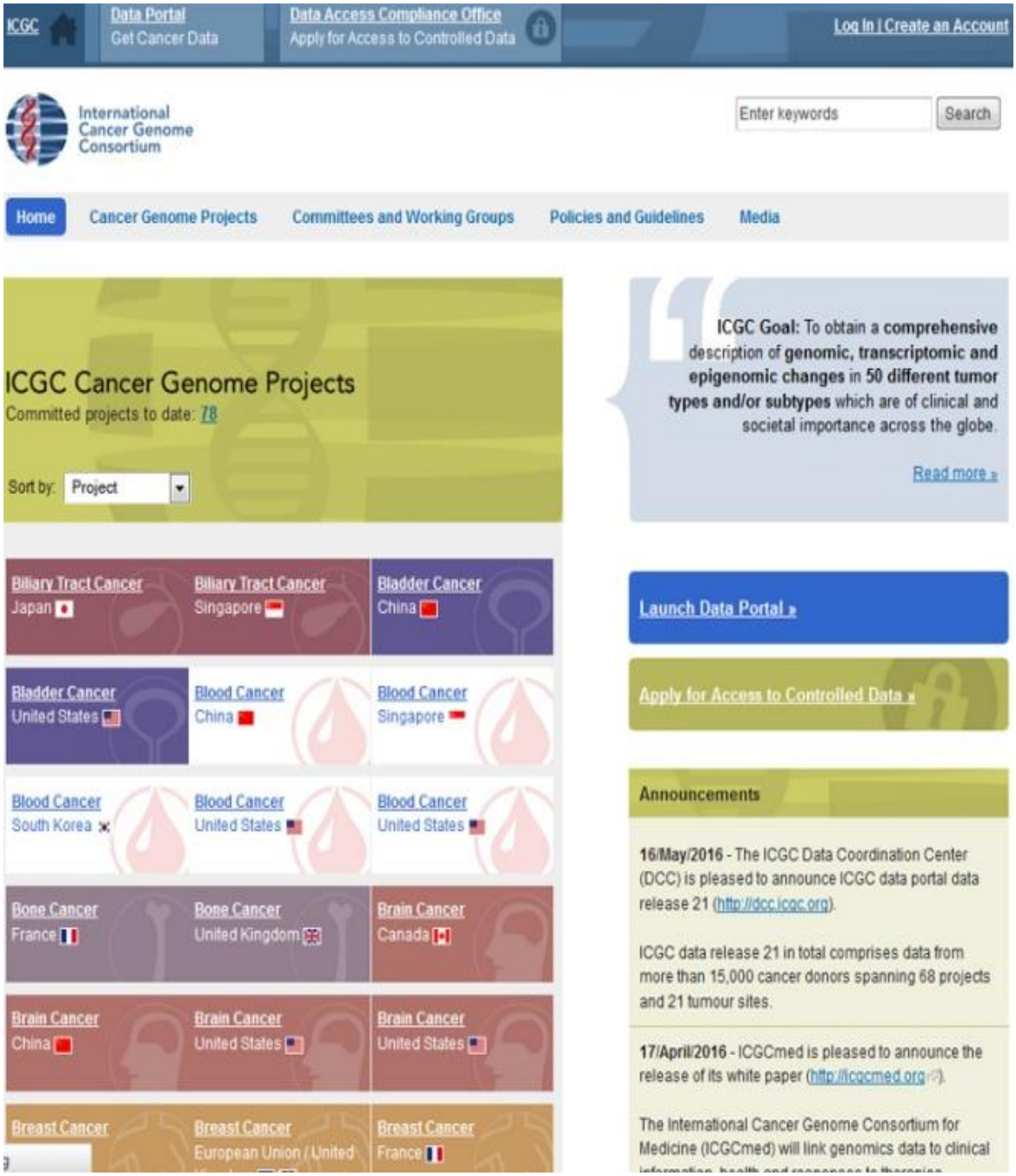


Figure 13: ICGC International Cancer Genome consortium: Home page (https://icgc.org/)

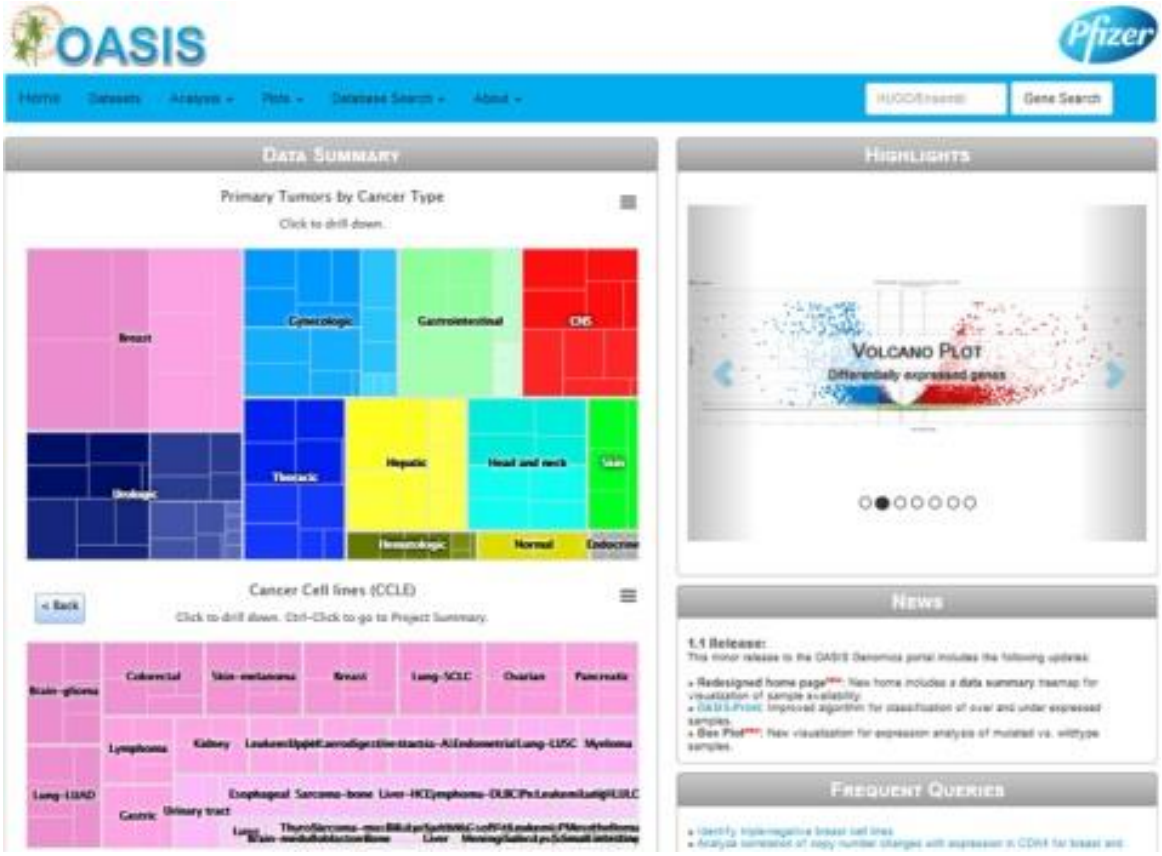


Figure 14: OASIS portal Home page (http://www.oasis-genomics.org/)

VIII-4 Firebrowse (http://firebrowse.org/)

This portal developed at the Broad Institute presents 38 cancer cohorts and 14,729 samples, mainly from the TCGA program, and provides an option to browse reports, clinical analysis, copy number variation, mutation, expression, and to download data for further analysis (Figure 15) See the tutorial for a complete view of possibilities (http://firebrowse.org/tutorial/FireBrowse-Tutorial.pdf).

VIII-5 GDC (https://gdc.nci.nih.gov/)

The NCI's Genomic Data Commons (GDC) provides the cancer research community with a unified data repository that enables data sharing across cancer genomic studies in support of precision medicine. (note added in proof, June 6, 2016). The GDC supports several cancer genome programs at the NCI Center for Cancer Genomics (CCG), including The Cancer Genome Atlas (TCGA), Therapeutically Applicable Research to Generate Effective Treatments (TARGET), and the Cancer Genome Characterization Initiative (CGCI). The GDC Data Portal provides a platform for efficiently querying and downloading high quality and complete data. The GDC also provides a GDC Data Transfer Tool and a GDC API for programmatic access.

IX- Impact on diseases

IX-1 OMIM

"Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive and X-linked Phenotypes" was first published in 1966 by Victor A. McKusick (Johns Hopkins University Press), after a catalog of X-linked traits, published in 1962. In parallel, the "Human Gene Mapping" was first organized in New Haven in 1973, and mapped 119 and 100 loci respectively to confirmed or provisional/tentative chromosome assignements (Birth Defect, 1974).

The first edition of the "Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive and X-linked Phenotypes" had 1487 entries and no mapped autosomal loci. Victor A. McKusick published 12 editions, the last one in 1998, of his catalog. "Online Mendelian Inheritance in Man" (OMIM, http://omim.org/) was consequently published online. OMIM is a continuously updated catalog of human genes and genetic disorders and traits, with particular focus on the molecular relationship between genetic variation and phenotypic expression (Amberger JS et al., 2015).

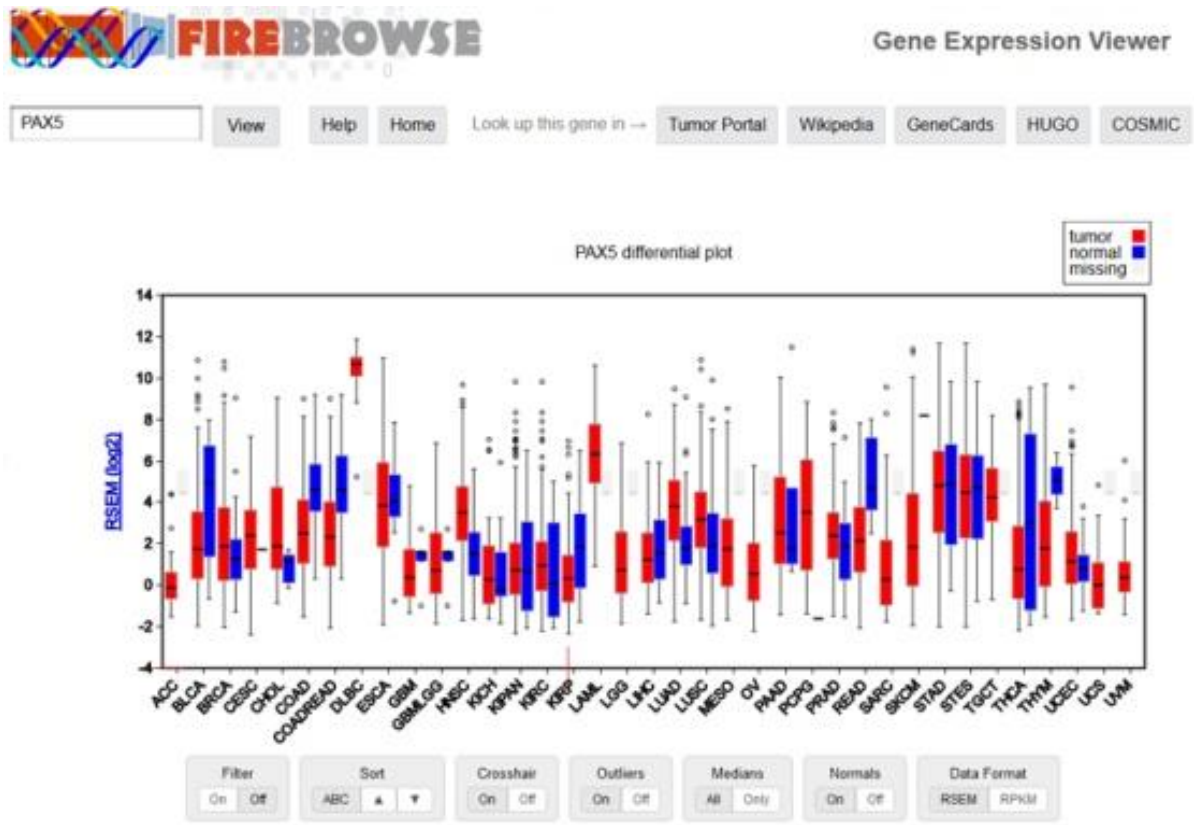


Figure 15: PAX5 expression at FireBrowse (http://firebrowse.org/viewGene.html?gene=PAX5).

As of April 2016, it consists of 23,460 entries: 15,237 gene descriptions, 4,705 phenotypes with known molecular basis, an additional 1,626 phenotypes with unknown molecular basis, and 1892 other entries. Gene entries start at: * 100640. Aldehyde dehydrogenase 1 family, member A1; ALDH1A1 Cytogenetic location: 9q21.13, Genomic coordinates (GRCh38): 9:72,900,661-72,953,316, and ends with * 616906. Cancer susceptibility candidate 1; CASC1. Phenotypes with known molecular basis entries start at: # 100100. Prune belly syndrome; PBS, Cytogenetic location: 1q43, and ends with # 616903. Nucleoside diphosphate-linked moiety X motif 15 deficiency; NUDT15D. OMIM describes somatic mutations in genes (11,139 entries for the term "mutation"). It is a very well curated database, with excellent reliability. Unfortunately the addition process of data as literature is published, by successive layerings/sedimentation makes it sometimes a laborious consultation. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine.

IX-2 ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/intro/)

ClinVar is designed to provide a public archive of reports of the relationships among human variations and phenotypes, with supporting evidence. By doing so, ClinVar facilitates access to and communication about the relationships asserted between human variation and observed health status, and the history of that interpretation. ClinVar collects reports of variants found in patient samples, and assertions made regarding their clinical significance. The alleles described in submissions are mapped to reference sequences, and reported according to the HGVS standard. ClinVar then presents the data for interactive users in daily workflow and other local applications. ClinVar works in collaboration with interested organizations to meet the needs of the medical genetics community as efficiently and effectively as possible (Harrison SM et al., 2016).

IX-3 MedGen (http://www.ncbi.nlm.nih.gov/medgen/)

MedGen is an NCBI portal of information about human disorders and other phenotypes having a genetic component. MedGen is structured to serve health care professionals, medical genetics community and other interested parties by providing centralized access to diverse content. MedGen aggregates the plethora of terms used for particular disorders into a specific concept, providing a "Rosetta stone" for stakeholders who may use different names for the same disorder.

Maintaining a clearly defined set of concepts and terms for phenotypes is essential in supporting characterization of genetic variation by its specific phenotypes effect. The assignment of identifiers for those concepts allows computational access to phenotypic information, an essential requirement for the large-scale analysis of genomic data. (NCBI Resource Coordinators, 2016).

IX-4 dbGaP (http://www.ncbi.nlm.nih.gov/dbgap/)

The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the data and results from studies where the interaction of genotype and phenotype in Humans has been investigated.

IX-5 SNPs3D (http://www.snps3d.org/)

SNPs3D is a website which assigns molecular functional effects of non-synonymous SNPs based on structure and sequence analysis. The site presents a data mining method to infer candidate SNP for 16 types of cancer (e.g. more than 1,000 genes potentially implicated in breast cancer: http://www.snps3d.org/modules.php?name=Candidate&disease=BREAST%20CANCER) (Yue P and Moult J, 2006).

IX-6 GTR (http://www.ncbi.nlm.nih.gov/gtr/)

The Genetic Testing Registry (GTR) provides a central location for voluntary submission of genetic test information by providers. The scope includes purpose of the test, methodology, validity, evidence of its usefulness and laboratory contacts and credentials. The overarching goal of the GTR is to advance public health research to include the genetic basis of health and disease (Rubinstein WS et al., 2013).

IX-7 ClinGen (https://www.clinicalgenome.org/)

ClinGen is a National Institutes of Health (NIH)-funded resource dedicated to building an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research. This resource has several goals for building a genomic knowledge base to improve patients care.

X- Pathology

X-1 Authoritative books in pathology are the following:

The "Rosai and Ackerman's Surgical Pathology" was first published in 1953. The tenth edition was published in 2011 by Elsevier, and contains 2892 pages. It includes clinical features, morphologic, immunohistochemical and molecular genetic features and prognosis, with a very large iconography. "WHO/IARC Classification of Tumours series" (http://publications.iarc.fr/Book-And-Report-Series/Who-Iarc-Classification-Of-Tumours) is not on free access, except editions prior to 2006, which are on free access in pdf format. The Armed Forces Institute of Pathology (AFIP) publishes series of the "AFIP Atlas of Tumor Pathology".

The WHO/OMS code, the ICD-O3, is not used by all databases (e.g. the Mitelman database or the COSMIC database have their own classification system, with no apparent matching). This is an obstacle for the integration of data by new resources.

X-2 Atlas of Genetics and Cytogenetics in Oncology and Haematology

The Atlas provides complete description of diseases, with papers similar to those found in the "Rosai and Ackerman's Surgical Pathology" or the "WHO/IARC Classification of Tumours series" (see above) with following restrictions: on the one hand, many tumor types are still missing from the Atlas; on the other hand, articles on genes closely related to these diseases are found, right next, in the Atlas, but not in the Rosai nor in the WHO's books, since this is out of their purpose.

X-3 PathologyOutlines (http://pathologyoutlines.com/)

PathologyOutlines provides information to practicing pathologists, with gross and microscopic images and summaries on CD markers and immunohistochemical stains and molecular markers.

X-4 The United States and Canadian Academy of Pathology (USCAP, http://www.uscap.org/)

USCAP is a provider of continuing medical education (CME) for pathologists to improve their practices. The Virtual Slide Box (http://uscapknowledgehub.org/index.htm?vsbindex.htm) and Juan Rosai's collection (http://rosaicollection.org/) are collections of several hundred slides with case reports and diagnoses.

XI- Cancer Registries

Cancer registries are organizations seeking to collect, store, analyze, and report data on various cancers for epidemiological purposes, for providing statistics on the occurrence of cancer in a defined population, and for obtaining a framework to assess the impact of cancer in a given population. Cancer registries are crucial for healthcare policy planning. They are key data source for clinical research, (epidemiology, study of carcinogens, evaluation of

treatments), providing the assessment of the care structures and care pathways, and research tools for social sciences and humanities (see https://www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/CancerEpi-17.pdf)

XI-1 International Agency for Research on Cancer (IARC, http://www.iarc.fr/)

The IARC is the outcome of an initiative by a group of leading French public figures; it was created on 20 May 1965, by a resolution of the World Health Assembly (http://www.iarc.fr/en/about/iarc-history.php). IARC is the specialized cancer agency of the World Health Organization (WHO/OMS). The objective of the IARC is to promote international collaboration in cancer research. The Agency is inter-disciplinary. Emphasis is placed on elucidating the role of environmental and lifestyle risk factors and studying their interplay with genetic background. IARC publishes the "Cancer Incidence in Five Continents" series and GLOBOCAN (Figure 16). The aim of the GLOBOCAN project (http://globocan.iarc.fr/Default.aspx) is to provide contemporary estimates of the incidences of, mortality and prevalence of major types of cancer, at national level, for 184 countries of the world.

XI-2 International Association of Cancer Registries (IACR, http://www.iacr.com.fr/).

The IACR (not to be confused with the IARC) was founded in 1966 as a professional society dedicated to fostering the aims and activities of cancer registries. It is a non-governmental organization which has held official relation with the World Health Organization since January 1979. With IACR IARC has developed with CanReg5, an open source tool to input, store, check and analyze cancer registry data. IACR has developed classifications (the successive editions of the International Classification of Diseases for Oncology, published by WHO), guidelines for registry practices and standard definitions. quality control, consistency checks and basic analysis of data, making data comparable between registries.

XI-3 Examples: The European Network of Cancer Registries (ENCR, http://www.encr.eu/)

has the same role in Europe as IACR has worldwide. The National Program of Cancer Registries (NPC, http://www.cdc.gov/cancer/), maintained by the Centers for disease control and prevention (CDC), collects data on cancer occurrence (including the type, extent, and location of the cancer), the type of initial treatment, and outcomes in the USA. The Surveillance, Epidemiology, and End Results (SEER, http://seer.cancer.gov/) program of the National Cancer Institute provides information on cancer. Research is supported by grants from the SEER. Quality improvement is another part of the SEER activities and it is dedicated to improving data quality by performing rigorous quality control studies and various data assessments. Union for International Cancer Control (UICC, http://www.uicc.org/). Founded in 1933, UICC brings together 900 organisations (cancer societies, ministries of health, research institutes and patient groups) over across 155 countries.

XII- Patient associations and interfaces between science and patients - freely accessible services

XII-1 Associations of parents and friends of patients

These associations of parents of patients with a rare disease are precious. They give moral support and help, and offer practical guidances and information about social benefits, subsidies and day-to-day life to families affected by illness. They often establish a program of grants for research (e.g. Xeroderma Pigmentosum Society (http://www.xps.org/, Sarcoma Foundation of America (http://www.curesarcoma.org/), Union for International Cancer Control (UICC) (http://www.uicc.org/)).

XII-2 interfaces between science and patients

GeneTests (https://www.genetests.org)

GeneTests (Pagon RA, 2006) provides information for 4,552 disorders (https://www.genetests.org/disorders/), 5,385 genes (https://www.genetests.org/genes/) (e.g. RUNX1 RUNX1 and contacts with/for: 79,009 laboratory tests (https://www.genetests.org/tests/), 680 laboratories (https://www.genetests.org/laboratories/) and 1,067 clinics (https://www.genetests.org/clinics/).

NORD (http://rarediseases.org)

NORD provides information on more than 1,300 rare diseases (e.g. Carcinoid syndrome http://rarediseases.org/rare-diseases/carcinoid-syndrome/), state health insurance information, guides for physicians, and patient assistance programs. They also provide grants to academic scientists for translational or clinical studies to help patients obtain life-saving or life-sustaining medication they could not otherwise afford.

Orphanet (http://www.orpha.net/)

Orphanet (Rath et al., 2012) offers an inventory of rare diseases with data on 5,833 diseases (e.g. Fanconi anaemia), an inventory of orphan drugs, list of 6,636 expert centres and 3,280 laboratories,

19,894 professionals for genetic counselling and medical management. Orphanet does not provide gene annotations. They hold a large partnership from 38 countries participating in the Orphanet consortium. They maintain large disease registries in Europe.

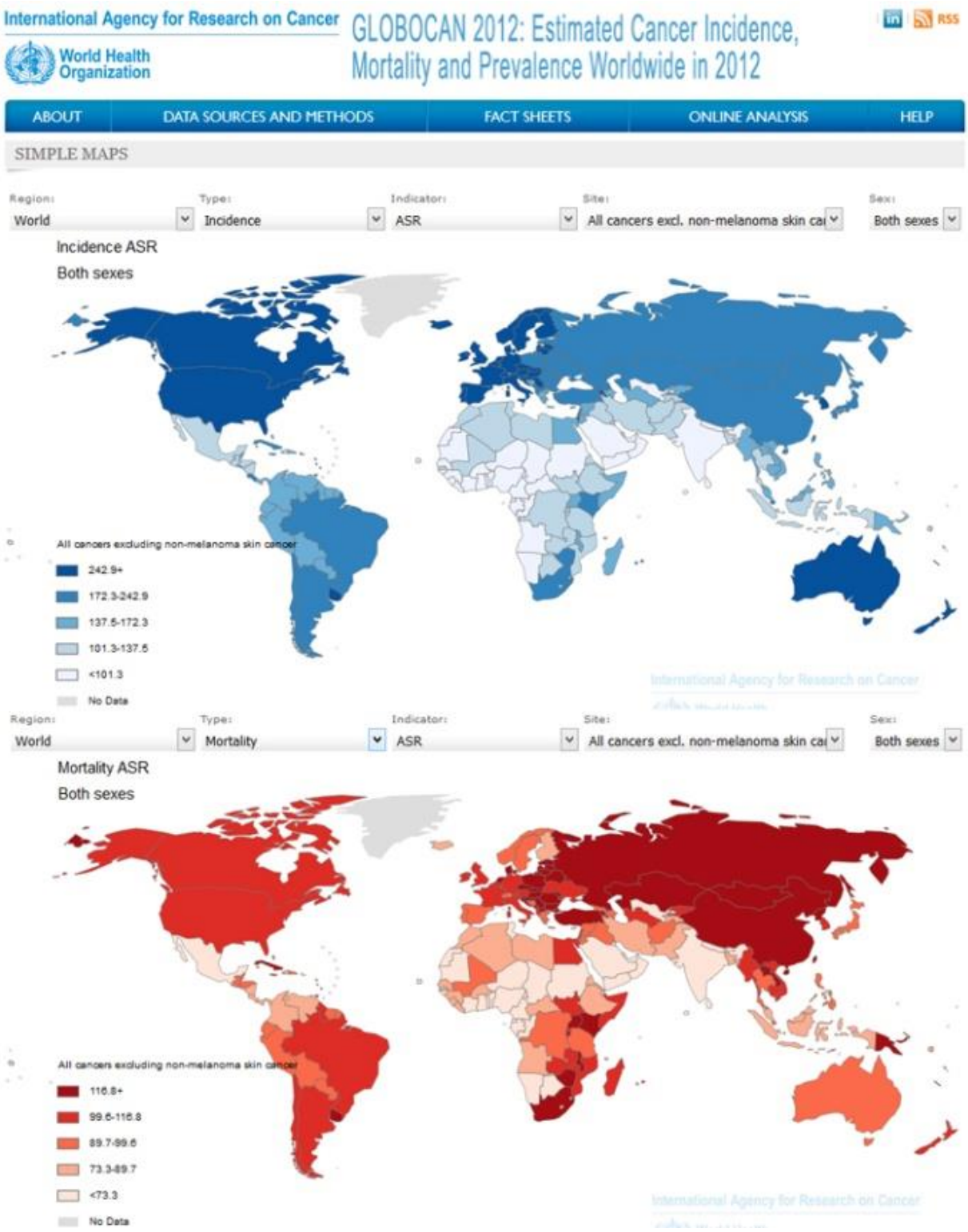


Figure 16: Surveillance system for cancers: GLOBOCAN 2012. Top: Incidence; Bottom: Mortality. ((http://globocan.iarc.fr/Default.aspx)

References

Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009 Apr 9;458(7239):719-24

Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer. 2015 Jun;15(6):371-81

Boveri T.. Zur Frage der Enstehung maligner Tumoren 1914 Gustav Fischer

Nowell PC, Hungerford DA. A minute Chromosome in Human Chronic Ganulocytic Leukemia Science 1960 132:1497

Caspersson T, Zech L, Modest EJ. Fluorescent labeling of chromosomal DNA: superiority of quinacrine mustard to quinacrine Science 1970 Nov 13;170(3959):762

Rowley JD. Identificaton of a translocation with quinacrine fluorescence in a patient with acute leukemia Ann Genet 1973 Jun;16(2):109-12

de Klein A, van Kessel AG, Grosveld G, Bartram CR, Hagemeijer A, Bootsma D, Spurr NK, Heisterkamp N, Groffen J, Stephenson JR. A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukaemia Nature 1982 Dec 23;300(5894):765-7

Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining Nature 1973 Jun 1;243(5405):290-3

Zech L, Haglund U, Nilsson K, Klein G. Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas Int J Cancer 1976 Jan 15;17(1):47-56

Berger R, Bernheim A, Weh HJ, Flandrin G, Daniel MT, Brouet JC, Colbert N. A new translocation in Burkitt's tumor cells Hum Genet 1979;53(1):111-2

Miyoshi I, Hiraki S, Kimura I, Miyamoto K, Sato J. 2/8 translocation in a Japanese Burkitt's lymphoma Experientia 1979 Jun 15;35(6):742-3

Van Den Berghe H, Gosseye CP, Englebienne V, Cornu G, Sokal G. Variant translocation in Burkitt lymphoma Cancer Genetics and Cytogenetics 1960, 1; 9-14

Oshimura M, Freeman AI, Sandberg AA. Chromosomes and causation of human cancer and leukemia XXVI Binding studies in acute lymphoblastic leukemia (ALL)

Rowley JD, Golomb HM, Dougherty C. 15/17 translocation, a consistent chromosomal change in acute promyelocytic leukaemia Lancet 1977 Mar 5;1(8010):549-50

Fukuhara S, Rowley JD, Variakojis D, Golomb HM. Chromosome abnormalities in poorly differentiated lymphocytic lymphoma Cancer Res 1979 Aug;39(8):3119-28

Ohno S, Babonits M, Wiener F, Spira J, Klein G, Potter M. Nonrandom chromosome changes involving the Ig gene-carrying chromosomes 12 and 6 in pristane-induced mouse plasmacytomas Cell 1979 Dec;18(4):1001-7

Seidal T, Mark J, Hagmar B, Angervall L. Alveolar rhabdomyosarcoma: a cytogenetic and correlated cytological and histological study Acta Pathol Microbiol Immunol Scand A 1982 Sep;90(5):345-54

Aurias A, Rimbaut C, Buffe D, Dubousset J, Mazabraud A. [Translocation of chromosome 22 in Ewing's sarcoma] C R Seances Acad Sci III 1983;296(23):1105-7

Turc-Carel C, Philip I, Berger MP, Philip T, Lenoir G. [Chromosomal translocation (11; 22) in cell lines of Ewing's sarcoma] C R Seances Acad Sci III 1983;296(23):1101-3

de Jong B, Molenaar IM, Leeuw JA, Idenberg VJ, Oosterhuis JW. Cytogenetics of a renal adenocarcinoma in a 2-year-old child Cancer Genet Cytogenet 1986 Mar 15;21(2):165-9

Stenman G, Sandros J, Dahlenfors R, Juberg-Ode M, Mark J. 6q- and loss of the Y chromosome--two common deviations in malignant human salivary gland tumors Cancer Genet Cytogenet 1986 Aug;22(4):283-93

Mark J, Dahlenfors R, Ekedahl C, Stenman G. The mixed salivary gland tumor — A normally benign human neoplasm frequently showing specific chromosomal abnormalities. Cancer Genetics and Cytogenetics 1980 2, 231-24

Heim S, Mandahl N, Kristoffersson U, Mitelman F, Rser B, Rydholm A, Willn H. Reciprocal translocation t(3;12)(q27;q13) in lipoma Cancer Genet Cytogenet 1986 Dec;23(4):301-4

Turc-Carel C, Dal Cin P, Rao U, Karakousis C, Sandberg AA. Cytogenetic studies of adipose tissue tumors I A benign lipoma with reciprocal translocation t(3;12)(q28;q14)

Bernard O, Lecointe N, Jonveaux P, Souyri M, Mauchauff M, Berger R, Larsen CJ, Mathieu-Mahul D. Two site-specific deletions and t(1;14) translocation restricted to human T-cell acute leukemias disrupt the 5' part of the tal-1 gene Oncogene 1991 Aug;6(8):1477-88

Barr FG, Nauta LE, Davis RJ, Schfer BW, Nycum LM, Biegel JA. In vivo amplification of the PAX3-FKHR and PAX7-FKHR fusion genes in alveolar rhabdomyosarcoma Hum Mol Genet 1996 Jan;5(1):15-21

Simon MP, Pedeutour F, Sirvent N, Grosgeorge J, Minoletti F, Coindre JM, Terrier-Lacombe MJ, Mandahl N, Craver RD, Blin N, Sozzi G, Turc-Carel C, O'Brien KP, Kedra D, Fransson I, Guilbaud C, Dumanski JP. Deregulation of the platelet-derived growth factor B-chain gene via fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans and giant-cell fibroblastoma Nat Genet 1997 Jan;15(1):95-8

Sinclair PB, Nacheva EP, Leversha M, Telford N, Chang J, Reid A, Bench A, Champion K, Huntly B, Green AR. Large deletions at the t(9;22) breakpoint are common and may identify a poor-prognosis subgroup of patients with chronic myeloid leukemia Blood 2000 Feb 1;95(3):738-43

Mller E, Hornick JL, Magnusson L, Veerla S, Domanski HA, Mertens F. FUS-CREB3L2/L1-positive sarcomas show a specific gene expression profile with upregulation of CD24 and FOXL1 Clin Cancer Res 2011 May 1;17(9):2646-56

Gelsi-Boyer V, Trouplin V, Adélïde J, Aceto N, Remy V, Pinson S, Houdayer C, Arnoulet C, Sainty D, Bentires-Alj M, Olschwang S, Vey N, Mozziconacci MJ, Birnbaum D, Chaffanet M. Genome profiling of chronic myelomonocytic leukemia: frequent alterations of RAS and RUNX1 genes BMC Cancer 2008 Oct 16;8:299

Van Vlierberghe P, van Grotel M, Tchinda J, Lee C,

Beverloo HB, van der Spek PJ, Stubbs A, Cools J, Nagata K, Fornerod M, Buijs-Gladdines J, Horstmann M, van Wering ER, Soulier J, Pieters R, Meijerink JP. The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia Blood 2008 May 1;111(9):4668-80

Mullighan CG, Collins-Underwood JR, Phillips LA, Loudin MG, Liu W, Zhang J, Ma J, Coustan-Smith E, Harvey RC, Willman CL, Mikhail FM, Meyer J, Carroll AJ, Williams RT, Cheng J, Heerema NA, Basso G, Pession A, Pui CH, Raimondi SC, Hunger SP, Downing JR, Carroll WL, Rabin KR. Rearrangement of CRLF2 in B-progenitor- and Down syndrome-associated acute lymphoblastic leukemia Nat Genet 2009 Nov;41(11):1243-6

Santo EE, Ebus ME, Koster J, Schulte JH, Lakeman A, van Sluis P, Vermeulen J, Gisselsson D, ra I, Lindner S, Buckley PG, Stallings RL, Vandesompele J, Eggert A, Caron HN, Versteeg R, Molenaar JJ. Oncogenic activation of FOXR1 by 11q23 intrachromosomal deletion-fusions in neuroblastoma Oncogene 2012 Mar 22;31(12):1571-81

Paszczyca A, Nilsson J, Magnusson L, Brosj O, Larsson O, Vult von Steyern F, Domanski HA, Lilljebjrn H, Fioretos T, Tayebwa J, Mandahl N, Nord KH, Mertens F. Fusions involving protein kinase C and membrane-associated proteins in benign fibrous histiocytoma Int J Biochem Cell Biol 2014 Aug;53:475-81

De Braekeleer E, Douet-Guilbert N, Morel F, Le Bris MJ, Meyer C, Marschalek R, Frec C, De Braekeleer M. FLNA, a new partner gene fused to MLL in a patient with acute myelomonoblastic leukaemia Br J Haematol 2009 Sep;146(6):693-5

Meyer C, Hofmann J, Burmeister T, Grger D, Park TS, Emerenciano M, Pombo de Oliveira M, Renneville A, Villarese P, Macintyre E, Cav H, Clappier E, Mass-Malo K, Zuna J, Trka J, De Braekeleer E, De Braekeleer M, Oh SH, Tsaur G, Fechina L, van der Velden VH, van Dongen JJ, Delabesse E, Binato R, Silva ML, Kustanovich A, Aleinikova O, Harris MH, Lund-Aho T, Juvonen V, Heidenreich O, Vormoor J, Choi WW, Jarosova M, Kolenova A, Bueno C, Menendez P, Wehner S, Eckert C, Talmant P, Tondeur S, Lippert E, Launay E, Henry C, Ballerini P, Lapillone H, Callanan MB, Cayuela JM, Herbaux C, Cazzaniga G, Kakadiya PM, Bohlander S, Ahlmann M, Choi JR, Gameiro P, Lee DS, Krauter J, Cornillet-Lefebvre P, Te Kronnie G, Schfer BW, Kubetzko S, Alonso CN, zur Stadt U, Sutton R, Venn NC, Izraeli S, Trakhtenbrot L, Madsen HO, Archer P, Hancock J, Cerveira N, Teixeira MR, Lo Nigro L, Mricke A, Stanulla M, Schrappe M, Sedk L, Szczepaski T, Zwaan CM, Coenen EA, van den Heuvel-Eibrink MM, Strehl S, Dworzak M, Panzer-Grmayer R, Dingermann T, Klingebiel T, Marschalek R. The MLL recombinome of acute leukemias in 2013 Leukemia 2013 Nov;27(11):2165-76

Speicher MR, Carter NP. The new cytogenetics: blurring the boundaries with molecular biology Nat Rev Genet 2005 Oct;6(10):782-92

Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer Nat Genet 2005 Jun;37 Suppl:S11-7

De Braekeleer E, Douet-Guilbert N, De Braekeleer M. Genetic diagnosis in malignant hemopathies: from cytogenetics to next-generation sequencing Expert Rev Mol Diagn 2014 Mar;14(2):127-9

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer Science 2005 Oct 28;310(5748):644-8

West RB, Rubin BP, Miller MA, Subramanian S, Kaygusuz G, Montgomery K, Zhu S, Marinelli RJ, De Luca A, Downs-Kelly E, Goldblum JR, Corless CL, Brown PO, Gilks CB, Nielsen TO, Huntsman D, van de Rijn M. A landscape effect in tenosynovial giant-cell tumor from activation of CSF1 expression by a translocation in a minority of tumor cells Proc Natl Acad Sci U S A 2006 Jan 17;103(3):690-5

Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzel R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL, Polakiewicz RD, Rush J, Comb MJ. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer Cell 2007 Dec 14;131(6):1190-203

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer Nature 2007 Aug 2;448(7153):561-6

Wang L, Motoi T, Khanin R, Olshen A, Mertens F, Bridge J, Dal Cin P, Antonescu CR, Singer S, Hameed M, Bovee JV, Hogendoorn PC, Socci N, Ladanyi M. Identification of a novel, recurrent HEY1-NCOA2 fusion in mesenchymal chondrosarcoma based on a genome-wide screen of exon-level expression data Genes Chromosomes Cancer 2012 Feb;51(2):127-39

Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurles ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing Nat Genet 2008 Jun;40(6):722-9

Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM. Transcriptome sequencing to detect gene fusions in cancer Nature 2009 Mar 5;458(7234):97-101

Maher CA, Palanisamy N, Brenner JC, Cao X, Kalyana-Sundaram S, Luo S, Khrebtukova I, Barrette TR, Grasso C, Yu J, Lonigro RJ, Schroth G, Kumar-Sinha C, Chinnaiyan AM. Chimeric transcript discovery by paired-end transcriptome sequencing Proc Natl Acad Sci U S A 2009 Jul 28;106(30):12353-8

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerd A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Brresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR. Complex landscapes of somatic rearrangement in human breast cancer genomes Nature 2009 Dec 24;462(7276):1005-10

Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma Nature 2013 Jul 4;499(7456):43-9

Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers

Nature 2012 Sep 27;489(7417):519-25

Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma Nature 2014 Mar 20;507(7492):315-22

Cancer Genome Atlas Research Network, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, Yau C, Laird PW, Ding L, Zhang W, Mills GB, Kucherlapati R, Mardis ER, Levine DA. Integrated genomic characterization of endometrial carcinoma Nature 2013 May 2;497(7447):67-73

Steidl C, Shah SP, Woolcock BW, Rui L, Kawahara M, Farinha P, Johnson NA, Zhao Y, Telenius A, Neriah SB, McPherson A, Meissner B, Okoye UC, Diepstra A, van den Berg A, Sun M, Leung G, Jones SJ, Connors JM, Huntsman DG, Savage KJ, Rimsza LM, Horsman DE, Staudt LM, Steidl U, Marra MA, Gascoyne RD. MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers Nature 2011 Mar 17;471(7338):377-81

Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, Wallis J, Chen K, Payton JE, Fulton RS, Veizer J, Schmidt H, Vickery TL, Heath S, Watson MA, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Ley TJ, Wilson RK. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene JAMA 2011 Apr 20;305(15):1577-84

Roberts KG, Morin RD, Zhang J, Hirst M, Zhao Y, Su X, Chen SC, Payne-Turner D, Churchman ML, Harvey RC, Chen X, Kasap C, Yan C, Becksfort J, Finney RP, Teachey DT, Maude SL, Tse K, Moore R, Jones S, Mungall K, Birol I, Edmonson MN, Hu Y, Buetow KE, Chen IM, Carroll WL, Wei L, Ma J, Kleppe M, Levine RL, Garcia-Manero G, Larsen E, Shah NP, Devidas M, Reaman G, Smith M, Paugh SW, Evans WE, Grupp SA, Jeha S, Pui CH, Gerhard DS, Downing JR, Willman CL, Loh M, Hunger SP, Marra MA, Mullighan CG. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia Cancer Cell 2012 Aug 14;22(2):153-66

Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, Verhaak RG. The landscape and therapeutic relevance of cancer-associated transcript fusions Oncogene 2015 Sep 10;34(37):4845-54

Mitelman F, Johansson B, Merten sF. Mitelman database of chromosome aberrations and genes fusions in Cancer Mitelman F, Johansson B and Mertens F (Eds.) 2016, http://cgap.nci.nih.gov/Chromosomes/Mitelman

Huret JL, Ahmad M, Arsaban M, Bernheim A, Cigna J, Desangles F, Guignard JC, Jacquemot-Perbal MC, Labarussias M, Leberre V, Malo A, Morel-Pair C, Mossafa H, Potier JC, Texier G, Vigui F, Yau Chun Wan-Senon S, Zasadzinski A, Dessen P. Atlas of genetics and cytogenetics in oncology and haematology in 2013 Nucleic Acids Res 2013 Jan;41(Database issue):D920-4

Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation Nat Rev Cancer 2007 Apr;7(4):233-45

Kalyana-Sundaram S, Shankar S, Deroo S, Iyer MK, Palanisamy N, Chinnaiyan AM, Kumar-Sinha C. Gene fusions associated with recurrent amplicons represent a class of passenger aberrations in breast cancer Neoplasia 2012 Aug;14(8):702-8

Gingeras TR. Implications of chimaeric non-co-linear transcripts Nature 2009 Sep 10;461(7261):206-11

Rickman DS, Pflueger D, Moss B, VanDoren VE, Chen CX, de la Taille A, Kuefer R, Tewari AK, Setlur SR, Demichelis F, Rubin MA. SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer Cancer Res 2009 Apr 1;69(7):2734-8

Meyer C, Kowarz E, Hofmann J, Renneville A, Zuna J, Trka J, Ben Abdelali R, Macintyre E, De Braekeleer E, De Braekeleer M, Delabesse E, de Oliveira MP, Cav H, Clappier E, van Dongen JJ, Balgobind BV, van den Heuvel-Eibrink MM, Beverloo HB, Panzer-Grmayer R, Teigler-Schlegel A, Harbott J, Kjeldsen E, Schnittger S, Koehl U, Gruhn B, Heidenreich O, Chan LC, Yip SF, Krzywinski M, Eckert C, Mricke A, Schrappe M, Alonso CN, Schfer BW, Krauter J, Lee DA, Zur Stadt U, Te Kronnie G, Sutton R, Izraeli S, Trakhtenbrot L, Lo Nigro L, Tsaur G, Fechina L, Szczepanski T, Strehl S, Ilencikova D, Molkentin M, Burmeister T, Dingermann T, Klingebiel T, Marschalek R. New insights to the MLL recombinome of acute leukemias Leukemia 2009 Aug;23(8):1490-9

Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, Nordentoft I, Birkenkamp-Demtrder K, Kruhffer M, Hager H, Knudsen B, Andersen CL, Srensen KD, Pedersen JS, rntoft TF, Dyrskjt L. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue PLoS One 2014 May 30;9(5):e98187

Fletcher CD. The evolving classification of soft tissue tumours - an update based on the new 2013 WHO classification Histopathology 2014 Jan;64(1):2-11

Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD, Jaffe ES. The 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasms Blood 2016 Mar 15

Hokland P, Ommen HB. Towards individualized follow-up in adult acute myeloid leukemia in remission Blood 2011 Mar 3;117(9):2577-84

Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood Nat Rev Clin Oncol 2013 Aug;10(8):472-84

Karabacak NM, Spuhler PS, Fachin F, Lim EJ, Pai V, Ozkumur E, Martel JM, Kojic N, Smith K, Chen PI, Yang J, Hwang H, Morgan B, Trautwein J, Barber TA, Stott SL, Maheswaran S, Kapur R, Haber DA, Toner M. Microfluidic, marker-free isolation of circulating tumor cells from blood samples Nat Protoc 2014 Mar;9(3):694-710

Watanabe M, Serizawa M, Sawada T, Takeda K, Takahashi T, Yamamoto N, Koizumi F, Koh Y. A novel flow cytometry-based cell capture platform for the detection, capture and molecular characterization of rare tumor cells in blood J Transl Med 2014 May 23;12:143

Yu KH, Ricigliano M, Hidalgo M, Abou-Alfa GK, Lowery MA, Saltz LB, Crotty JF, Gary K, Cooper B, Lapidus R, Sadowska M, O'Reilly EM. Pharmacogenomic modeling of circulating tumor and invasive cells for prediction of chemotherapy response and resistance in pancreatic cancer Clin Cancer Res 2014 Oct 15;20(20):5281-9

Baccelli I, Schneeweiss A, Riethdorf S, Stenzinger A, Schillert A, Vogel V, Klein C, Saini M, Buerle T, Wallwiener M, Holland-Letz T, Hfner T, Sprick M, Scharpff M, Marm F, Sinn HP, Pantel K, Weichert W, Trumpp A. Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay Nat Biotechnol 2013 Jun;31(6):539-44

Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, Antipova A, Lee C, McKernan K, De La Vega FM, Kinzler KW, Vogelstein B, Diaz LA Jr, Velculescu VE. Development of personalized tumor biomarkers using massively parallel sequencing Sci Transl Med 2010 Feb 24;2(20):20ra14

Druker BJ, Sawyers CL, Kantarjian H, Resta DJ, Reese SF, Ford JM, Capdeville R, Talpaz M. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome N Engl J Med 2001 Apr 5;344(14):1038-42

Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S, Sawyers CL. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia N Engl J Med 2001 Apr 5;344(14):1031-7

Rutkowski P, Van Glabbeke M, Rankin CJ, Ruka W, Rubin BP, Debiec-Rychter M, Lazar A, Gelderblom H, Sciot R, Lopez-Terrada D, Hohenberger P, van Oosterom AT, Schuetze SM; European Organisation for Research and Treatment of Cancer Soft Tissue/Bone Sarcoma Group; Southwest Oncology Group. Imatinib mesylate in advanced dermatofibrosarcoma protuberans: pooled analysis of two phase II clinical trials J Clin Oncol 2010 Apr 1;28(10):1772-9

Joensuu H. Adjuvant treatment of GIST: patient selection and treatment strategies Nat Rev Clin Oncol 2012 Apr 24;9(6):351-8

Lee HJ, Thompson JE, Wang ES, Wetzler M. Philadelphia chromosome-positive acute lymphoblastic leukemia: current treatment and future perspectives Cancer 2011 Apr 15;117(8):1583-94

Kohno T, Tsuta K, Tsuchihara K, Nakaoku T, Yoh K, Goto K. RET fusion gene: translation to personalized lung cancer therapy Cancer Sci 2013 Nov;104(11):1396-400

Shaw AT, Hsu PP, Awad MM, Engelman JA. Tyrosine kinase gene rearrangements in epithelial malignancies Nat Rev Cancer 2013 Nov;13(11):772-87

Malik R, Khan AP, Asangani IA, Cielik M, Prensner JR, Wang X, Iyer MK, Jiang X, Borkin D, Escara-Wilke J, Stender R, Wu YM, Niknafs YS, Jing X, Qiao Y, Palanisamy N, Kunju LP, Krishnamurthy PM, Yocum AK, Mellacheruvu D, Nesvizhskii AI, Cao X, Dhanasekaran SM, Feng FY, Grembecka J, Cierpicki T, Chinnaiyan AM. Targeting the MLL complex in castration-resistant prostate cancer Nat Med 2015 Apr;21(4):344-52

Chen CW, Koche RP, Sinha AU, Deshpande AJ, Zhu N, Eng R, Doench JG, Xu H, Chu SH, Qi J, Wang X, Delaney C, Bernt KM, Root DE, Hahn WC, Bradner JE, Armstrong SA. DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia Nat Med 2015 Apr;21(4):335-43

Dawson MA, Prinjha RK, Dittmann A, Giotopoulos G, Bantscheff M, Chan WI, Robson SC, Chung CW, Hopf C, Savitski MM, Huthmacher C, Gudgin E, Lugo D, Beinke S, Chapman TD, Roberts EJ, Soden PE, Auger KR, Mirguet O, Doehner K, Delwel R, Burnett AK, Jeffrey P, Drewes G, Lee K, Huntly BJ, Kouzarides T. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia Nature 2011 Oct 2;478(7370):529-33

McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra A 3rd, Diaz E, LaFrance LV, Mellinger M, Duquenne C, Tian X, Kruger RG, McHugh CF, Brandt M, Miller WH, Dhanak D, Verma SK, Tummino PJ, Creasy CL. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations Nature 2012 Dec 6;492(7427):108-12

Fillmore CM, Xu C, Desai PT, Berry JM, Rowbotham SP, Lin YJ, Zhang H, Marquez VE, Hammerman PS, Wong KK, Kim CF. EZH2 inhibition sensitizes BRG1 and EGFR mutant lung tumours to TopoII inhibitors Nature 2015 Apr 9;520(7546):239-42

Bernheim A, Huret JL, Guillaud-Bataille M, Brison O, Couturiers J; Groupe Franais de Cytogntique Oncologique. [Cytogenetics, cytogenomics and cancer: 2004 update] Bull Cancer 2004 Jan;91(1):29-43

BEADLE GW. Genetics and metabolism in Neurospora Physiol Rev 1945 Oct;25:643-63

Burks C, Fickett JW, Goad WB, Kanehisa M, Lewitter FI, Rindone WP, Swindell CD, Tung CS, Bilofsky HS. The GenBank nucleic acid sequence database Comput Appl Biosci 1985 Dec;1(4):225-33

Burks C, Cassidy M, Cinkosky MJ, Cumella KE, Gilna P, Hayden JE, Keen GM, Kelley TA, Kelly M, Kristofferson D, et al. GenBank Nucleic Acids Res 1991 Apr 25;19 Suppl:2221-5

Benton D. Recent changes in the GenBank On-line Service Nucleic Acids Res 1990 Mar 25;18(6):1517-20

Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank Nucleic Acids Res 2015 Jan;43(Database issue):D30-5

Cook CE, Bergman MT, Finn RD, Cochrane G, Birney E, Apweiler R. The European Bioinformatics Institute in 2016: Data growth and integration Nucleic Acids Res 2016 Jan 4;44(D1):D20-6

Pundir S, Magrane M, Martin MJ, O'Donovan C; UniProt Consortium. Searching and Navigating UniProt Databases Curr Protoc Bioinformatics 2015 Jun 19;50:1

Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames org: the HGNC resources in 2015 Nucleic Acids Res

NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information Nucleic Acids Res 2016 Jan 4;44(D1):D7-19

Fishilevich S, Zimmerman S, Kohn A, Iny Stein T, Olender T, Kolker E, Safran M, Lancet D. Genic insights from integrated human proteomics in GeneCards Database (Oxford) 2016 Apr 5;2016

Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, Haussler D, Zhu J. The UCSC Cancer Genomics Browser: update 2015 Nucleic Acids Res 2015 Jan;43(Database issue):D812-7

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Girn CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Keenan S, Lavidas I, Martin FJ, Maurel T, McLaren W, Murphy DN, Nag R, Nuhn M, Parker A, Patricio M, Pignatelli M, Rahtz M, Riat HS, Sheppard D, Taylor K, Thormann A, Vullo A, Wilder SP, Zadissa A, Birney E, Harrow J, Muffato M, Perry E, Ruffier M, Spudich G, Trevanion SJ, Cunningham F, Aken BL, Zerbino DR, Flicek P. Ensembl 2016 Nucleic Acids Res 2016 Jan 4;44(D1):D710-6

Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma Nature 2011 Jun 29;474(7353):609-15

Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B, Wong-Erasmus M,

Yao L, Kasprzyk A. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data Database (Oxford) 2011 Sep 19;2011:bar026

Chin L, Hahn WC, Getz G, Meyerson M.. Making sense of cancer genomic data. Genes Dev. 2011 Mar 15;25(6):534-55. doi: 10.1101/gad.2017311.

Pavlopoulou A, Spandidos DA, Michalopoulos I. Human cancer databases (review) Oncol Rep 2015 Jan;33(1):3-18

Klonowska K, Czubak K, Wojciechowska M, Handschuh L, Zmienko A, Figlerowicz M, Dams-Kozlowska H, Kozlowski P. Oncogenomic portals for the visualization and analysis of genome-wide cancer data Oncotarget 2016 Jan 5;7(1):176-92

Brookes AJ, Robinson PN. Human genotype-phenotype databases: aims, challenges and opportunities Nat Rev Genet 2015 Dec;16(12):702-15

Yang Y, Dong X, Xie B, Ding N, Chen J, Li Y, Zhang Q, Qu H, Fang X. Databases and web tools for cancer genomics study Genomics Proteomics Bioinformatics 2015 Feb;13(1):46-50

Niroula A, Vihinen M. Variation Interpretation Predictors: Principles, Types, Performance, and Choice Hum Mutat 2016 Jun;37(6):579-97

Diehl AG, Boyle AP. Deciphering ENCODE Trends Genet 2016 Apr;32(4):238-49

Wu J, Wu M, Li L, Liu Z, Zeng W, Jiang R. dbWGFP: a database and web server of human whole-genome single nucleotide variants and their functional predictions Database (Oxford) 2016 Mar 17;2016

Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells Science 2015 Sep 25;349(6255):1483-9

Sverre Heim and Felix Mitelman. Cancer Cytogenetics: Chromosomal and Molecular Genetic Abberations of Tumor Cells 2015, Wiley-Blackwell , New-York

Dorkeld F, Bernheim A, Dessen P, Huret JL. A database on cytogenetics in haematology and oncology Nucleic Acids Res 1999 Jan 1;27(1):353-4

A PROPOSED standard system of nomenclature of human mitotic chromosomes. Lancet 1960 May 14;1(7133):1063-5 PubMed PMID: 13857542

Shaffer LG, McGowen-Jordan J, Schmid M, editors. An International System for Human Cytogenetic Nomenclature 2013, Basel: S. Karger

Mitelman F, Johansson B, Mertens F. Mitelman database of chromosome aberrations and genes fusions in Cancer

Kim N, Kim P, Nam S, Shin S, Lee S. ChimerDB--a knowledgebase for fusion sequences Nucleic Acids Res 2006 Jan 1;34(Database issue):D21-4

Kim P, Yoon S, Kim N, Lee S, Ko M, Lee H, Kang H, Kim J, Lee S. ChimerDB 2 0--a knowledgebase for fusion genes updated Nucleic Acids Res

Novo FJ, de Mend IO, Vizmanos JL. TICdb: a collection of gene-mapped translocation breakpoints in cancer BMC Genomics 2007 Jan 26;8:33

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders Nucleic Acids Res

Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U, Campbell PJ. COSMIC: exploring the world's knowledge of somatic mutations in human cancer Nucleic Acids Res 2015 Jan;43(Database issue):D805-11

Frenkel-Morgenstern M, Gorohovski A, Lacroix V, Rogers M, Ibanez K, Boullosa C, Andres Leon E, Ben-Hur A, Valencia A. ChiTaRS: a database of human, mouse and fruit fly chimeric transcripts and RNA-sequencing data Nucleic Acids Res 2013 Jan;41(Database issue):D142-51

Frenkel-Morgenstern M, Gorohovski A, Vucenovic D, Maestre L, Valencia A. ChiTaRS 2 1--an improved database of the chimeric transcripts and RNA-seq data with novel sense-antisense chimeric RNA transcripts Nucleic Acids Res

Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, Degenhardt J, Mayba O, Gnad F, Liu J, Pau G, Reeder J, Cao Y, Mukhyala K, Selvaraj SK, Yu M, Zynda GJ, Brauer MJ, Wu TD, Gentleman RC, Manning G, Yauch RL, Bourgon R, Stokoe D, Modrusan Z, Neve RM, de Sauvage FJ, Settleman J, Seshagiri S, Zhang Z. A comprehensive transcriptional portrait of human cancer cell lines Nat Biotechnol 2015 Mar;33(3):306-12

Wang Y, Wu N, Liu J, Wu Z, Dong D. FusionCancer: a database of cancer fusion genes derived from RNA-seq data Diagn Pathol 2015 Jul 28;10:131

Lvf M, Thomassen GO, Bakken AC, Celestino R, Fioretos T, Lind GE, Lothe RA, Skotheim RI. Fusion gene microarray reveals cancer type-specificity among fusion genes Genes Chromosomes Cancer 2011 May;50(5):348-57

Skotheim RI, Thomassen GO, Eken M, Lind GE, Micci F, Ribeiro FR, Cerveira N, Teixeira MR, Heim S, Rognes T, Lothe RA. A universal assay for detection of oncogenic fusion transcripts by oligo microarray analysis Mol Cancer 2009 Jan 19;8:5

Urakami K, Shimoda Y, Ohshima K, Nagashima T, Serizawa M, Tanabe T, Saito J, Usui T, Watanabe Y, Naruoka A, Ohnami S, Ohnami S, Mochizuki T, Kusuhara M, Yamaguchi K. Next generation sequencing approach for detecting 491 fusion genes from human cancer Biomed Res 2016;37(1):51-62

Babiceanu M, Qin F, Xie Z, Jia Y, Lopez K, Janus N, Facemire L, Kumar S, Pang Y, Qi Y, Lazar IM, Li H. Recurrent chimeric fusion RNAs in non-cancer tissues and cells Nucleic Acids Res 2016 Apr 7;44(6):2859-72

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors Science 1992 Oct 30;258(5083):818-21

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dhner H, Cremer T, Lichter P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances Genes Chromosomes Cancer 1997 Dec;20(4):399-407

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays Nat Genet 1998 Oct;20(2):207-11

Commo F, Fert C, Soria JC, Friend SH, Andr F, Guinney J. Impact of centralization on aCGH-based genomic profiles

for precision medicine in oncology Ann Oncol 2015 Mar;26(3):582-8

Clough E, Barrett T. The Gene Expression Omnibus Database Methods Mol Biol 2016;1418:93-110

Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, Fllgrabe A, Fuentes AM, Jupp S, Koskinen S, Mannion O, Huerta L, Megy K, Snow C, Williams E, Barzine M, Hastings E, Weisser H, Wright J, Jaiswal P, Huber W, Choudhary J, Parkinson HE, Brazma A. Expression Atlas update--an integrated database of gene and protein expression in humans, animals and plants Nucleic Acids Res 2016 Jan 4;44(D1):D746-52

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M. The landscape of somatic copy-number alteration across human cancers Nature 2010 Feb 18;463(7283):899-905

Kim TM, Xi R, Luquette LJ, Park RW, Johnson MD, Park PJ. Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes Genome Res 2013 Feb;23(2):217-27

Cao Q, Zhou M, Wang X, Meyer CA, Zhang Y, Chen Z, Li C, Liu XS. CaSNP: a database for interrogating copy number alterations of cancer genome from SNP array data Nucleic Acids Res 2011 Jan;39(Database issue):D968-74

Cai H, Gupta S, Rath P, Ai N, Baudis M. arrayMap 2014: an updated cancer genome resource Nucleic Acids Res 2015 Jan;43(Database issue):D825-30

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome Nat Genet 2004 Sep;36(9):949-51

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzlez JR, Gratacs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Global variation in copy number in the human genome Nature 2006 Nov 23;444(7118):444-54

MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome Nucleic Acids Res 2014 Jan;42(Database issue):D986-92

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources Am J Hum Genet 2009 Apr;84(4):524-33

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation Nature 2015 Oct 1;526(7571):68-74

International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PI, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurles M, Korn JM, Kristiansson K, Lee C, McCarrol SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghori MJ, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations Nature 2010 Sep 2;467(7311):52-8

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM; Broad GO; Seattle GO; NHLBI Exome Sequencing Project. Evolution and functional impact of rare coding variation from deep sequencing of human exomes Science 2012 Jul 6;337(6090):64-9

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes Nat Rev Cancer 2004 Mar;4(3):177-83

Cooper DN, Krawczak M. Human Gene Mutation Database Hum Genet 1996 Nov;98(5):629

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v. 2.0: the next generation in gene variant databases. Hum Mutat. May;32(5):557-63

Deng M, Brgelmann J, Schultze JL, Perner S. Web-TCGA: an online platform for integrated analysis of molecular cancer data sets BMC Bioinformatics 2016 Feb 6;17:72

Gundem G, Perez-Llamas C, Jene-Sanz A, Kedzierska A, Islam A, Deu-Pons J, Furney SJ, Lopez-Bigas N. IntOGen: integration and data mining of multidimensional oncogenomic data Nat Methods 2010 Feb;7(2):92-3

Wu TJ, Shamsaddini A, Pan Y, Smith K, Crichton DJ, Simonyan V, Mazumder R. A framework for organizing cancer-related variations from existing databases, publications and NGS data using a High-performance Integrated Virtual Environment (HIVE) Database (Oxford) 2014 Mar 25;2014:bau022

Minikel EV, Vallabh SM, Lek M, Estrada K, Samocha KE, Sathirapongsasuti JF, McLean CY, Tung JY, Yu LP, Gambetti P, Blevins J, Zhang S, Cohen Y, Chen W, Yamada M, Hamaguchi T, Sanjo N, Mizusawa H, Nakamura Y, Kitamoto T, Collins SJ, Boyd A, Will RG, Knight R, Ponto C, Zerr I, Kraus TF, Eigenbrod S, Giese A, Calero M, de Pedro-Cuesta J, Hak S, Laplanche JL, Bouaziz-Amar E, Brandel JP, Capellari S, Parchi P, Poleggi A, Ladogana A, O'Donnell-Luria AH, Karczewski KJ, Marshall JL, Boehnke M, Laakso M, Mohlke KL, Khler A, Chambert K, McCarroll S, Sullivan PF, Hultman CM, Purcell SM, Sklar P, van der Lee SJ, Rozemuller A,

Jansen C, Hofman A, Kraaij R, van Rooij JG, Ikram MA, Uitterlinden AG, van Duijn CM; Exome Aggregation Consortium (ExAC), Daly MJ, MacArthur DG. Quantifying prion disease penetrance using large population control cohorts Sci Transl Med 2016 Jan 20;8(322):322ra9

. Birth Defects Cytogenet Cell Genet. 1974;13(3):1-216

Harrison SM, Riggs ER, Maglott DR, Lee JM, Azzariti DR, Niehaus A, Ramos EM, Martin CL, Landrum MJ, Rehm HL. Using ClinVar as a Resource to Support Variant Interpretation Curr Protoc Hum Genet 2016 Apr 1;89:8

Yue P, Moult J. Identification and analysis of deleterious human SNPs J Mol Biol 2006 Mar 10;356(5):1263-74

Rubinstein WS, Maglott DR, Lee JM, Kattman BL, Malheiro AJ, Ovetsky M, Hem V, Gorelenkov V, Song G, Wallin C, Husain N, Chitipiralla S, Katz KS, Hoffman D, Jang W, Johnson M, Karmanov F, Ukrainchik A, Denisenko M, Fomous C, Hudson K, Ostell JM. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency Nucleic Acids Res 2013 Jan;41(Database issue):D925-35

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation Nucleic Acids Res 2016 Jan 4;44(D1):D733-45

Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, Dreszer TR, Fujita PA, Guruvadoo L, Haeussler M, Harte RA, Heitner S, Hickey G, Hinrichs AS, Hubley R, Karolchik D, Learned K, Lee BT, Li CH, Miga KH, Nguyen N, Paten B, Raney BJ, Smit AF, Speir ML, Zweig AS, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2015 update Nucleic Acids Res 2015 Jan;43(Database issue):D670-81

Wu C, Jin X, Tsueng G, Afrasiabi C, Su AI. BioGPS: building your own mash-up of gene annotations and expression profiles Nucleic Acids Res 2016 Jan 4;44(D1):D313-6

UniProt Consortium. UniProt: a hub for protein information Nucleic Acids Res 2015 Jan;43(Database issue):D204-12

Gaudet P, Michel PA, Zahn-Zabal M, Cusin I, Duek PD, Evalet O, Gateau A, Gleizes A, Pereira M, Teixeira D, Zhang Y, Lane L, Bairoch A. The neXtProt knowledgebase on human proteins: current status Nucleic Acids Res 2015 Jan;43(Database issue):D764-70

Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations Nucleic Acids Res 2015 Jan;43(Database issue):D512-20

Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. New and continuing developments at PROSITE Nucleic Acids Res 2013 Jan;41(Database issue):D344-7

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future Nucleic Acids Res 2016 Jan 4;44(D1):D279-85

Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJ, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD. The InterPro protein families database: the classification resource after 15 years Nucleic Acids Res 2015 Jan;43(Database issue):D213-21

Pagon RA. GeneTests: an online genetic information resource for health care providers J Med Libr Assoc 2006 Jul;94(3):343-8

Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users Hum Mutat 2012 May;33(5):803-8

Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, Reddy J, Borges-Rivera D, Lee TI, Jaenisch R, Porteus MH, Dekker J, Young RA. Activation of proto-oncogenes by disruption of chromosome neighborhoods Science 2016 Mar 25;351(6280):1454-8

Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations Proc Natl Acad Sci U S A 2016 Apr 19;113(16):E2326-34

Lawler M, Siu LL, Rehm HL, Chanock SJ, Alterovitz G, Burn J, Calvo F, Lacombe D, Teh BT, North KN, Sawyers CL; Clinical Working Group of the Global Alliance for Genomics and Health (GA4GH). All the World's a Stage: Facilitating Discovery Science and Improved Cancer Care through the Global Alliance for Genomics and Health Cancer Discov 2015 Nov;5(11):1133-6

This article should be referenced as such:

De Braekeleer E, Huret JL, Mossafa H, Hautaviita K, Dessen P. General resources in Genetics and/or Oncology. Atlas Genet Cytogenet Oncol Haematol. 2016; 20(5):289-315.