Progenetix.net: an online repository for molecular cytogenetic aberration data

Michael Baudis [1, 2,*] and Michael L. Cleary [2]

[1]*Medizinische Klinik und Poliklinik V der Universität Heidelberg, Germany and*
[2]*Department of Pathology, Stanford University Medical Center, Stanford, CA 94305, USA*

ABSTRACT

Summary: Through sequencing projects and, more recently, array-based expression analysis experiments, a wealth of genetic data has become accessible via online resources. In contrast, few of the (molecular-) cytogenetic aberration data collected in the last decades are available in a format suitable for data mining procedures. www.progenetix.net is a new online repository for previously published chromosomal aberration data, allowing the addition of band-specific information about chromosomal imbalances to oncologic data analysis efforts.
Availability: http://www.progenetix.net
Contact: mbaudis@stanford.edu

Neoplastic transformation and progression is the result of genetic defects arising in normal cells and giving rise to a malignant clone. During the process of oncogenesis, some of the usually multiple steps required for acquisition of the full neoplastic phenotype may represent themselves as numerical or structural abnormalities in the chromosomes of the transformed cells.

Over the last decades, the analysis of chromosomal abnormalities in malignant cells has gained importance in oncologic research as well as in clinical practice. A vast number of genetic abnormalities has been identified in the virtually complete range of human neoplasias. Several attempts have been undertaken for collection and classification of those abnormalities, the most widely recognized being the catalog by Mitelman and co-workers (Mitelman, 1994; online access through http://cgap.nci.nih.gov/Chromosomes/Mitelman).

In addition to metaphase analysis of short-term cultivated tumor cells or tumor cell lines, molecular cytogenetic techniques have recently been applied to the analysis of chromosomal abnormalities in primary tumor tissues. One of the more widely used screening techniques is Comparative Genomic Hybridization (CGH; Kallion-iemi *et al.*, 1992; du Manoir *et al.*, 1993). Briefly, this method is based on the competitive *in-situ* hybridization of differentially labeled tumor versus normal genomic DNA to normal human metaphase spreads. The calculation of the intensity ratios of the two fluorochromes gives an overview about relative gains and losses of DNA in the tumor genome with mapping to the respective chromosomal bands. The identification of frequently imbalanced regions in tumor entities may point towards tumor suppressor gene or proto-oncogenes mapping to the respective chromosomal bands. Usually, the result of those experiments is communicated either in text format according to the International System for Cytogenetic Nomenclature (Mitelman, 1995) or graphically, with aberration bars next to chromosomal ideograms for the representation of chromosomal gains and losses.

Because in each experiment CGH analysis covers the whole number of chromosomes, the comparision of data sets from related malignancies could lead to the delineation of common as well as divergent genetic pathways defining the respective malignant phenotypes. Although an extremely large number of malignant tumors has been analyzed using this technique, no comprehensive CGH database with band-specific chromosomal aberration information is publicly available[†].

A minimal requirement for such a database would be the conversion of the text or graphical information used in publications to data tables, representing the information about the aberration status of single chromosomal bands for each case. For the site discussed here, this process includes: (1) the transformation of the published results in a format adapted from the ISCN, and (2) the automatic generation of the band specific aberration table.

Due to format variations of the published data, step 1 consists of the manual conversion of the text data or evaluation and conversion of the graphical representations, respectively. Due to the (in computational terms) odd

---

*To whom correspondence should be addressed.

[†] Links to a number of online CGH resources with different scopes can be found at www.progenetix.net.

conventions of the cytogenetic nomenclature, a complex string transformation algorithm had to be developed for the second step (data not shown). Evaluation and conversion of the datasets is performed 'offline', with upload of the expanded dataset following each update.

For addition of a set of CGH data to the *progenetix.net* site, two basic criteria have to be fulfilled. First, the included results have to consist of previously published original data. Second, the presentation of the results has to allow for a band specific assignment of the chromosomal imbalances on a case-by-case basis.

Due to their clinical and pathological highly variable appearance and the number of publications in this area, malignant lymphomas were selected as a starting point for the data collection. The first 11 publications included data from 250 different malignant lymphoma species and covered a broad range of histological subtypes (Bentz *et al.*, 1995, 1996; Joos *et al.*, 1996; Monni *et al.*, 1996; Dierlamm *et al.*, 1997; Autio *et al.*, 1998; Barth *et al.*, 1998; Siu *et al.*, 1999; Bentz *et al.*, 2000; Weber *et al.*, 2000; Franke *et al.*, 2001; for updates see www. progenetix.net[‡]). Using this data set, a number of different presentation modes were tested:

(1) a list of the included projects with links to the corresponding PubMed entries;

(2) the list of all cases including the respective chromosomal aberrations in an ISCN adapted format;

(3) a bar chart summarizing the chromosomal imbalances on each chromosomal arm for all included cases, as well as the same kind of representation for all single projects and ICD-O-3 entities;

(4) the band-specific graphical representation of imbalances, summarizing data from all projects for each chromosome; with an additional list naming the abberant cases per chromosomal band; the cases were linked to the descriptive entries of the respective projects.

These presentation modes are intended to give users a fast overview of the underlying data. A table containing the aberration status of each chromosomal band (324 bands resolution) is available for download. This data format allows the inclusion of the cytogenetic data in data mining projects, such as linkage to expression data, or application of statistical procedures such as cluster analysis.

For the future, the extension of the dataset to non-hematological malignancies as well as the integration of active lookup procedures is intended.

[‡] On 7 July 2001, the dataset consists of 490 cases from 19 publications. Additional to malignant lymphoma, data from lymphoblastic leukemias and myeloid neoplasias have been added.

REFERENCES

Autio,K., Aalto,Y. *et al.* (1998) Low number of DNA copy number changes in small lymphocytic lymphoma. *Haematologica*, **83**, 690–692.

Barth,T.F., Dohner,H. *et al.* (1998) Characteristic pattern of chromosomal gains and losses in primary large B-cell lymphomas of the gastrointestinal tract. *Blood*, **91**, 4321–4330.

Bentz,M., Huck,K. *et al.* (1995) Comparative genomic hybridization in chronic B-cell leukemias shows a high incidence of chromosomal gains and losses. *Blood*, **85**, 3610–3618.

Bentz,M., Plesch,A. *et al.* (2000) t(11;14)-positive mantle cell lymphomas exhibit complex karyotypes and share similarities with B-cell chronic lymphocytic leukemia. *Genes Chromosomes Cancer*, **27**, 285–294.

Bentz,M., Werner,C.A. *et al.* (1996) High incidence of chromosomal imbalances and gene amplifications in the classical follicular variant of follicle center lymphoma. *Blood*, **88**, 1437–1444.

Dierlamm,J., Rosenberg,C. *et al.* (1997) Characteristic pattern of chromosomal gains and losses in marginal zone B-cell lymphoma detected by comparative genomic hybridization. *Leukemia*, **11**, 747–758.

du Manoir,S., Speicher,M.R. *et al.* (1993) Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization. *Hum. Genet.*, **90**, 590–610.

Franke,S., Wlodarska,I. *et al.* (2001) Lymphocyte predominance Hodgkin disease is characterized by recurrent genomic imbalances. *Blood*, **97**, 1845–1853.

Joos,S., Otano-Joos,M.I. *et al.* (1996) Primary mediastinal (thymic) B-cell lymphoma is characterized by gains of chromosomal material including 9p and amplification of the REL gene. *Blood*, **87**, 1571–1578.

Kallioniemi,A., Kallioniemi,O.-P. *et al.* (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, **258**, 818–821.

Mitelman,F. (1994) *Catalog of Chromosome Aberrations in Cancer*. Wiley–Liss, New York.

Mitelman,F. (1995) *ISCN (1995): an International System for Human Cytogenetic Nomenclature*. Karger, Basel.

Monni,O., Joensuu,H. *et al.* (1996) DNA copy number changes in diffuse large B-cell lymphoma–comparative genomic hybridization study. *Blood*, **87**, 5269–5278.

Siu,L.L., Wong,K.F. *et al.* (1999) Comparative genomic hybridization analysis of natural killer cell lymphoma/leukemia. Recognition of consistent patterns of genetic alterations. *Am. J. Pathol.*, **155**, 1419–1425.

Weber,T., Weber,R.G. *et al.* (2000) Characteristic chromosomal imbalances in primary central nervous system lymphomas of the diffuse large B-cell type. *Brain Pathol.*, **10**, 73–84.