

Deep Insight Section

Internet databases and resources for cytogenetics and cytogenomics

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/Internet_DatabasesID20143.html

Printable original version : http://documents.irevues.inist.fr/bitstream/handle/2042/62675/04-2015-Internet_DatabasesID20143.pdf

DOI: 10.4267/2042/62675

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 2.0 France Licence.

© 2017 *Atlas of Genetics and Cytogenetics in Oncology and Haematology*

Abstract

Databases devoted *stricto sensu* to cancer cytogenetics are the "Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer" (<http://cgap.nci.nih.gov/Chromosomes/Mitelman>), the "Atlas of Genetics and Cytogenetics in Oncology and Haematology" (<http://atlasgeneticsoncology.org>), and COSMIC (<http://cancer.sanger.ac.uk/cosmic>). However, cancer being a complex multi-step process, cytogenetics are broadened to "cytogenomics", with complementary resources, including resources on proteins and cancer. These resources are essential to both practical and theoretical knowledge in cytogenomics of cancer. Must be briefly reviewed: general databases (nucleic acid and protein sequences databases and bibliographic ones), cancer genomic portals associated to recent international integrated programs, such as TCGA or ICGC, fusion genes databases, genomic sequences and transcripts databases (with different cartography browsers), array CGH databases and structural variation databases for copy number, polymorphisms and mutation databases, databases on proteins (structure and function with implication of mutations and rearrangements), databases on diseases, databases and books on pathology, cancers, and patient associations and interfaces between science and patients. Other resources such as the International System for Human Cytogenetic Nomenclature (ISCN), the International Classification of Diseases for Oncology (ICD-O), the Human Gene Nomenclature Database (HGNC), and the Nomenclature for the description of sequence variations allow a common language. Data within the scientific/medical community should be freely available. However, most of the institutional stakeholders are now gradually disengaging, and well known databases are forced to beg or to disappear (which may happen!).

Keywords: Cytogenetic, Cancer, Database, Mitelman Database, Atlas of Genetics and Cytogenetics in Oncology and Haematology, COSMIC, PubMed, GenBank, TCGA, ICGC, UniProt, OMIM, IARC, ISCN, ICD-O, HGNC.

INTRODUCTION

In each cancer case there is a genetic event present (Stratton MR et al., 2009). Cytogenetics has been a major player in understanding genetics behind cancer, providing specific keys for diagnosis and prognostic assessments, as well as enabling the sub-classification of otherwise seemingly identical

disease entities (Mertens F et al., 2015). This "Deep Insight" is dedicated to cytogenetics resources will highlight the various facets used in the current strategies in theoretical understanding of cancer and the consequent practical strategies in treating the disease.

Brief history

In 1914, Boveri stated that the heritable acquired characteristics of cancer cells are brought about by a disturbance of the normal chromosomal balance (Boveri, 1914). This theory was supported by a wealth of experimental data showing that cancer originates in a single cell through acquired genetic changes. The investigation in the 1950s, on ascites tumors that were induced experimentally or observed in patients tended to confirm that cytogenetic aberrations are an important and integral part of tumor development and evolution. These cytogenetic studies demonstrated that certain laws could direct neoplasia-associated chromosomal variability. Like selective pressures, where any changes in the surrounding tumor would modify the equilibrium, causing a change where the most viable chromosomal profile is prevailing in the new environment.

The importance of cytogenetics boomed since the discovery of the first chromosomal anomaly reported by Peter Nowell and David Hungerford in 1960, linking the Philadelphia (Ph) chromosome to chronic myeloid leukaemia (CML) (Nowell and Hungerford, 1960). It was the first assessment in detecting chromosomal anomaly in human leukaemia and seemed reasonable that it was the cause of origin for CML. This discovery was the first strong assessment to Boveri's theory. This observation stimulated the field to find other karyotypic anomalies in other cancers. Unfortunately, a heterogeneous panel of chromosomal rearrangements was detected in what seemed to be the same cancer. This was a terrible setback for the arguments stating karyotypic anomalies as the origin of cancer. The explanation was that chromosomal rearrangements were an epiphenomenon that could appear during tumor progression without having any pathogenetic consequences.

In the 1970's the situation changed dramatically when chromosomal banding techniques invented by Caspersson and Zech (Caspersson T et al., 1970) were introduced. This process gave an option to identify individual chromosomes, which were defined by a unique banding pattern. The description of chromosomal rearrangements immediately became clearer providing more gravity to the conclusions drawn. This was a new era for cancer cytogenetics showing an increase in the numbers of aberrant human malignant and benign karyotypes.

In the 1980s, the onset of molecular genetics techniques intensely widened our understanding of the pathogenetic progression underlying the neoplastic process. These techniques provided an opportunity to characterise the chromosomal breakpoints at the molecular level and has highlighted two classes of genes implicated in these karyotypical rearrangements: the oncogenes and the tumor suppressing genes.

MYC and BCR/ABL1: One of the first oncogenes described as activated by chromosomal rearrangement is MYC, which was characterised in Burkitt lymphoma studies.

Another example is the translocation between ABL1 and BCR. Peter Nowell and David Hungerford first described a recurrent presence of an extra-chromosome in CML patients in 1960 (Nowell PCH and Hungerford DA, 1960). In 1973 Janet D. Rowley used quinacrine coloration to prove this chromosome to be the result of a translocation between chromosomes 9 and 22 (Rowley JD, 1973a). Only as late as 1982 de Klein et al. showed that the genes ABL1 and BCR were fused together giving rise to an abnormal gene (de Klein A et al., 1982). With these new techniques, each chromosome and chromosome region could be identified on the basis of their unique banding pattern, giving daylight to previously undetectable subtle rearrangements. By this technique she identified the recurrent translocation t(8;21)(q22;q22) (Rowley JD, 1973b). These findings evoked interest in the cytogenetic analysis of other haematological malignancies. The number of reported balanced rearrangements has increased, in particular translocations including t(8;14)(q24;q32), t(2;8)(p11;q24) and t(8;22)(q24;q11) in Burkitt lymphoma (Zech L et al., 1976 ; Berger R et al., 1979 ; Miyoshi I et al., 1979 ; Van Den Berghe H et al., 1979), t(4;11)(q21;q23) in acute lymphoblastic leukaemia (ALL) (Oshimura M et al., 1977) t(15;17)(q22;q21) in acute promyelocytic leukaemia (APL) (Rowley JD et al., 1977), and t(14;18)(q32;q21) in follicular lymphoma (Rowley JD et al., 1977). During this fruitful period the first specific translocation in an animal model was found, a mouse plasmacytoma, which is a B cell malignancy displaying similar characteristics to human Burkitt lymphomas (Ohno S et al., 1979).

The following decade witnessed a rise in number of results from malignant solid tumors, mainly sarcomas but also a few carcinomas. Several of the aberrations identified were as specific as the ones that were previously described in haematological cancer: t(2;13)(q36;q14) in alveolar rhabdomyosarcoma (ARMS) (Seidal T et al., 1982), t(11;22)(q24;q12) in Ewing sarcoma (Aurias A et al., 1983 ; Turc-Carel C et al., 1983), t(X;1)(p11;q21) in Kidney cancer (de Jong B et al., 1986) and t(6;9)(q23;p23) in alivary gland tumors (ACC) of the salivary glands (Stenman G et al., 1986). Evidence was showing that many benign tumors were bearing characteristic rearrangements, including reciprocal translocations such as t(3;8)(p21;q12) in salivary gland adenoma (SGA) (Mark J et al., 1980) and t(3;12)(q27;q13) in lipoma (Heim S et al., 1986 ; Turc-Carel C et al., 1986).

Although the vast majority of fusion genes are formed by balanced translocations, they can also be produced by interstitial deletions. These were first

identified in the 1990s, amongst them the fusion between genes STIL (STIL/TAL1 interrupting locus) in T-ALL (Bernard O et al., 1991). Since then, many others were observed with more or less extensive deletions, duplications and/or amplifications in the breakpoint regions (Barr FG et al., 1996 ; Simon MP et al., 1997 ; Sinclair PB et al., 2000 ; Müller E et al., 2011). Gene fusion can also arise from copy number shifts like in the aforementioned fusion gene USP16/RUNX1 (ubiquitin specific peptidase 16 and runt related transcription factor 1) in chronic myelomonocytic leukemia (Gelsi-Boyer V et al., 2008) and in the fusion gene SET/NUP214 (SET nuclear proto-oncogene and nuclear pore complex protein Nup214) in T-ALL (Van Vlierbergh P et al., 2008 ; Mullighan CG et al., 2009 ; Santo EE et al., 2012 ; Plaszczyca A et al., 2014) (Figure 1).

Technical developments

In the late 1970s, various technical developments helped in solving what molecular consequences the oncogenic chromosomal rearrangements could have. These techniques enabled the identification and characterisation of genes that were located at the breakpoints of chromosomal rearrangements. The genes implicated in MPC, Burkitt lymphoma and CML proved to be pivotal for the comprehension of the mechanism underlying chromosomal

rearrangements. The engineering of fluorescence in-situ hybridization (FISH) enabled several chromosomal structures to be identified simultaneously. This significantly improved the location of breakpoints on chromosomes. It also considerably reduced the scale of which chromosomes could be observed and broadened the type of rearrangements that could be observed (cryptic rearrangements). The big advantage of the FISH technique is that it can also be used for non-dividing cells (interphase nuclei). FISH probes of a specific gene can identify new partner genes, like in the case of mixed lineage leukemia (MLL, KMT2A) gene (De Braekeleer E et al., 2009; Meyer C et al., 2013).

Although cytogenetic analyses are unquestionably crucial for the identification of fusion genes and rearrangements, there are certain limits to this technique. Firstly, revealing chromosome bandings requires having access to in-vitro living, dividing cells so that metaphases can be observed. Secondly, some tumor types can have very complex genomes which makes it difficult to understand the full story and distinguish the primary aberrations and origin of the cancer development from the bulk of the rearrangements (Speicher MR and Carter NP, 2005).

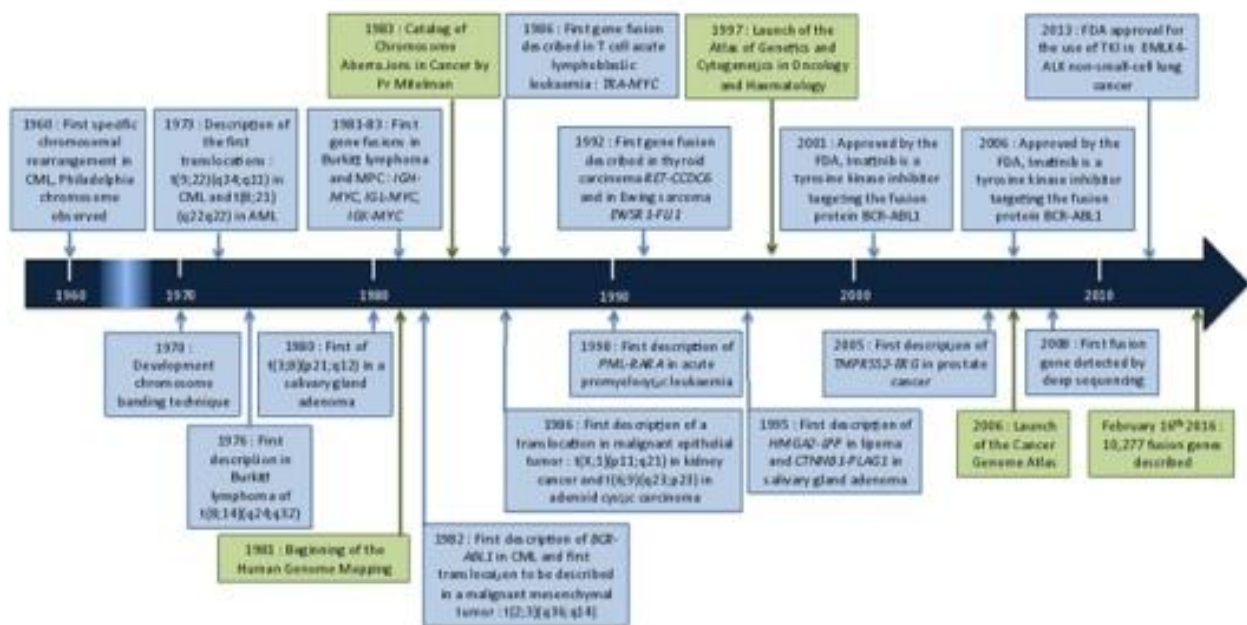


Figure 1: Timeline of important discoveries concerning fusion genes, chromosomal rearrangements and the establishment of databases regrouping all these chromosomal abnormalities.

In the 1990s, the progress of high throughput tools for global genetic analyses, such as array based platforms for gene expression and copy number profiling, gave rise to new methods for observing

chromosomal rearrangements. These techniques were not ideal either since balanced chromosomal rearrangements could pass undetected or the analysis of expression profiles could prove to be tricky. On

the other hand, they presented a higher level of resolution than in chromosome banding and didn't require prior cell culturing (Pinkel D and Albertson DG, 2005; De Braekeleer E et al., 2014). The first novel gene fusion detected with the analysis of gene expression pattern of a tumor was the fusion of the transcription factor PAX3 gene with the nuclear receptor co-activator 1, NCOA1 gene. By focusing on genes presenting outlined values of expression, the fusions genes implicating the transmembrane protease serine 2 gene (TMPRSS2) with two genes encoding ETS transcription factors. The first is v-ets avian erythroblastosis virus E26 oncogene homolog (ERG) and the second is ets variant 1 (ETV1) (Tomlins SA et al., 2005). It was the first report of specific fusion genes implicated and representing a major subset of a common epithelial malignancy. By using a modification of this method, other fusion genes were discovered in many tumor types, such as tenosynovial giant cell tumor, lung cancer and chondrosarcoma (West RB et al., 2006; Rikova K et al., 2007; Soda M et al., 2007; Wang L et al., 2012). The introduction of deep sequencing technologies a few years ago gave a new insight to identify new fusions genes either at DNA or RNA level. The combination of detailed information (base pair level) and broad (genome-wide) on DNA, transcriptome, structural variants and fusion transcripts could be obtained without any prior information on the cytogenetic features of the cancer cells. The initial study using deep sequencing to detect fusion genes or chromosomal rearrangements were done on established cell lines (Campbell PJ et al., 2008). The analysis of primary samples from common cancer (Maher CA et al., 2009a; Maher CA et al., 2009b), such as carcinomas of the breast (Stephens PJ et al., 2009), colon (Cancer Genome Atlas Research Network, 2013), lung (Cancer Genome Atlas Research Network, 2012), prostate (Cancer Genome Atlas Research Network, 2014), uterus (Cancer Genome Atlas Research Network et al., 2013) as well as leukaemias and lymphomas (Steidl C et al., 2011; Welch JS et al., 2011; Roberts KG et al., 2012), came afterward. One study draws a bridge between over several several cancers by cumulating the bioinformatics data of 4,366 cancers from 13 different tumor types that were previously studied within the Cancer Genome Atlas (TCGA) network. The outcome was the description of 8,600 different fusion transcripts (Yoshihara K et al., 2015). These results have dramatically changed the gene fusion landscape with the identification of more than 10,000 fusion genes with more than 90% of these having been identified by various deep-sequencing approaches during the last 5 years (Mitelman F et al., 2016; Huret JL et al., 2013). The high resolution of deep sequencing gave the possibility to identify the vast majority of genes implicated in chromosomal rearrangements that

would have been complicated or impossible to identify by conventional cytogenetic techniques. Indeed, 75% of the genes fusions first detected by deep sequencing are intrachromosomal and approximately 50% are between genes located in the same chromosome band (Mitelman F et al., 2016). Large majority of genes, - already described in the literature before the deep sequencing era- were embedded in extensive networks like MLL in leukaemias, EWS RNA-binding protein 1 (EWSR1) in sarcomas and rearranged during Transfection Protooncogene (RET) in carcinomas (Mitelman F et al., 2007). However, this picture has somewhat changed with the massive increase of fusion genes that were added with genome-wide studies. The fact that these studies were mainly focusing on previously uncharacterized tumor types brought a lot of new networks emerging from rarer gene fusions than leukaemias, lymphomas and sarcomas. Furthermore, carcinomas often show highly rearranged genomes, with numerous mutations at the gene and chromosome levels and it may be that the genes detected by deep sequencing are the results of chance events caused by chromosomal instability, as vast majority of fusion transcripts were associated with amplification or deletion events at the DNA level (Yoshihara K et al., 2015; Mitelman H et al., 2016; Huret JL et al., 2013; Mitelman F et al., 2007; Kalyana-Sundaram S et al., 2012). Transcription-induced gene fusion (TIGF) or Trans-TIGF, when they happen on different chromosomes, results in the fusion of transcripts from non-adjacent genes without a corresponding fusion at DNA level (Gingeras TR, 2009; Rickman DS et al., 2009; Meyer C et al., 2009; Hedegaard J et al., 2014). Certain have been shown to have no impact, since they were expressed in normal tissues like the fusion genes JAZF zinc finger 1 (JAZF1)/SUZ12 a polycomb repressive complex 2 subunit and PAX3/. Others, implicating the gene MLL, to be the driving mutation (Meyer C et al., 2009).

The prognostic and treatment value of chromosomal rearrangements and mutated genes: The high correlation between recurrent gene fusions and tumor subtypes has made them the ideal maker for diagnostic purposes. This correlation is also important in treatment stratification, the best example being the different fusion of MLL in AML (Meyer C et al., 2009). The routine molecular strategy to detect these fusion genes is the use of cytogenetics, FISH, RT-PCR and deep sequencing. The mounting knowledge of the clinical importance of gene fusions, as well as various chromosomal rearrangements, has gradually led to an increasing emphasis on genetic features in the classification of tumors. The latest World Health Organisation (WHO) classification, translocation and/or gene fusion status is mandatory for the diagnosis of some types of tumors, such as "AML with

t(8;21)(q22;q22), RUNX1/RUNX1T1" and "B lymphoblastic leukaemia/lymphoma with t(5;14)(q31;q32), IL3/IGH". For other cancers, such as alveolar soft part sarcoma and synovial sarcoma, it is considered as a distinctive defining element of the neoplasm (Fletcher CD, 2014; Swerdlow SH et al., 2016). Since fusion genes are diagnostic markers, they can also be used as markers for monitoring minimal residual disease following treatment (De Braekeleer E et al., 2014 ; Hokland P, Ommen HB and Hokland P, 2011). Currently, this strategy is in clinical use mainly for haematological disorders but the improvements in the detection and enrichment of circulating cancer cells and DNA suggest that solid tumors with gene fusions might also be monitored in a similar way (Crowley E et al., 2013; Karabacak NM et al., 2014; Watanabe M et al., 2014; Yu KH et al., 2014; Baccelli I et al., 2013). It is important to mention that the detection of the fusion gene can be used to monitor the progression or the relapse of the cancerous cells but it doesn't need to be an important actor in the neoplastic phenomenon, as long as they are representative and specific of the neoplastic cells (Leary RJ et al., 2010).

Research on fusion genes paved the way to develop specific drugs targeting chimeric proteins. The tyrosine kinase inhibitor Imatinib, approved in 2001, was the first drug specifically designed to target the chimeric protein BCR/ABL1 in CML (Druker BJ et al., 2001; Druker BJ et al., 2001) by blocking its kinase activity. This drug dramatically improved the lifespan and life quality of patients bearing CML. The immense success of imatinib spurred interest in developing new compounds against the chimeric proteins, all of which are kinase inhibitors. Different tumors have shown to display various fusions involving kinase-encoding genes, such as ALK, BRAF, Fibroblast growth factor receptor 3 (FGFR3), neurotrophic tyrosine kinase receptor type 1 (NTRK1), RET and ROS1 (Yoshihara K et al., 2015; ; Huret JL et al., 2013; Kohno T et al., 2013 ; Shaw AT et al., 2013). These fusion genes are occurring at low frequencies but if merged they represent a considerable number of patients. Stratification strategies considering the genotype and phenotype of the tumor would contribute greatly to identifying patients with these very promising treatment targets. Many new compounds are currently being tested in clinical models although others have reached the Phase 1 and Phase 2 stages in clinical trials, for example, chromatin modifier such as MLL (MEN1 (Malik R et al., 2015), DOT1L (Chen CW et al., 2015), BRD4 (Dawson MA et al., 2011) or EZH2 (McCabe MT et al., 2012; Fillmore CM et al., 2015).

The need for organising data banks

Since discovering their involvement in cancer initiation, progression and evolution,

chromosomal rearrangements have triggered wide, increasing interest in understanding them better. The amount of genes involved has increased, the network underlying certain genes has been resolved and the mechanistic aspect is unravelled. Unfortunately, a lot of work has to be done before cancer has been eradicated. One of the steps is to synthesise all the information and make it available in order to increase the common knowledge of genes that are implicated and their interactions with other pathways in the cell. The importance of creating data banks and reporting various chromosomal rearrangements has been recognised since the 80's.

1981: Human Genome Mapping

The information on chromosome modification in cancer has been included as part of the Human Genome Mapping (HGM) workshop since 1981. The provision of up-to-date information of all chromosomal rearrangements was the initial goal. This means including all case reports, which are suspected to be the starting point of tumor development or a contribution to the proliferation but also complex karyotypes with several cytogenetic anomalies or secondary modifications leading to the evolution and resistance to treatment. The increasing number of cases, reports and the multitude of cytogenetically abnormal neoplasms made it too challenging to include everything in the database. In 1991, the HGM decided to focus only on aberrations repeatedly found as sole anomalies in a few given tumor types. As a consequence the number of recurrent changes was severely underestimated, especially in solid tumors where single anomalies are a rare finding. This illustration of chromosomal anomalies mainly represents the tip of the iceberg since the generalisation and improvements in classic cytogenetic techniques and the development of new techniques have considerably increased the number of reports of chromosomal rearrangements in different types of tumors. Several of these anomalies may be of diagnostic and prognostic importance, as well as a large amount of details of molecular analysis.

1983: Catalog of Chromosome Aberrations in Cancer

In 1983, Felix Mitelman published a colossal manuscript that was a supplement to Cytogenetics and Cell Genetics. The goal of this publication was to catalogue all known chromosomal rearrangements. The complexity of the data pushed the laboratories and institutes to adopt computerised methods to compile, revise and index the information. Many cytogeneticist, clinicians and cell biologists were in the demand for a systematic, concise and uniform presentation of material. The vast body of literature was making it complicated to evaluate if a chromosomal abnormality had been described before or not. To facilitate this process,

Mitelman presented a compilation of 3,844 published and unpublished cases from colleagues or from his own laboratory. The two volumes presented all the implicated genes, chromosomes and rearrangements known. This set of two books was the first of its kind but far from the last. For several years these two volumes accompanied the bookshelves of several cytogenetists and oncologists. A re-edition of this work took place in 1985 with data of new cases and improved data of cases already described. The number of cases had now increased to a bit more than 5,000. The number of cases increased with each edition so that by the fifth edition it was composed of two large volumes of more than 4,000 pages, making it arduous to use. The sixth edition had already more than 30,000 cases in it. To make it more user friendly it was then published as a CD. The number of cases would still continue to increase and this information was not freely available. Felix Mitelman then had the idea to display the information on the Internet, rendering it freely available. In 2000, the catalogue became accessible for the public under the name Mitelman Database of Chromosome Aberrations in Cancer associated to the Cancer Genome Anatomy Project internet site and under the supervision of the National Cancer Institute (see below).

1997: Atlas of Genetics and Cytogenetics in Oncology and Haematology

How did the idea of the Atlas come about? Prognosis for leukaemia depends on the genes involved: 5 years survival rate: 6% in the inv(3)(q21q26) RPN1/MECOM leukemia, 100% in the dic(9;12)(p13;p13) PAX5/ETV6 leukemia. Treatment depends on the severity of the disease. However, thousands of genes were discovered to be implicated in cancer (14,000 unique fusion transcripts have been detected), and 1,200 types of solid tumors exist. Some cancers are frequent while many others are very rare (many with only 1 published case). This is particularly true for leukemia subtypes of which there are more than 1,000! 25,000 new publications concerning human cancer genetics are added each year to PubMed. No-one has the whole required knowledge, necessary to guide the treatment procedure in case of a rare disease. The following conclusion was made that huge databases were required to collect and summarize data on these rare diseases in order to produce meta-analyses. The Atlas has been established for that reason; to contribute to 'meta-medicine', meaning the mediation between the knowledge and the knowledge users in medicine. Besides resources dedicated only to cytogenetics, a quick overview of resources in surrounding areas "Cancer Cytogenetics", stricto sensu, deals with chromosomes and cancer. "Cytogenetics" means "Cell Genetics" ("cyto" comes from $\kappa\upsilon\tau\omicron\varsigma$, in the meaning of the term "the cell"); "Cytogenomics", as

coined by Alain Bernheim, (Bernheim A et al., 2004) (from a princeps paper in French in 1998), means the "genetics -as a whole- of the cell", with complex interconnections and interactions between these operators. As is known for long, "one-gene-one-reaction" (Beadle GW, 1945) (understood today as "one-gene-one-protein"), and we can infer from "Cyto-genomics" to the terms "Cyto-transcriptome" and "Cyto-proteomics", or, in a more holistic approach, (and more simply) "Cell Biology".

Cancer is now known as being a multi-step process, with genetic events at almost each step. Therefore, the "Cancer Cytogenetics" research field should incorporate knowledge of the "Cell Biology" of normal and cancerous cells, gene fusions, mutations or copy number variation, epigenetics, protein domains, metabolic or signaling pathways, as well as consequences of these cytogenomic rearrangements and disorders in the pathogenesis of cancer, from gross and microscopic pathological presentation to patients and diseases, clinical pictures, and, even, to epidemiological data given by cancer registries.

It is useful for the cancer cytogeneticists to have an easy and quick access to databases and books of these surrounding subject areas. Therefore, besides resources of cancer cytogenetics, we will mention other resources, including resources on proteins and resources on cancer.

Presently, Internet provides access to a vast and complex network of knowledge that can make it challenging for you to find the answer to your questions. Several databases are freely accessible, but unfortunately not all of them are user friendly. We will briefly describe the main resources in the following pages.

Recent reviews on cancer databases

In complement, and not to duplicate good recent publications in the last months, some reviews on cancer databases list most of the Internet resources in the general field of cancer genomics. A review of L. Chin gives an overview of the current state of cancer genomics (L. Chin et al., 2011). Regardless of a wide spectrum of references, the topic of cytogenetic resources is absent (Pavlopoulou A et al., 2015 ; Klonowska K et al., 2016 ; Brookes AJ, Robinson PN and Brookes AJ, 2015 ; Yang Y et al., 2015 ; Niroula A and Vihinen M, 2016 ; Diehl AG and Boyle AP, 2016; Martincorena I, et al. 2015). There are also many descriptions of database (and particularly in cancer) in all special issues of Nucleic Acid Research (each year in January).

2. GENERAL RESOURCES

Note: a detailed description of General resources in Genetics and/or Oncology may be found at http://atlasgeneticsoncology.org/Deep/General_ResourcesID20144.html

I- Bibliography

PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) is a widely used and free search engine and database of

biomedical citations and abstracts, based essentially on the MEDLINE database of references on life sciences and biomedical topics. Medline is the U.S. National Library of Medicine (NLM) premier bibliographic database. PubMed Central (<http://www.ncbi.nlm.nih.gov/pmc/>) is an archive of biomedical and life sciences journal literature. 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. Scopus (<http://www.scopus.com/>) is a database owned by Elsevier.

II- Nomenclatures

Gene Nomenclature: The HUGO Gene Nomenclature Committee (HGNC, <http://www.genenames.org/>) is the authority that assigns standardised nomenclature to human genes. Nomenclature for the description of sequence variations (<http://www.hgvs.org/mutnomen/>) is maintained by the Human Genome Variation Society (HGVS). International System for Human Cytogenetic Nomenclature (ISCN): The ISCN is the language used to describe abnormal karyotypes. International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3): The WHO/OMS has established a code, which provides a topographical (organ) identifier and an identifier for the detailed pathology.

III- Nucleic acid, genes and protein databases

Nucleic acid databases: GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) is a DNA sequence database. The need to have (in parallel to the genome projects) the best representation of genomic and transcript sequences (for diverse species) has been at the origin of consensus databases (as RefSeq, UCSC, Ensembl) with several methods of optimisation. Genomic sequences and transcripts: RefSeq

(<http://www.ncbi.nlm.nih.gov/refseq/>) maintains and curates a database of annotated genomic, transcript, and protein sequence records. Ensembl (<http://www.ensembl.org/>) developed a software which produces and maintains automatic annotation on selected eukaryotic genomes. The UCSC Genome Browser database (see above) is a large collection containing genome assemblies of various species. Proteins: In addition to the amino acid sequence, protein name and description with domains, these databases may provide a brief annotation information, others are only computationally analysed. These databases are the following: UniProt (<http://www.uniprot.org/>), a hub consisting of two sections: "TrEMBL" and "Swiss-Prot"; neXtProt (<http://www.nextprot.org/db/>); PhosphoSitePlus (<http://www.phosphosite.org/homeAction.action>), an excellent resource providing comprehensive information and tools for the study of protein post-

translational modifications; PROSITE (<http://prosite.expasy.org/>) Pfam (<http://pfam.xfam.org/>) and InterPro (<http://www.ebi.ac.uk/interpro/>). The Atlas of Genetics and Cytogenetics in Oncology and Haematology presents highly curated paragraphs with the description of the protein, but on a restricted sample.

IV- Cards

Entrez Gene (<http://www.ncbi.nlm.nih.gov/gene/>) is NCBI's primary text search and retrieval system that integrates the PubMed database and molecular databases including DNA and protein sequence, structure, gene, genome, genetic variation and gene expression. Genecards (<http://www.genecards.org/>) is a database that provides information on all annotated and predicted human genes.

V- Genome cartography

The cartography of genes on a genome has always been a fundamental mean of representation of 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: The UCSC Genome Browser website (<http://genome.ucsc.edu/>) contains the reference sequence for a large collection of genomes. The Genome Browser zooms and scrolls over chromosomes, "Blat" quickly maps a sequence to the genome. The UCSC Cancer Browser <https://genome-cancer.ucsc.edu/proj/site/help/>) allows researchers to interactively explore cancer genomics data and its associated clinical information. Ensembl (<http://www.ensembl.org>) generates genomic datasets and distributes created datasets and promote standards and interoperability between genomic resources.

VI- Structural variation databases

Genomic structural variation (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. Several databases have integrated data produced in the literature on copy number variation of DNA sequences: dbVar (<http://www.ncbi.nlm.nih.gov/dbvar/>), DGV - Genomic Variants (<http://dgv.tcag.ca/dgv/app/home>), DECIPHER (<https://decipher.sanger.ac.uk/>) and 1000 Genomes (<http://www.1000genomes.org/>).

VII- Polymorphism databases

It is important to distinguish polymorphisms due to 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. Polymorphism databases are: dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/overview.html>), HAPMAP

(<http://hapmap.ncbi.nlm.nih.gov/index.html.en>), 1000 Genomes Project (<http://www.1000genomes.org/>) and Exome Variant server (EVS) (<http://evs.gs.washington.edu/EVS/>).

VIII- Portals/Working consortiums

The primary goals of these projects are to generate catalogues of genomic abnormalities (somatic mutations, SNP genotyping, copy number variation profiling, abnormal expression of genes, epigenetic modifications) of series of genes in tumors from different cancer types. The main portals are: TCGA (<http://cancergenome.nih.gov/>), ICGC: (<https://icgc.org/>), OASIS (<http://www.oasis-genomics.org/>) and Firebrowse (<http://firebrowse.org/>).

IX- Impact on diseases

"Online Mendelian Inheritance in Man" (OMIM, <http://omim.org/>) is a catalog of human genes and genetic disorders; other databases providing information about human disorders and other phenotypes having a genetic component ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/intro/>), MedGen (<http://www.ncbi.nlm.nih.gov/medgen/>), dbGaP (<http://www.ncbi.nlm.nih.gov/dbgap/>), SNPs3D (<http://www.snps3d.org/>) and GTR (<http://www.ncbi.nlm.nih.gov/gtr/>).

X- Pathology

Authoritative books in pathology includes clinical features, morphologic, immunohistochemical and molecular genetic features and prognosis, with a very large iconography. They are the following: the "Rosai and Ackerman's Surgical Pathology" and the "WHO/IARC Classification of Tumours series" (<http://publications.iarc.fr/Book-And-Report-Series/Who-Iarc-Classification-Of-Tumours>). The Armed Forces Institute of Pathology (AFIP) publishes series of the "AFIP Atlas of Tumor Pathology". The Atlas of Genetics and Cytogenetics in Oncology and Haematology provides complete description of diseases, but again on a limited sample; on the other hand, articles on genes closely related to these diseases are found, right next, in the Atlas. As a product of collaborative work, the usefulness of the Atlas is dependent on colleague participation in updating and completing it. PathologyOutlines (<http://pathologyoutlines.com/>) provides iconography. To be also noted, the United States and Canadian Academy of Pathology (USCAP, <http://www.uscap.org/>). The International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) gives ICD-O codes for each cancer, with an ICD-O3-TOPO, which provides a topographical (organ) identifier and an ICD-O3-MORPH, which provides the basic and detailed pathology.

XI- Cancer Registries

Cancer registries are organizations seeking to collect, store, analyze, and report data on various cancers for epidemiological purposes. The International Agency for Research on Cancer

(IARC, <http://www.iarc.fr/>) is the specialized cancer agency of the World Health Organization (WHO/OMS). It publishes the "Cancer Incidence in Five Continents" series and GLOBOCAN (<http://globocan.iarc.fr/Default.aspx>).

The International Association of Cancer Registries (IACR, <http://www.iacr.com.fr/>) has developed classifications (the ICD-O), guidelines for registry practices and standard definitions, quality control, consistency checks and basic analysis of data, making data comparable between registries. The European Network of Cancer Registries (ENCR, <http://www.enrcr.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 in the USA. The Surveillance, Epidemiology, and End Results (SEER, <http://seer.cancer.gov/>) is a program of the National Cancer Institute. To be cited as well, the Union for International Cancer Control (UICC, <http://www.uicc.org/>).

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

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/>)). Interfaces between science and patients: These sites provide information for patients, including information on diseases, professionals for genetic counselling, laboratories, and laboratory tests: GeneTests (<https://www.genetests.org/>); NORD (<http://rarediseases.org/>); Orphanet (<http://www.orpha.net/>).

3. CYTOGENOMICS RESOURCES

Note: a detailed description of Cancer Cytogenomics resources may be found at Cancer Cytogenomics resources

I- Chromosome rearrangements/Hybrid genes

Mitelman Database:

The database of chromosome aberrations in cancer counts a total number of cases amounting to more than 60,000, implicating more than 10,000 gene fusions, culled from the literature and organized into distinct sub-databases: The "Cases Quick Searcher" and the "Cases Full Searcher" contain the data related to chromosomal aberrations in individual cases. The "Molecular Biology Associations Searcher" collects cases according to the gene

rearrangements. The "Clinical Associations Searcher" is based on tumor characteristics, related to chromosomal aberrations and/or gene rearrangements. This free access database shows raw data and is reliable.

Atlas of Genetics and Cytogenetics in Oncology and Haematology:

The Atlas (<http://atlasgeneticsoncology.org>) is a peer reviewed on-line journal encyclopaedia and database with free access on the Internet. It is an integrated structure and comprises the following topics: genes, cytogenetics and clinical entities in cancer, and cancer-prone diseases. The Atlas combines various types of knowledge all on one site: genes, gene rearrangements, cytogenetics, protein domains, function, cell biology, pathways. It also contains clinical genetics, including hereditary diseases which are cancer-prone conditions, and diseases, focusing on cancers, but also listing other medical conditions. The Atlas is mainly composed of structured review articles or "cards" (original monographs written by invited authors), The Atlas contributes to the cytogenetic diagnosis and may guide treatment decision making

COSMIC (<http://cancer.sanger.ac.uk/cosmic>) is a catalog of somatic mutations in cancer. It includes all abnormalities, from single nucleotide variations to chromosome rearrangements / fusion genes.

Other resources:

chimerDB 2.0
<http://biome.ewha.ac.kr:8080/FusionGene/> is a database of fusion genes with PubMed references and some information about the structure of chimeric genes. TICdb (<http://www.unav.es/genetica/TICdb/>) is a database of Translocation breakpoints In Cancer with the fusion sequences at the nucleotide level. ChiTARS (<http://chitars.bioinfo.cnio.es/>) is a database of chimeric transcripts. TCGA Fusion gene Data Portal (<http://54.84.12.177/PanCanFusV2/>) presents an analysis across tumor types of the TCGA program. Other resources are OMIM (<http://www.omim.org/>), Fusion cancer (<http://donglab.ecnu.edu.cn/databases/FusionCancer/>). "Cancer Cytogenetics: Chromosomal and Molecular Genetic Abberations of Tumor Cells" is a book authored by Sverre Heim and Felix Mitelman.

II- Data for SKY and FISH

Fluorescence in-situ hybridization (FISH) technique enables identification of chromosomal structures to be identified using specific probes. This significantly improves the localisation of breakpoints on chromosomes. FISH technique can also be used on non-dividing cells (interphase nuclei). The Cancer Chromosome Aberration Project (CCAP) has generated a set of BAC clones that have been mapped cytogenetically by FISH and physically by STSs to the human genome. The BAC data is

integrated into various databases ([http://cgap.nci.nih.gov/Chromosomes/CCAPBAC Clones](http://cgap.nci.nih.gov/Chromosomes/CCAPBACClones)), (<http://mkweb.bcgsc.ca/bacarray/>). All BAC can be located on the UCSC genome browser (<http://genome.ucsc.edu>). BAC from the fishClones file can be visualized on the chromosomal bands on the Atlas (<http://atlasgeneticsoncology.org/Bands/>). More recently, several commercial companies have developed more specific catalogs of FISH clones as oligonucleotides probes.

III- Comparative genomic hybridization (CGH) resources

This technique detects disequilibria between a disease sample and a normal sample. Several sites are repositories for these CGH/SNP profiles: GEO, (<http://www.ncbi.nlm.nih.gov/geo/>), Array Express (<http://www.ebi.ac.uk/arrayexpress/>), Tumorscape (<http://www.broadinstitute.org/tcga/home>), MetaCGH

(<http://compbio.med.harvard.edu/metacgh/>), CaSNP (<http://cistrome.org/CaSNP/>), Cell line project (http://cancer.sanger.ac.uk/cell_lines), Cancer Cell Line Encyclopedia (<http://www.broadinstitute.org/ccll/home>) and ArrayMap (<http://www.arraymap.org>)

IV- Mutation databases

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. Even in haematological malignancies, where the chromosome rearrangements have shown to bear a major role, nonetheless, it appears now that some mutations at the nucleotide level can still be very important in determining treatments in relation to patient outcome (e.g. ASXL1, ATM, BCL6, BRAF, KRAS and NRAS, CBL, CCND3, CDKN2A and CDKN2C, CEBPA, CRLF2, ETV6, FLT3, GATA2, ID3, IDH1, IDH2, IKZF1, JAK1, KIT, MYD88, NOTCH1, NPM1, RUNX1, TP53). The main mutation databases are: COSMIC (<http://cancer.sanger.ac.uk/cosmic>), CENSUS (<http://cancer.sanger.ac.uk/census/>), HGMD (<http://www.hgmd.cf.ac.uk/ac/index.php>), LOVD (<http://www.lovd.nl/3.0/home>), TCGA cBioPortal (<http://www.cbioportal.org/>), ICGC Data Portal (<http://dcc.icgc.org/>), OASIS Portal (see above), IntOGen (<http://www.intogen.org>), BioMuta v2 (<https://hive.biochemistry.gwu.edu/tools/biomuta/>), DoCM (<http://docm.genome.wustl.edu/>), CIViC (<https://civic.genome.wustl.edu/#/home>), and ExAC (<http://exac.broadinstitute.org>).

Resource	URL	Location	PMID	Last Update
Nomenclature				
The Human Gene Nomenclature Database (HUGO)	http://www.genenames.org/	(Hinxton, UK)	26842383	2016
International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3)	http://www.who.int/classifications/icd/adaptations/oncology/en/	(IARC, Lyon, Fr)	-	2016
Human Genome Variation Society	http://www.hgvs.org/	(Melbourne, Aus)	-	2016
Cards				
Atlas of Genetics in Oncology and Haematology	http://atlasgeneticoncology.org/	(INIST-CNRS, Nancy, Fr)	23161685	2016
Entrez_Gene (NCBI)	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene	(NCBI, Bethesda, US)	26615191	2016
European Bioinformatic Institute (EBI)	https://www.ebi.ac.uk	(EBI, Hinxton, UK)	26673705	2016
EnSembI (Sanger - EBI)	http://www.ensembl.org/	(Sanger_EBI, Hinxton, UK)	26687719	2016
Genecode	http://www.genecodegenes.org/	(Sanger Institute, Hinxton, UK)	22955987	2015
GeneCards: human genes, proteins and diseases	http://www.genecards.org/	(Weizmann, Rehovot, Is)	27048349	2016
SOURCE	http://source-search.princeton.edu/cgi-bin/source/sourceSearch	(Princeton, US)	-	2015
AcceView	http://www.ncbi.nlm.nih.gov/EBI-Research/Assembly/	(NCBI, Bethesda, US)	16925834	2012
GENATLAS	http://www.dsi.univ-paris5.fr/genatlas	(Imagine, Paris, Fr)	9835018	2016
WikiGenes	http://www.wikigenes.org/	(De)	18728691	-
H-invDB	http://www.h-invitational.jp/hinv/hg-db/index.jsp	(Tokai University,Ja)	23197657	2015
Cancer Portals				
Genomic Data Commons	https://gdc.nci.nih.gov	(NCI, Bethesda, US)	-	2016
ICGC Data Portal	https://dcc.icgc.org/	(ICRC, Ontario, Ca)	23930502	2015
TCGA cBioPortal	http://www.cbioportal.org/public-portal/	(MSKCC, New York, US)	26852330	2016
Broad Tumor Portal	http://www.tumorportal.org/	(Broad Institute, Boston, US)	24380350	2014
FireBrowse GDAC	http://firebrowse.org/	(Broad Institute, Boston, US)	-	2016
GTEx Portal	http://www.gtexportal.org/home/	(Broad Institute, Boston, US)	25954001	2016
Integrative Onco Genomics (mOgen)	http://www.intogen.org/	(Barcelona, Es)	20111033	2014
OASIS Portal	http://www.oasis-genomics.org/	(Pfizer, US)	-	2015
Cancer Browser (UCSC)	https://genome-cancer.ucsc.edu/	(San Diego, US)	25382408	2015
canSAR	https://cansar.icr.ac.uk/	(ICR, London, UK)	26673713	2016
IHEC Data Portal	http://epigenomesportal.ca/ihcc/	(Mc Gill, Ca)	-	-
Genomic and Cartography				
Human Genome Browser (UCSC)	http://genome.ucsc.edu/goldenPath	(San Diego, US)	26580259	2015
Ensembl Map view (Sanger_EBI)	http://www.ensembl.org/	(Sanger_EBI, Hinxton, UK)	26687719	2015
NCBI Map Viewer : Homo_sapiens genome view	http://www.ncbi.nlm.nih.gov/mapview/	(NCBI, Bethesda, US)	26615191	2015
CGAP - Bac clones	http://ccgap.scri.nh.gov/Chromosomes/CCAP_BAC_Clones	(NCI, Bethesda, US)	-	-
Gene, transcription and regulation				
NIH Genetic Sequence database (GenBank)	http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html	(NCBI, Bethesda, US)	26615191	2016
Reference Sequence database (RefSeq)	http://www.ncbi.nlm.nih.gov/refseq/	(NCBI, Bethesda, US)	26615191	2016
Encyclopedia of DNA Elements (ENCODE)	https://www.encodeproject.org/	(NHGRI, US)	26962025	2016
The Consensus CDS Project (CCDS)	http://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi	(UCSC, San Diego, US)	24217909	2016
Unigene (NCBI)	http://www.ncbi.nlm.nih.gov/UniGene/	(NCBI, Bethesda, US)	26615191	2016
ASG - Alternative Splicing Gallery (ASG)	http://statgen.ncsu.edu/asg/	(NCSSU, US)	15292448	2004
Gene Expression Atlas (EBI)	http://www.ebi.ac.uk/psa/The_Array/Express	(EBI, Hinxton, UK)	26481351	2016
Gene Expression Omnibus (NCBI)	http://www.ncbi.nlm.nih.gov/geo	(NCBI, Bethesda, US)	27006011	2016
Exploration of Expression Compendium (SEEK)	http://seek.princeton.edu/	(Princeton, US)	25581801	2015
Multi Experiment Matrix (MEM)	http://bit.ca.ut.ec/mem/index.cgi	(Tartu, Est)	19961599	2012
BioGPS	http://biogps.org/	(Scrpps, US)	26578587	2016
Human Epigenome	http://www.epigenome.org/	(Sanger, UK)	15550966	-
http://ihcc-epigenomes.org/	http://ihcc-epigenomes.org/	(Mc Gill, Ca)	-	-

Protein : sequence, function, domain, 3D structure				
Atlas of Genetics in Oncology and Haematology	http://atlasgeneticsoncology.org/	(INIST-CNRS, Nancy, Fr)	23161685	2016
UniProt	http://www.uniprot.org/	(EBI, Hinxton, UK)	26088053	2016
SwissProt	http://www.expasy.ch/swprot/	(SIB, Geneva, Ch)	26088053	2016
Swiss-VAR	http://swissvar.expasy.org/	(SIB, Geneva, Ch)	20106818	2016
NextProt	http://www.nextprot.org/	(SIB, Geneva, Ch)	25593349	2016
ENZYME	http://www.expasy.org/enzyme/	(SIB, Geneva, Ch)	10592255	2016
PhosphoSitePlus	http://www.phosphosite.org/	(Danvers, US)	25514926	2014
Prosite: Protein signature	http://www.expasy.ch/prosite/	(SIB, Geneva, Ch)	23161676	2016
Protein families (PFAM)	http://pfam.xfam.org/	(Sanger, Hinxton, UK)	26673716	2015
A Conserved Domain Database (CDD)	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml	(NCBI, Bethesda, US)	18984618	2015
Domain mapping of disease mutations (DMDM)	http://bcmf.umc.edu/dmdm/	(Baltimore, US)	20681956	2014
Protein domain (PRODOM)	http://prodom.probs.fr/	(PRABI, Lyon, Fr)	12230031	2015
Protein Data Bank (PDB)	http://www.rcsb.org/	(San Diego, US)	25428375	2016
Pictorial database of 3D structures (PDBeur)	http://www.ebi.ac.uk/hinxton/scv/databases/pdbeur/	(EBI, Hinxton, UK)	15601193	2016
Jena Library of Biological Macromolecules (JMB)	http://jmb.fh-leibniz.de/	(Jena, De)	11752308	2015
Structural Biology Knowledgebase (SBKB)	http://sbkb.org/pdbid/	(Rutgers, US)	21472436	2016
Structural Classification of Proteins (SCOP)	http://scop.berkeley.edu/search?ver=2.05	(Berkeley, US)	24304899	2015
Classification of protein structures (CATH)	http://www.cathdb.info/	(UCL, London, UK)	25348408	2013
Human Protein Atlas	http://www.proteinatlas.org/	(Uppsala, Su)	21572409	2016
Human Protein Reference Database (HPRD)	http://www.hprd.org/	(John Hopkins, Baltimore, US)	22159132	2010
Differentially Expressed Proteins in Human Cancer (dbDEPC)	http://dbdepc.hiujiao.org/	(Shanghai, Cn)	22096234	2011
Protein Interaction databases				
Database of Interacting Proteins (DIP)	http://dip.doe-mbi.ucla.edu/dip/Query.cgi	(UCLA, US)	14681454	2014
Molecular Interaction Database (IntAct)	http://www.ebi.ac.uk/intact/index.jsp	(EBI, Hinxton, UK)	22121220	2016
Functional coupling (FunCoup)	http://funcoup.sbc.su.se/	(KTH, Stockholm, Su)	24183702	2013
Biological General Repository for Interaction Datasets (BioGRID)	http://www.biogrid.org/	(Toronto, Ca)	26728913	2016
Ontologies - Pathways				
Gene Ontology	http://www.geneontology.org/	-	23378336	2016
QuickGO	http://www.ebi.ac.uk/QuickGO/	(EBI, Hinxton, UK)	19744993	2016
Kegg Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/	(Kyoto, Jp)	26476454	2016
BioCarta Pathways	http://scop.aci.nih.gov/Pathways/BioCarta_Pathways	(NCI, Bethesda, US)	-	-
Reactome	http://www.reactome.org/	(OICR, ca; New-York, US; EBI, UK)*	26656494	2016
Network Data Exchange (NDEX)	http://www.ndexbio.org/	(University of California, US)	26594663	2016
Atlas of Cancer Signaling Networks global map (ACSN)	http://acsn.curie.fr/	(Curie, Paris, Fr)	26192618	2013
WikiPathways	http://www.wikipathways.org/index.php/		22096230	2016
Orthology - Evolution				
The Hierarchical Catalog of Eukaryotic Orthologs (OrthoDB)	http://cegg.unige.ch/orthodb/	(Univ. Geneva, Ch)	25428351	2011
TREEFAM:	http://www.treefam.org/	(EBI, Hinxton, uk)	24194607	2013
Homologene (NCBI)	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene	(NCBI, Bethesda, US)	26612181	2016
Gene Sifter (UCSC)	http://genome.ucsc.edu/goldenPath/help/hsNearHelp.html	(UCSC, San Diego, US)	15867434	2016
Ortholog groups with inparalogs (InParanoid)	http://inparanoid.sbc.su.se/cgi-bin/index.cgi	(Stockholm, Su)	23429972	2013
HOVERGEN	http://www.prabi.fr/ovip.php?article=38	(Prabi, Lyon, Fr)	19534752	2009
Gene fusions - Chromosomal Rearrangements				
Atlas of Genetics in Oncology and Haematology	http://atlasgeneticsoncology.org/	(INIST-CNRS, Nancy, Fr)	23161685	2016
Mitelman Database of Chromosome Aberrations in Cancer	http://scop.aci.nih.gov/Chromosomes/Mitelman	(NCBI, Bethesda, US)	-	2016
Catalog of somatic mutations in Cancer (COSMIC)	http://cancer.sanger.ac.uk/cosmic/	(Sanger Center, Hinxton, UK)	25355519	2016
Database of Translocation breakpoints in Cancer (TCdb)	http://www.ontario.ca/genetics/TCdb/	(Univ Navarre, Sp)	17257420	2013
ChTARS	http://chitars.bioinfo.cim.uva.es/	(Barcelona, Es)	21414346	2014
TCGA Fusion gene Data Portal	http://24.84.12.177/PanCafusV2/	(MDACC, Houston, US)	25500544	2014
Fusion Cancer	http://donatlab.cmu.edu/donations/FusionCancer/index.html	(Beijing, Ca)	26212638	2014
ChimerDB 2.0	http://biome.ewha.ac.kr/B09/FusionGene/	(Ewha Womans University, Kr)	19906715	2010
Database of Chromosomal Rearrangements In Disease (dCRID)	http://vcl.acornscience.com/dCRID/	(Houston, US)	21051346	2010
arrayMap - genomic arrays for copy number profiling in human cancer	http://www.arraymap.org/	(UZH-SIB, Zurich, Ch)	23428357	2016
CONAN : Cell lines Project: Copy Number Analysis	http://www.sanger.ac.uk/cgi-bin/genetics/CGP/conan/search.cgi	(Sanger Center, Hinxton, UK)	-	2016
Polymorphisms : SNP, mutations				
Single Nucleotide Polymorphism (dbSNP)	http://www.ncbi.nlm.nih.gov/SNP/summaryview.html	(NCBI, Bethesda, Us)	26615191	2016
The International HapMap Project	http://www.hapmap.org/index.html.en	(NCBI, Bethesda, US)	20811451	2011
A Deep Catalog of Human Genetic Variation (1000 Genomes)	http://www.1000genomes.org/	(EBI, Hinxton, UK)	26432245	2016
Exome Aggregation Consortium (ExAC)	http://exac.broadinstitute.org/	(Broad, US)	26791959	2016
Exome Variant server (EVS)	http://evs.washington.edu/	(Washington, US)	22604720	2014
ICGC Data Portal	http://dcc.icgc.org/	(OICR, Ontario, Ca)	21930502	2015
TCGA Copy Number Portal	http://www.broadinstitute.org/tcga/home	(Broad Institute, Boston, US)	20164920	2013
Cancer Gene Census	http://cancer.sanger.ac.uk/cancergenome/projects/census/	(Sanger Center, Hinxton, UK)	14993899	2016
Catalog of somatic mutations in Cancer (COSMIC)	http://www.sanger.ac.uk/perl/CGP/cosmic	(Sanger Center, Hinxton, UK)	23355519	2016
Leiden Open Variation Database (LOVD 3.0)	http://www.lovd.nl/3.0/home	(Leiden, Ne)	21520333	2016
BioData v2	https://hive.biochemistry.gwu.edu/tools/biomdata/	(George Washington Univ, Washington DC, US)	24667251	2014
DocM Database of curated mutations	http://docm.genome.wustl.edu/	(WUSTL, US)	-	2016
CTVc Clinical Interpretations of Variants in Cancer	https://ctvc.genome.wustl.edu/#home	(WUSTL, US)	-	2016
NCG5	http://ncg.kcl.ac.uk/		26516186	2016
Cancer3D	http://www.cancer3d.org/		23392415	2015
Integrative Onco Genomics (IntOGen)	http://www.intogen.org/	(Barcelona, Es)	20111033	2014
Human Gene Mutation Database (HGMD) (*)	http://www.hgmd.cf.ac.uk/ac/	(Institute of Medical Genetics, Cardiff, UK)	24077912	2016
Database of genomic structural variation (dbVar)	http://www.ncbi.nlm.nih.gov/dbvar/	(Bethesda, NCBI, US)	26615191	2016
DGV - Genomic Variants	http://dgv.tong.ca/dgv-app/home	(Toronto, Ca)	24124517	2015
DECIPHER	https://decipher.sanger.ac.uk/	Sanger Centre, Hinxton, UK)	19344873	2016
SNP3D	http://www.snp3d.org/	(UMD, US)	16412461	2006
Human Genome Variant Society	http://www.hgvs.org/content/guidelines	(Melbourne, Aus)	26931183	2016

Diseases				
Atlas of Genetics in Oncology and Haematology	http://atlasgeneticsoncolore.org/	(INIST-CNRS, Nancy, Fr)	23161685	2016
Online Mendelian Inheritance in Man (OMIM)	http://www.omim.org	(John Hopkins, Baltimore, US)	25428349	2016
Human medical genetics (MedGen)	http://www.ncbi.nlm.nih.gov/medgen/	(NCBI, Bethesda, US)	26615191	2016
Database of Genotypes and Phenotypes (dbGaP)	http://www.ncbi.nlm.nih.gov/gap/	(NCBI, Bethesda, US)	26615191	2016
Human variations and phenotypes (ClinVar)	http://www.ncbi.nlm.nih.gov/clinvar/intro	(NCBI, Bethesda, US)	27037489	2016
The Genetic Testing Registry (GTR)	http://www.ncbi.nlm.nih.gov/gtr/	(NIH, Bethesda, US)	23193275	2016
A medical genetics information resource (GeneTests)	https://www.genetests.org/discovers/	(Elmwood Park, New Jersey, US)	18973587	2016
HuGE Navigator	https://pharkb.cdc.gov/HuGENavigator/home.do	(CDC, Atlanta, US)	-	2016
Database of rare diseases and orphan drugs (ORPHANET)	http://www.orpha.net/	(INSERM, Paris, Fr)	12655825	2016
National Organization for Rare Diseases (NORD)	http://rarediseases.org/	US	-	2016
Clinical trial, drugs and therapy				
BioCentury BC1Q	http://www.biocentury.com/	(Redwood City, US)	-	2016
The Drug Gene Interaction Database (DgiDB)	http://dgidb.genome.wustl.edu/	(WUSTL, US)	26311824	2016
Comparative Genomics Database (CTD)	http://ctd.mdibl.org/	(NC State Univ, US)	25326323	2016
Pharmacogenomics Knowledge	http://www.pharmgkb.org/	(Stanford, US)	-	2016
Genomics of Drug Sensitivity in Cancer (Wellcome Trust)	http://www.cancersites.org/	(Sanger, Haxton, UK)	23180760	2012
Clinical Trial	http://clinicaltrials.gov/ct2/search/browse/browse-intn_cat_A23ee	(NIH, Bethesda, US)	-	2016
Cancer registries				
International Agency for Research on Cancer	http://www.iarc.fr/	(Lyon, Fr)	-	2016
International Association of Cancer Registries	http://www.iacr.com.fr/	(Lyon, Fr)	-	2016
European Network of Cancer Registries	http://www.enrc.eu/	-	-	2016
National Program of Cancer Registries	http://www.cdc.gov/cancer/	US	-	2016
Union for International Cancer Control	http://www.uicc.org/	US	-	2016
Bibliography, Data mining				
COREMINE	http://www.coremine.com/medical/#search	(Oslo, No)	-	2016
Information Hyperlinked over Proteins (IHOP)	http://www.ihop-net.org/UnPub/IHOP/	(MSKCC, New-York, US)	18478678	2016
Gene Interactions in Cancer (ZODIAC)	http://www.compenome.org/ZODIAC/	(Evanston, US)	27976326	2016
PubMed	http://www.ncbi.nlm.nih.gov/pubmed/	(NIH, Bethesda, US)	26615191	2016
PubMed Central	http://www.ncbi.nlm.nih.gov/pmc/	(NIH, Bethesda, US)	26615191	2016
PubChem	http://pubchem.ncbi.nlm.nih.gov/	(NIH, Bethesda, US)	26615191	2016
GoPubMed	http://www.gopubmed.org/web/gopubmed/	(NIH, Bethesda, US)	26615191	2016

TABLE 1: Internet resources

4. PRACTICAL EXERCISE

See online Practical Exercises.

5. DISCUSSION

We have briefly discussed the various databases useful for clinicians, students, and researchers in finding answers to their questions and in determining which field in cancer research still needs to be studied. Only a handful of databases or portals take the cytogenetic information into consideration although being one of the first observation points confirming that the cell has transformed into a cancerous cell. Over the years (1960-2016), chimeric genes and fusion proteins have been discovered mainly by cytogenetic means. This has led to understanding of major cancerogenetic processes, and, later on, to the concept of treatments targets for many cases. Cytogenetics, or rather, cytogenomics of cancer, is a major contributor for the concept of "personalized medicine for cancer", together with other tools for other mechanisms.

Dorothy Warburton wrote her colleagues in the American Cytogenetics Forum List: Subject: Is Cytogenetics Dying? "I have been told by my department chairman and other advisers that cytogenetics is dying at least three times: in 1968 (just before banding), in 1984 (just before FISH) and in 2001 (just before microarrays) (...) Chromosome changes are being recognised as the cause of more and more abnormalities, not fewer. This is true for both cancer and constitutional cytogenetics (...) There are also a great many basic facts about how chromosome abnormalities originate that we know little about and that are great fields for research".

Besides what is known, there is still a lot to learn about how a mutation or a chromosomal rearrangement is influencing the cellular mechanism.

A decade ago, conventional wisdom was that the expressed genes were regulating all the processes in the cell and the rest of the genome was considered as "junk DNA".

The main purpose of this junk DNA was to act as a buffer and to protect the coding part from any kind of aggression. Progressively, we realized that the mutations happening outside the coding regions could influence the expression of nearby genes or in some cases several kilobases away. These mutations were located in regulatory regions recruiting or blocking transcription factors binding DNA.

More elements were found in the "junk DNA" with the discovery of the microRNAs (miRNA) and long non-coding RNAs (lincRNA). Both proved influential in regulating the genome and the expression of several genes.

These miRNA and lincRNA are deregulated during the carcinogenic process and studies have demonstrated that by re-establishing the normal expression of one or few miRNAs or lincRNAs, the cancer cells transform back into normal cells. Afterwards, it was observed that genes which were not mutated could be silenced and vice versa. This was due to the epigenetic regulation of the genes and the recruitment of epigenetic regulators in

different regions of the genome.

The cancerous cells actively modify their own epigenetic marks on the genome to increase the

expression of stemness genes providing them with proliferative or resistance advantages.

The chromosome structure in the nucleus is dependent on many variables modulating the chromatin interactions across the whole genome. Techniques like chromatin conformation capture and interphase fluorescent in situ hybridization (FISH) detect spatial associations between specific genes. In the future, they are likely to be the techniques for detecting abnormalities associated with progression of tumours. Similarly, when modifying the 3D structure of chromosomes, it has been demonstrated that the disruption of chromosome neighbourhoods via mutations in insulated neighbourhood boundaries (cohesin CTCF interactions) activates proto-oncogenes in cancer cells (Hnisz D et al., 2016).

As mentioned above, the last build of the human genome (GRCh38, Dec 2013) is more precise than the previous ones and takes into account more haplotypes shared by parts of the human population. Nevertheless, due to a great part of repeated sequences of various classes (50% of the genome), there are some gene families only present at low frequencies, e.g. the insertions of endogenous retroviruses, which represent 8% of the genome and are mostly not localized on GRCh38 for a great part (Wildschutte JH et al., 2016). We need to recognize that such a diversity may be a factor leading to susceptibility to cancer or other diseases.

There is an increasing realisation that our environment, our nutrition and our way of life have an impact on how our cells mutate and how these mutations are repaired or how the abnormal cells are eliminated by our organism.

It is difficult to fully understand and appreciate the complexity of the cellular mechanism and the various levels that can be deregulated in cancer makes. The use of databases makes this process easier as it condenses the complex information and provides links to other relevant databases providing even more specialized information.

Sequencing of thousands of genomes of patients bearing different types of tumours and genomes of normal persons, generates a huge amount of data demonstrating the complexity of tumours. Focus points are the sensitive parts of the genome which are more prone to be mutated, the expression profiles, the function of the proteins, the pathways the proteins are located in, and the protein interactions (in normal cases and abnormal cases). The same is done at the single cell level, bringing a huge amount of data demonstrating the heterogeneity of the tumour and the interaction with the niche. In principle, this wealth of integrated genomic data and clinical information could reveal the genetic bases of cancer, inherited diseases, and drug responses-illnesses and remedies that have touched nearly every person and family across the

globe. This will constitute the databases of tomorrow.

Interpreting these data requires a larger evidence base than any one party alone can develop. However, existing technologies, regulations and approaches are currently not designed for sharing and interpreting this wealth of information effectively, especially across diseases and nations. Databases will need to integrate information increasingly in the upcoming years but also stay interoperable with other databases. This corroborates the idea of having a common nomenclature and language to avoid mistakes but also have common and reliable information

(http://www.aacr.org/Research/Research/Pages/aacr-project-genie.aspx#.Vx_mi8f90qx;

<https://genomicsandhealth.org>) in the same manner in which the Working Group for Planetary System Nomenclature maintains the "astronomical naming conventions and planetary nomenclature", any scientific field, confronted with the exponential growth of data, has to create a system of classification and thesaurus for naming of new items, and nosological definitions. This often implies the use of a structured grammar and census of accepted terms/objects.

Resources such as the International System for Human Cytogenetic Nomenclature (ISCN), the International Classification of Diseases for Oncology (ICD-O), the Human Gene Nomenclature Database (HGNC), and the Nomenclature for the description of sequence variations (from the HGVS) are indispensable research tools allowing a common language.

Interoperability: In the context of interoperability between recent cancer projects, such as TCGA and ICGC as well as all data produced by thousands laboratories or hospitals, the Global Alliance for Genomics and Health (Global Alliance, <https://genomicsandhealth.org/>) (Lawler M et al., 2015) was created to accelerate the potential of genomic medicine. This association brings together over 375 leading institutions working together to generate a common framework of harmonized approaches in enabling the data sharing. In particular, it works to establish interoperable technical standards (standardised language and tools) to management of genomic and clinical data and for a better representation of genotype-phenotype associations.

Open data (open source, open hardware, open content, and open access) is a concept so dear to many, in particular in the medical and scientific world. This concept has recently had a renewed vigour with/since the advent of the Internet. Maintaining the data open and free remains a daily struggle. Everything has a cost, even a free database, from upkeeping and updating the database to

providing for the scientific staff necessary to produce expertized data.

Data should remain freely available. However, a business model remains to be established. Although economic investment by the public sector would benefit the whole of mankind, as well as economically profitable in the end, most of the institutional stakeholders are gradually disengaging, and well-known databases are forced to beg for funds (see recent examples, useless to cite them!) or to disappear. This would be a regrettable drawback for the scientific and medical community - yet it may still happen.

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 Entstehung maligner Tumoren 1914 Gustav Fischer

Nowell PC, Hungerford DA. A minute Chromosome in Human Chronic Granulocytic 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. Identification 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 N 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, Willén 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, Lecoine N, Jonveaux P, Souyri M, Mauchauffé 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, Schäfer 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

Möller 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, Adélaïde 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, Øra 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
- P´aszczyca A, Nilsson J, Magnusson L, Brosjö O, Larsson O, Vult von Steyern F, Domanski HA, Lilljebjörn 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, Férec 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, Gröger D, Park TS, Emerenciano M, Pombo de Oliveira M, Renneville A, Villarese P, Macintyre E, Cavé 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, Schäfer 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, Mörücke A, Stanulla M, Schrappe M, Sedék L, Szczepa´ski T, Zwaan CM, Coenen EA, van den Heuvel-Eibrink MM, Strehl S, Dworzak M, Panzer-Grümayer 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, Hogendorn PC, Succi 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, Hurler 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, Khrebtkova 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, Langerød A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Børresen-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, Neriha 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, Klcó 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, Viguié 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, Cavé H, Clappier E, van Dongen JJ, Balgobind BV, van den Heuvel-Eibrink MM, Beverloo HB, Panzer-Grümayer R, Teigler-Schlegel A, Harbott J, Kjeldsen E, Schnittger S, Koehl U, Gruhn B, Heidenreich O, Chan LC, Yip SF, Krzywinski M, Eckert C, Mörücke A, Schrappe M, Alonso CN, Schäfer BW, Krauter J, Lee DA, Zur Stadt U, Te Kronnie G, Sutton R, Izraeli S, Trakhtenbrot L, Lo Nigro L, Tsaour 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-Demtröder K, Kruhøffer M, Hager H, Knudsen B, Andersen CL, Sørensen KD, Pedersen JS, Ørntoft TF, Dyrskjøt 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, Bäuerle T, Wallwiener M, Holland-Letz T, Höfner T, Sprick M, Scharpf M, Marmé 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, Berni 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 WJ, Robson SC, Chung CW, Hopf C, Savitski MM, Huthmacher C, Gudgin E, Lugo D, Beinke S, Chapman TD, Roberts EJ, Soden PE, Auger KR, Mirquet 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 Topolli inhibitors *Nature* 2015 Apr 9;520(7546):239-42
- Bernheim A, Huret JL, Guillaud-Bataille M, Brison O, Couturiers J; Groupe Français de Cytogéné 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, Girón 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-Kozłowska H, Kozłowski 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 Aberrations 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 Mendibíl 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, Boulosa 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
- Løv f 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, Kusuha 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, Döhner 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, Seagraves 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, Ferté C, Soria JC, Friend SH, André 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, Füllgrabe 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

Beroukhir 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, Taberero 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

Iafraite 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, Shaperro MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs 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, McCarroll 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, Ghorri 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, Brägelmann 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 K, de Pedro-Cuesta J, Haik 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, Kähler 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, Moutl 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, Ciufu 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, Cucho 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, Scheremetjov 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. Internet databases and resources for cytogenetics and cytogenomics. *Atlas Genet Cytogenet Oncol Haematol*. 2016; 20(4):218-237.
