Atlas of Genetics and Cytogenetics in Oncology and Haematology in 2013

Jean-Loup Huret, Mohammad Ahmad, Mélanie Arsaban, Alain Bernheim, Jérémy Cigna, François Desangles, Jean-Christophe Guignard, Marie-Christine Jacquemot-Perbal, Maureen Labarussias, Vanessa Leberre, et al.

HAL Id: hal-01587507
https://hal.sorbonne-universite.fr/hal-01587507

Submitted on 14 Sep 2017

Atlas of Genetics and Cytogenetics in Oncology and Haematology in 2013

Jean-Loup Huret[1,2,*], Mohammad Ahmad[1], Mélanie Arsaban[1], Alain Bernheim[3], Jérémy Cigna[1], François Desangles[1,4], Jean-Christophe Guignard[1], Marie-Christine Jacquemot-Perbal[5], Maureen Labarussias[1], Vanessa Leberre[1], Anne Malo[1], Catherine Morel-Pair[5], Hossein Mossafa[1,6], Jean-Claude Potier[7], Guillaume Texier[7], Franck Viguié[1,8], Sylvie Yau Chun Wan-Senon[1], Alain Zasadzinski[5] and Philippe Dessen[3]

[1]'Atlas in Poitiers Editorial Group', [2]Genetics, Department Medical Information, University and Hospital of Poitiers, INSERM U 935, 86021 Poitiers, France, [3]Génétique des Tumeurs, 985 INSERM, Institut Gustave Roussy, 94805 Villejuif, France, [4]HIA Val de Grâce, 75005 Paris, France, [5]INIST-CNRS, 95310 Nancy, France, [6]Laboratoire Cerba, 95310 Cergy-Pontoise, France, [7]CRITT Informatique, 86360 Poitiers, France and [8]Cytogenetics, CHU Saint Antoine, 75012 Paris, France

ABSTRACT

The Atlas of Genetics and Cytogenetics in Oncology and Haematology (http://AtlasGeneticsOncology. org) is a peer-reviewed internet journal/encyclopaedia/database focused on genes implicated in cancer, cytogenetics and clinical entities in cancer and cancer-prone hereditary diseases. The main goal of the Atlas is to provide review articles that describe complementary topics, namely, genes, genetic abnormalities, histopathology, clinical diagnoses and a large iconography. This description, which was historically based on karyotypic abnormalities and *in situ* hybridization (fluorescence *in situ* hybridization) techniques, now benefits from comparative genomic hybridization and massive sequencing, uncovering a tremendous amount of genetic rearrangements. As the Atlas combines different types of information (genes, genetic abnormalities, histopathology, clinical diagnoses and external links), its content is currently unique. The Atlas is a cognitive tool for fundamental and clinical research and has developed into an encyclopaedic work. In clinical practice, it contributes to the cytogenetic diagnosis and may guide treatment decision making, particularly regarding rare diseases (because they are numerous and are frequently encountered). Readers as well as the authors of the Atlas are researchers and/or clinicians.

INTRODUCTION

Why the Atlas?

Cancer, the second most common cause of death in the developed countries ($\sim$27%), is the first cause of untimely death (>35% of deaths before the age of 65 years). Cancer is a genetic disorder. The prognosis of a leukaemia depends on the genes involved; median survival is 3 months in case of an inv(3)(q21q26) (RPN1/EVI1), whereas 95% of patients with a dic(9;12)(p13;p13) (PAX5/ETV6) are cured. Treatments depend on the severity of the disease. However, there are >900 leukaemia entities! In addition to the small set of a hundred or so genes are known to play a major role in cancer, 2000–9000 other genes are possibly implicated in cancer, 1200 types of solid tumours exist and 'n' hereditary diseases are cancer-prone conditions. Also, 25 000 new publications on cancer genetics in man are added annually to PubMed.

The Atlas (1–5), which began in 1997, is an internet journal/encyclopaedia/database, and, as such, it is a hybrid formula. Like any scientific journal, it is peer reviewed and, like any internet site, it is easy to use (hyperlinks, updates, automation in data mining). The Atlas focuses on genes involved in cancer, cytogenetics and clinical entities in cancer and cancer-prone diseases. This is the place where complementary topics, namely, genes, genetic abnormalities, histopathology and clinical diagnoses are developed and hyperlinked. We also have established links to the major external databases in the related field(s). It is a collective undertaking by researchers and clinicians aimed at describing the state of the art in cancer

*To whom correspondence should be addressed. Tel: +33 5 49 44 45 46 or +33 5 49 45 47 67; Email: jean-loup.huret@chu-poitiers.fr or jlhuret@univ-poitiers.fr

genetics for the medical and scientific community and providing a cognitive tool for fundamental and clinical research.

Readers of the Atlas are consultants in hospitals, researchers, university teachers, but students of medicine and of the life sciences also use the Atlas.

DATABASE CONTENTS

About 2200 authors have so far contributed to the Atlas, making 2167 review articles available. The Atlas contains peer-reviewed articles on 1135 genes, 503 leukaemia entities, 177 solid tumours and 104 cancer-prone inherited diseases. It also contains 'automated cards' on 8190 other genes potentially implicated in cancer. Automated cards are produced in the following manner: a list of genes is obtained from the gene directory of the NCBI (6), filtered by keywords related to cancer in Description, GO Terms and GeneRIF fields. Location data are selected from the refGene.txt file on the UCSC Genome server (7). Automated cards contain data released into the public domain and hyperlinks towards the main databases. The automated cards section will be extended to encompass the entire genome, that is, ∼30 000 genes. The Atlas also contains traditional articles, called 'Deep Insights' (81 articles to date), dealing with topics in areas related to core subjects in the Atlas [e.g. chromothripsis, centrosome, autophagy and so forth (8–10)], 63 case reports on haematological malignancies and 124 chapters in the educational section (with ∼40 in English, Spanish and French). Whenever possible, articles on leukaemias or solid tumours contain data on the prognosis, to help clinicians, and an iconography of the chromosomes and histopathology specimens for the biologist's diagnosis. The Atlas contains an iconography of ∼11 800 images.

Articles can be accessed either by theme (genes, leukaemias, solid tumours, cancer-prone disease) or by chromosome. In the latter case, genes may be displayed in alphabetical order or in physical order from pter to qter, according to human genome assembly hg19 [February 2009, UCSC Genome browser (7)]. The 'Case Reports in haematology' section was created to allow the delineation of new leukaemia entities, describing and revealing the natural history/epidemiology of a rare disease with no previous clinical description and an unknown prognosis. This contributes to applied research in epidemiology, and it will allow better therapeutic decisions. A portal with numerous links devoted to genetics and/or cancer adds to the information available in the Atlas.

RECENT DEVELOPMENTS

Cell biology corner

The various articles related to a given topic in the field of cell biology (e.g. apoptosis, cell cycle, micro RNA, nuclear membrane), as well as topics related to physiopathology (e.g. epithelial–mesenchymal transition or angiogenesis) are collected and grouped together in this section, along with a selected iconography (Figure 1).

Atlas database

In 1997, the Atlas templates consisted of concise structured files. The model had to evolve with novel developments (new areas being explored, such as the cell biology corner, and the fact that full review articles gradually replaced the concise files). The templates of the articles in the database had to be modifiable. This was done by describing headings, sub-headings and the type of information they contain, with no limits on the hierarchy of the sub-headings, so that the database structure could evolve constantly. The concepts of meta object facilities (MOF) (11) http://dl.acm.org/citation.cfm?id = 1028976.1029004&coll = DL&dl = GUIDE were used to implement such a process. MOF makes it possible to describe not only the data but also the way it is structured in the same environment. MOF offers the possibility of changing dynamically allocated data structures and of applying these evolutions to the data without re-programming applications. It consists of a fixed meta-model that describes the data representation model and the resulting model that stores the data. Although tools exist to create such a database, they are still experimental and cumbersome to implement for users. A hybrid model has been created, using MOF exclusively for the concepts concerning the structuring of the articles. The other data structures used to manage the application are implemented in a relational database (Figure 2). In addition, an automatic survey of different databases permits an updated list of consistent identifiers that are used to generate the external links for all the genes defined in the Atlas.

Open access electronic journal

An 'open access journal' version of the Atlas is now available on the digital publishing platform I-Revues of the Institute for Scientific and Technical Information (INIST) of the French National Centre for Scientific Research (CNRS). It presents the archives of a quarterly journal since 1997, which became a bimonthly journal in 2008 and a monthly journal in 2009, comprising 2215 articles in 90 volumes, which constitutes a 7465-page collection, available at: http://documents.irevues.inist.fr/handle/2042/15655. DSpace software, an open source repository software package, is used. It is based on banks of digital objects described by a set of standardized metadata in extensible mark-up language format (Dublin Core qualified) and a standard uniform identifier (CNRI handle). Metadata are also available on the Web in Dublin Core extensible mark-up language format and are freely harvestable. DSpace supports the common interoperability standards used in the institutional repository domain, such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH protocol). A digital object identifier (DOI®), recorded at the International Agency CrossRef, is assigned to each article, and it ensures the sustainability of their ranking and enables publishers to create direct links between scientific articles and also with databases.

Figure 1. An example of a cell biology cluster, where articles and iconography concerning a given theme are assembled and are illustrated here by the nuclear membrane topic.
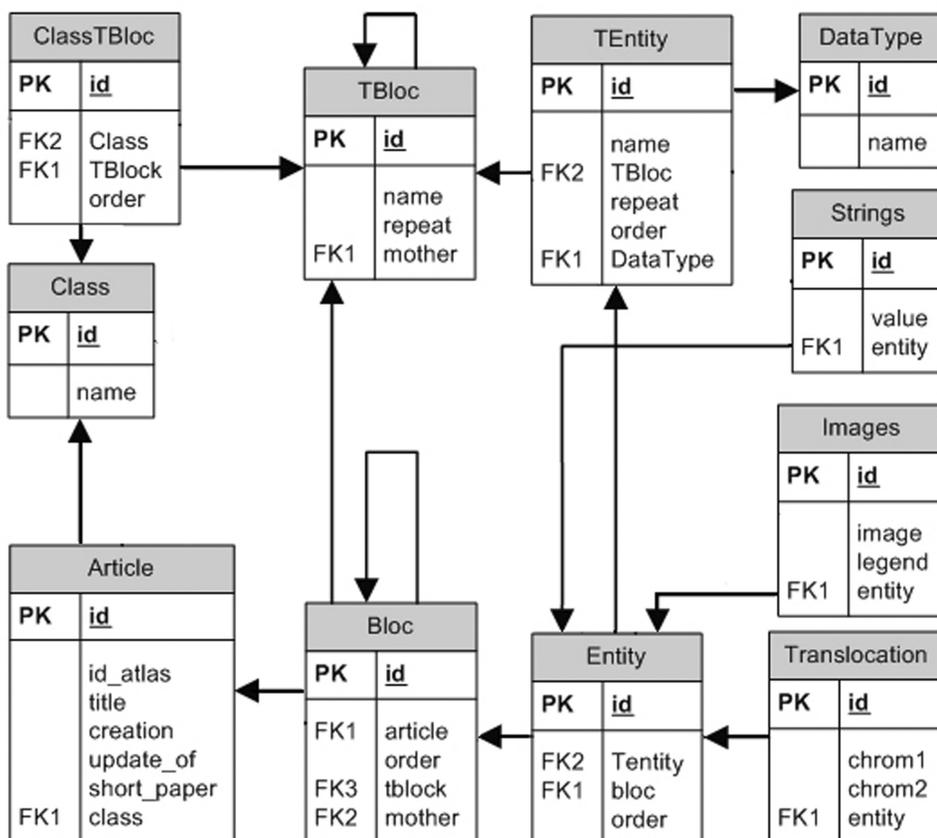

Figure 2. A simplified representation of the Atlas database. Structures are described in tables Tbloc, Tentity and datatype. The contents of the articles are stored in Tables Block, Entity, Images and so forth. DataType contains a list of data types that are used to create an article, such as text, images, links and so forth. The data are stored in a specific table for each type. Tentity describes the entities that make up a block. TBloc describes the structuring topic and sub-topic articles. The articles are classified into different categories (leukaemia, gene and so forth), each class having its own section structure. To create a new category of articles, the publishing tool will search for any structure that must be implemented by querying the database metadata (TBloc, TEntite, DataType).

FUTURE DIRECTIONS

Description of recurrent anomalies, gene fusions and other cytogenomic-acquired alterations

The main goal of the Atlas is to describe the recurrent anomalies in the human genome implicated in oncological processes, the genes involved and the various diseases in which they are implicated. This description, which was historically based on karyotypic abnormalities and *in situ* hybridization (fluorescence *in situ* hybridization) techniques, now benefits from comparative genomic hybridization and massive sequencing. The field of cytogenomic oncology has exploded with copy number, single-nucleotide polymorphism and loss of heterozygosity techniques and next-generation genome sequencing, uncovering a tremendous amount of acquired genomic alterations (polymorphisms, mutations, copy number alterations, rearrangements, duplications, inversions, translocations, with gene fusions or re-locations), confirming/revealing the extreme complexity of cellular alterations at the origin of cancer and during cancer progression. The result is a massive amount of new data. All of these data are currently scattered across various public databases or may be found exclusively in publications.

The Atlas plans to browse these data from the Mitelman database (12) http://cgap.nci.nih.gov/Mitelman/ and the Cancer Genome Project (13) and COSMIC (as a reference for mutations) (14) to integrate all these data sets and make them available by chromosomal band. The Atlas also intends to include the location of diagnostic probes.

Orphan chromosome anomalies

A chromosomal abnormality in a leukaemia patient is said to be a 'driver' anomaly (driving to carcinogenesis), either when a known oncogene is implicated in the rearrangement or when at least two cases have been described with the same breakpoints. Consequently, all the laboratories have 'single' orphan cases supposedly 'passenger' anomalies, pending on the back burner. An 'orphan chromosomal abnormalities' module will provide an area where cytogeneticists will be able to deposit the description of yet unknown anomalies, with complete clinical, cytological and cytogenetic data. As soon as a second case is deposited alongside a first case in the repository, the anomaly will become *ipso facto* a 'driver' anomaly, warranting further studies and scientific communications.

Scientific and professional societies

Finally, the Atlas has so far relied on institutional funds. It recently launched a public appeal for donation and grants because of cash flow problems, and some individuals as well as scientific societies responded positively. A further step we would welcome is that scientific and professional societies in genetics, cancer genetics and oncology areas become involved in the scientific (and financial) long-term development of the Atlas.

ELECTRONIC ADDRESSES

http://AtlasGeneticsOncology.org Atlas of Genetics and Cytogenetics in Oncology and Haematology Home Page.

http://chromosomesincancer.org The non-profit association ARMGHM's Home Page, whose goal is to host and handle the Atlas and receive donations.

http://documents.irevues.inist.fr/handle/2042/15655 Archives of the Atlas Journal.

CITING THE ATLAS

If you use the Atlas of Genetics and Cytogenetics in Oncology and Haematology in your published research, please, cite this article.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Huret,J.L., Minor,S.L., Dorkeld,F., Dessen,P. and Bernheim,A. (2000) Atlas of genetics and cytogenetics in oncology and haematology, an interactive database. *Nucleic Acids Res.*, **28**, 349–351.
2. Pearson,H. (2001) Lifelines: browsing the cancer catalogue. *Nature Science Updates*, 31 May 2001.
3. Kaiser,J. (2001) Fingerprinting a killer. *Science*, **292**, 1803.
4. Huret,J.L., Dessen,P. and Bernheim,A. (2003) Atlas of Genetics and Cytogenetics in Oncology and Haematology, year 2003. *Nucleic Acids Res.*, **31**, 272–274.
5. Huret,J.L., Dessen,P. and Bernheim,A. (2003) An internet database on genetics in oncology. *Oncogene*, **22**, 1907.
6. Maglott,D., Ostell,J., Pruitt,K.D. and Tatusova,T. (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
7. Dreszer,T.R., Karolchik,D., Zweig,A.S., Hinrichs,A.S., Raney,B.J., Kuhn,R.M., Meyer,L.R., Wong,M., Sloan,C.A., Rosenbloom,K.R. *et al.* (2012) The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.*, **40**, D918–D923.
8. Mehrpour,M., Botti,J. and Codogno,P. (2012) Mechanisms and regulation of autophagy in mammalian cells. *Atlas Genet. Cytogenet. Oncol. Haematol.*, **16**, 165–182.
9. Hoyer-Fender,S. (2012) Centrosomes in fertilization, early embryonic development, stem cell division, and cancer. *Atlas Genet. Cytogenet. Oncol. Haematol.*, **16**, 306–319.
10. Chen,J.M., Férec,C. and Cooper,D.N. (2012) Chromothripsis: a new molecular mechanism in cancer development. *Atlas Genet. Cytogenet. Oncol. Haematol.*, **16**, 381–384.

11. Bracha,G. and Ungar,D. (2004) Mirrors: design principles for meta-level facilities of object-oriented programming languages. *Proceeding OPSLA '04 Proceedings of the 19th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*, Vol. 39, ACM SIGPLAN Notices. ACM New York, NY, USA. pp. 331–334.

12. Mitelman,F., Johansson,B. and Mertens,F. (eds). (2012) *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer*, http://cgap.nci.nih.gov/Chromosomes/Mitelman (15 November 2012, date last accessed).

13. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.

14. Forbes,S.A., Bindal,N., Bamford,S., Cole,C., Kok,C.Y., Beare,D., Jia,M., Shepherd,R., Leung,K., Menzies,A. *et al.* (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.*, **39**, D945–D950.