BIOINFORMATICS ORIGINAL PAPER Vol. 22 no. 17 2006, pages 2066–2073
doi:10.1093/bioinformatics/btl359

Genome analysis

VAMP: Visualization and analysis of array-CGH, transcriptome and other molecular profiles

Philippe La Rosa[1,*], Eric Viara[1], Philippe Hupé[1,2], Gaëlle Pierron[3], Stéphane Liva[1], Pierre Neuvial[1], Isabel Brito[1], Séverine Lair[1], Nicolas Servant[1], Nicolas Robine[1,4], Elodie Manié[3], Caroline Brennetot[3], Isabelle Janoueix-Lerosey[3], Virginie Raynal[3], Nadège Gruel[3], Céline Rouveirol[2], Nicolas Stransky[2], Marc-Henri Stern[3], Olivier Delattre[3], Alain Aurias[3], François Radvanyi[2] and Emmanuel Barillot[1]

[1]Institut Curie, Service Bioinformatique, 26 rue d'Ulm, Paris, 75248 cedex 05, France, [2]Institut Curie, CNRS UMR 144, 26 rue d'Ulm, Paris, 75248 cedex 05, France, [3]Institut Curie, INSERM U509, 26 rue d'Ulm, Paris, 75248 cedex 05, France and [4]Institut Curie, CNRS, Université Pierre et Marie Curie UMR 7147, 26 rue d'Ulm, Paris, 75248 cedex 05, France

Received on January 11, 2006; revised on May 31, 2006; accepted on June 25, 2006
Advance Access publication July 4, 2006
Associate Editor: Nikolaus Rajewsky

ABSTRACT

Motivation: Microarray-based CGH (Comparative Genomic Hybridization), transcriptome arrays and other large-scale genomic technologies are now routinely used to generate a vast amount of genomic profiles. Exploratory analysis of this data is crucial in helping to understand the data and to help form biological hypotheses. This step requires visualization of the data in a meaningful way to visualize the results and to perform first level analyses.

Results: We have developed a graphical user interface for visualization and first level analysis of molecular profiles. It is currently in use at the Institut Curie for cancer research projects involving CGH arrays, transcriptome arrays, SNP (single nucleotide polymorphism) arrays, loss of heterozygosity results (LOH), and Chromatin ImmunoPrecipitation arrays (ChIP chips). The interface offers the possibility of studying these different types of information in a consistent way. Several views are proposed, such as the classical CGH karyotype view or genome-wide multi-tumor comparison. Many functionalities for analyzing CGH data are provided by the interface, including looking for recurrent regions of alterations, confrontation to transcriptome data or clinical information, and clustering. Our tool consists of PHP scripts and of an applet written in Java. It can be run on public datasets at http://bioinfo.curie.fr/vamp

Availability: The VAMP software (Visualization and Analysis of array-CGH,transcriptome and other Molecular Profiles) is available upon request. It can be tested on public datasets at http://bioinfo.curie.fr/vamp. The documentation is available at http://bioinfo.curie.fr/vamp/doc

Contact: vamp@curie.fr

1 INTRODUCTION

Array Comparative Genome Hybridization (array-CGH) is a recently developed technology based on DNA microarrays (Pinkel *et al*., 1998; Snijders *et al*., 2001; Solinas-Toldo *et al*., 1997; Ishkanian *et al*., 2004) that can be used to investigate DNA copy number differences between two samples. A CGH array-generally consists of spotted clones of genomic sequences (e.g. bacterial artificial chromosomes) that cover part or all of the genome. Both DNA samples are labeled with distinct fluorescent dyes and undergo competitive hybridization onto the CGH array. The array is then scanned with a scanner or a CCD camera, and the acquired image is analyzed (gridding, spot addressing, spot segmentation, spot quantification, outlier detection), normalized (to remove as much as possible any systematic spatial or intensity biases, e.g. Neuvial *et al*., (2005), duplicate statistical analysis is then carried out (each clone is generally spotted in several copies), and adequate statistical algorithms detect any loss or gain regions (Hupé *et al*., 2004; Olshen *et al*., 2004; Fridlyand *et al*., 2004; Jong *et al*., 2003; Picard *et al*., 2005; Eilers and de Menezes, 2005; Bilke *et al*., 2005). CGH arrays are often used in cancer research because chromosome aberrations are thought to be causal in tumor progression (Albertson *et al*., 2003; Pinkel and Albertson, 2005). Here, normal DNA is used as reference and the test sample would be tumoral biopsy DNA. The normal sample has two copies of each genomic region, whereas tumor DNA may show losses or gains in certain DNA regions. Measurement of the signal intensities of the reference and tumor samples for each clone makes it possible to determine the lost or gained regions in the tumor sample. Further analyses can include the determination of recurrent loss or gain of DNA regions, clustering of samples and determination of candidate oncogenes and candidate tumor suppressor genes within the altered regions (based on their annotations or on their transcription level). It is also possible to link array-CGH results to the clinical phenotype or to biological parameters through, for example, supervised classification or correlation analysis. The visualization of the data is a crucial step in the analysis procedure and is essential for hypothesis formulation and model-free reasoning. We have developed, in the framework of large-scale array-CGH projects, a graphical user interface that allows several visualization modes of the CGH profiles and offers several data analysis tools. The software also displays a large variety of genomic profiles, such as transcriptome,

*To whom correspondence should be addressed.

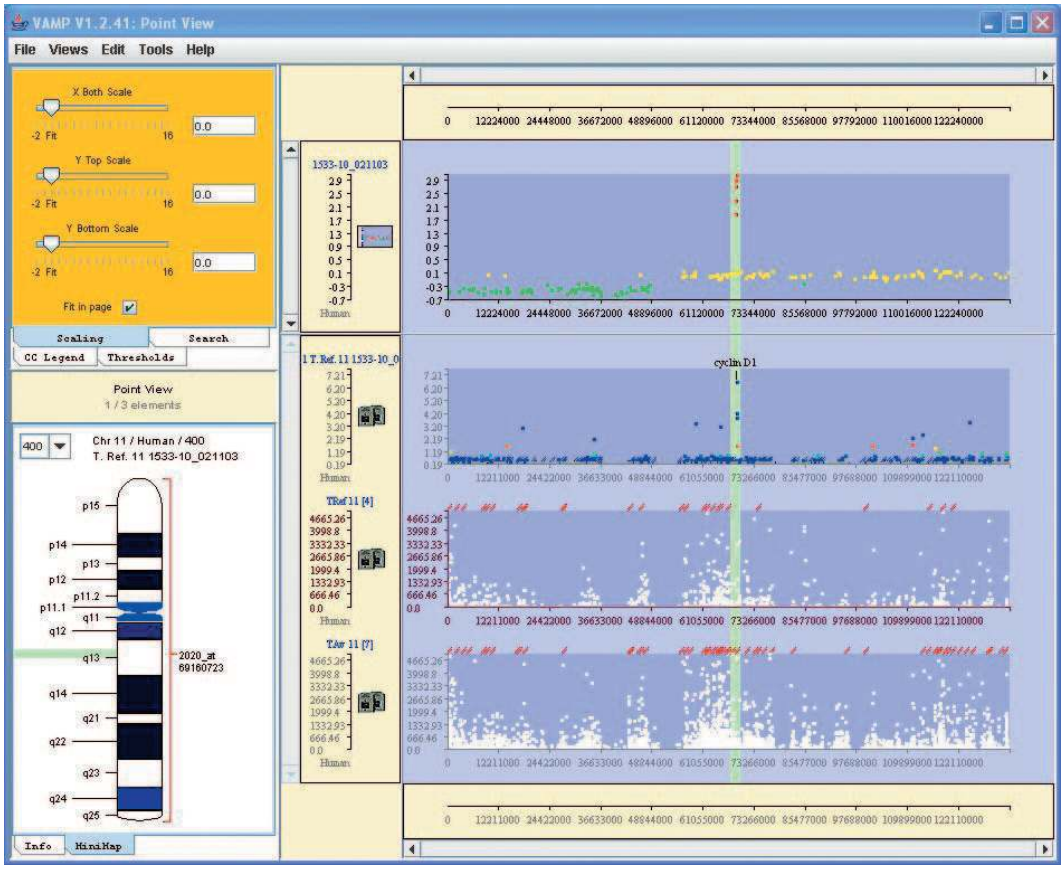

Fig. 1. Array-CGH (top profile) versus transcriptome ratio (second profile in descending order), computed for Affymetrix U95 array of a bladder tumor sample and of a reference sample. This confrontation pinpoints the probable implication of the oncogene cyclin D1 in this tumor. The third and fourth profiles in descending order correspond to a reference profile (average normal bladder tissue profile) and the profile of the tumor under study, respectively. The second profile is the ratio of the fourth to the reference profile.

Loss Of Heterozygosity (LOH), Vogelstein *et al*. (1989), Single nucleotide polymorphism (SNP) arrays (Bignell *et al*., 2004; Huang *et al*., 2004) and ChIP chip [Chromatin ImmunoPrecipitation coupled with microarrays, Buck and Lieb (2004)] profiles and allows addition of new tools for data treatment or analysis. We have called the software VAMP for 'Visualization and Analysis of Molecular Profiles'. In this article we first detail how data are visually presented in VAMP, and then we explain how the user interacts with the software and which functionalities are offered for data analysis. Finally, we describe the software architecture of VAMP.

2 RESULTS

2.1 Data representation

VAMP was designed to graphically represent any genomic profile along the genome axis. We started the development of VAMP for array-CGH data, but we have extended it to accept, on the same window, any kind of profile. We currently use the software for expression arrays, SNP arrays, LOH results and ChIP chip profiling, in addition to array-CGH. VAMP is currently used for three species (human, mouse and yeast) but the addition of a new species is straightforward. It is possible to visualize simultaneously, on the same window, different types of profiles for a given species, e.g. array-CGH and mRNA expression profiles of a tumor (Fig. 1). All profiles in a window are drawn on the *x*-axis with the same scale (the genome sequence), which allows an easy comparison of profiles.

A typical VAMP window is divided into three areas (Fig. 2): the main frame consists of the graphical display of the profiles; the top left frame controls zoom, search and drawing options; the bottom left frame offers the choice between textual information (Fig. 3) on the object under the mouse pointer, or context information, called MiniMap (Fig. 2).

2.1.1 Main frame VAMP currently offers several types of visualization that can be displayed in the main frame: (1) List View, (2) Profile View (Fig. 2) (3) Karyotype View (Fig. 3), (4) Dot Plot View (Fig. 4). These views all allow simultaneous visualization of several profiles (the only limitation is the memory size of the computer running VAMP, or more precisely, the memory allocated to the Java virtual machine: for example with an 800 Mb Java virtual machine memory, 700 microarrays (each with 3500 probes) can be loaded simultaneously).

- List View: the List View lists the names of all the arrays currently loaded and can be used for selecting or keeping track of the data under study.

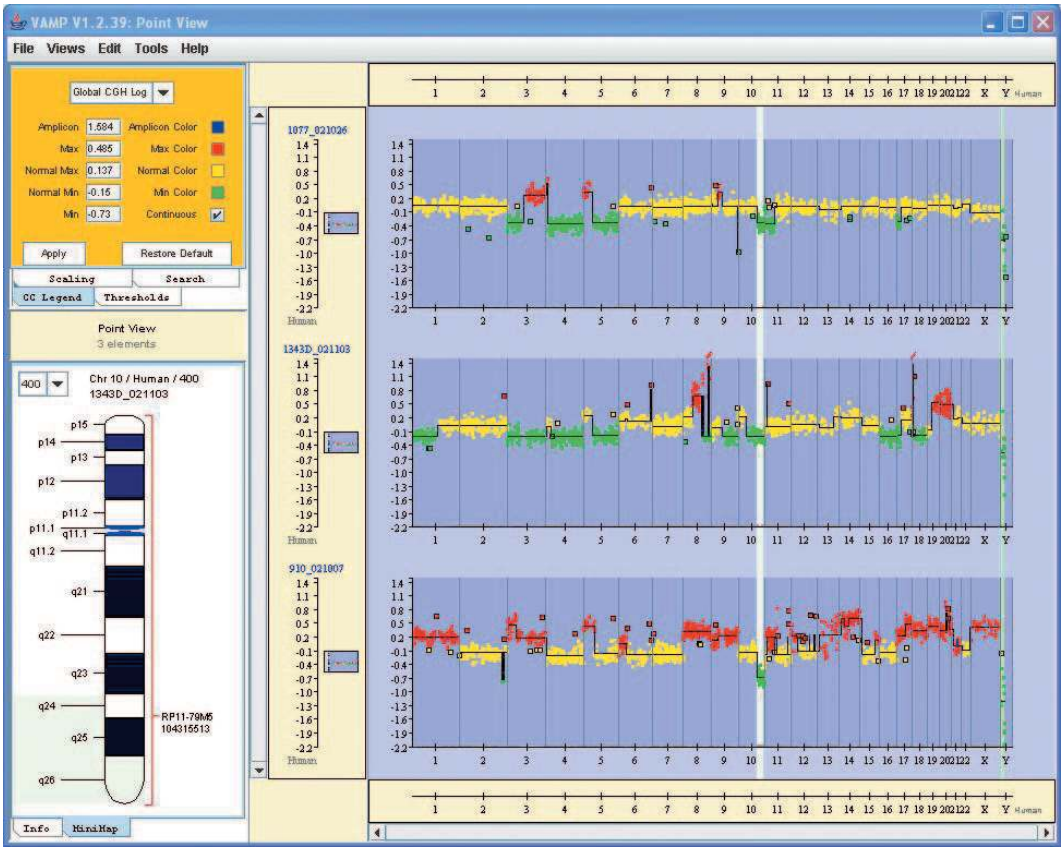

Fig. 2. Genomic View, main frame: profiles along all the concatenated chromosomes; top left: zoom control, search and drawing options; bottom left: textual information on the object under the mouse pointer or (in this figure) chromosome context information (MiniMap). The regions spanning the three tumors highlighted in green are those that are lost in all tumors (short arm of chromosome 10, and Y chromosome); these are called minimal regions.

- Profile View: the Profile View (Fig. 2) can display the profiles as points, barplots or curves. It can be split into two frames, as in Figure 1. The upper frame can, for example, contain a profile for reference when browsing a collection of profiles in the lower frame. The two frames have separate control of Y-scale and Y-scrolling, but have the same X-scale and X-scrolling. The Profile View can also display symbols for chromosome telomeres and centromeres, and can show the results of CGH ratio statistical analysis (e.g. breakpoints, or smoothed signal values, see Fig. 2).
- Karyotype View: the Karyotype View (Fig. 3) displays profiles having the well-known classical CGH rendering: vertical representations of chromosomes with cytogenetic banding and contiguous representation of sample profiles.
- Dot Plot View: the Dot Plot View does not consider the microarray probe positions on the genome, but only their ranks. It displays a collection of samples as a heat map based on the level of signal for each probe (Fig. 4).

By default, points or barplots are colored according to the signal intensity (generally using ratios of the two channels or log-ratios) using a continuous scale from red to yellow to green. All the previously mentioned views for the CGH data can be colored as a function of the array-CGH data analysis. Typically, gained DNA regions are displayed in red, lost regions in green, amplicons in blue and normal in yellow.

Whatever view is chosen, the profiles can be represented in Genomic mode or Chromosome mode. The Genomic mode simply depicts the profiles along all the concatenated chromosomes. It is the most usual representation, and allows comparison of profiles from different samples or comparison of different types of profiles from a given sample. The Chromosome mode is similar to the Genomic mode except that it only displays one particular chromosome. It is also possible to merge several chromosomes and to represent those chromosomes useful for the study.

- New Views: our object-oriented architecture easily allows us to add new types of views that can be associated with particular actions or data processing. For example, the Minimal Region functionality is associated with a particular type of view. Therefore, when profiles are pasted in the window, the Minimal Region View automatically displays the array-CGH profiles with the DNA regions recurrently lost or gained in the samples (Fig. 2).

2.1.2 Top left frame This frame controls zoom, search and drawing options. Zooming is independent on *X* and *Y* axes, and all profiles in the same window have the same zoom control, except

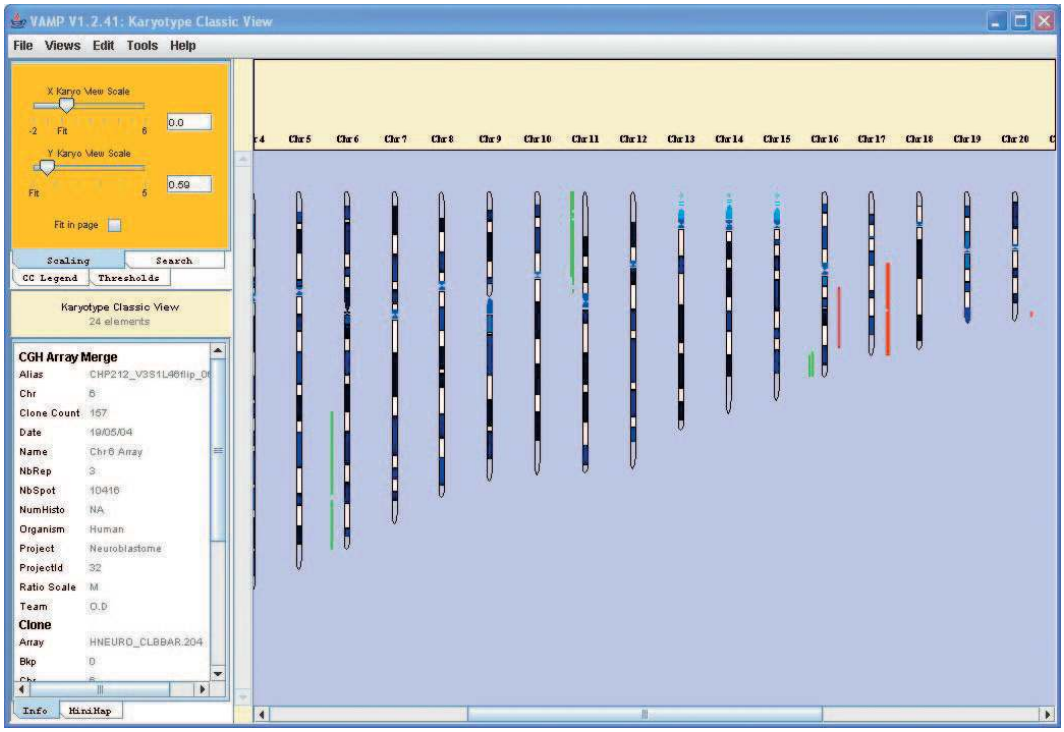

Fig. 3. Karyotype View, classic rendering of CGH data, loss regions in green, gain in red.

for *Y* zooming of the reference profile. The search can be carried out on any property attached to the arrays or the clones/probes held in an XML (eXtended Markup Language) data file or in the database (see Fig. 6 and the Software architecture presentation below). For XML data files, the list of properties is not limited,but is established at run time, leading to a very flexible search option. Drawing options include color-coding for signal values, and the threshold values to be applied; they can be either global to the application or restricted to one profile (local). User preferences can be saved on your computer in a XML configuration file.

2.1.3 Bottom left frame (Object information and context frame) The bottom left frame can either display textual information on the object under the mouse pointer (Fig. 3) or context information, called MiniMap (Fig. 2). The textual information consists of mandatory fields (object genomic position, signal value, project name, organism and data type) and any other type of complementary information stored in the XML data file. For example, in array-CGH profiles we currently display general information about the clone under the mouse pointer (name, chromosome, number of valid replicates, rank and position on the sequence, signal ratio and standard deviation, size of the clone, CGH status—gain/lost/normal) as well as information about the array (name, number of spots, number of clones, number of replicates, chromosomes covered, ratios or log-ratios) and information about the sample (sample id, project name, date). MiniMap is a special view type that gives some context on what the user is examining in the main frame: (1) a cytogenetic representation of the chromosome under the mouse pointing, with (2) a rule delimiting the region of the chromosome displayed on the main frame and (3) the name and position of the object (array-CGH clone, transcriptome microarray probe, etc.) under the mouse pointer. In this view, the display can be automatically updated when the user moves the mouse.

2.2 User interaction

All user actions are accessible either through a Menu on the menubar, or through pointing to or clicking objects. When using VAMP, the session can be saved in local XML files. Reloading the file later on allows the continuation of the analysis within the context of the previous work, or allows the exchange of results and data with colleagues. All user preferences can also be stored in local XML files. Drag and drop capability is offered for any profile, from one window to any other window, the rendering being automatically adapted (e.g. from a dot plot view to a karyotype view). An advanced printing function is offered, either in visible mode (only the profiles that are visible on the screen are printed), or in global mode (all profiles in the view are printed). A template is offered for defining the output of the printing (this can, for example, include several frames in an arbitrary composition, to which text or images can be added). It can be used for defining and printing standardized outputs. The user can also interactively monitor the print preferences.

2.3 Data analyses

VAMP allows addition of any new piece of software for data analysis and visualization of the results. Several functionalities have already been implemented either as plug-ins or within the VAMP Java source code. VAMP was initially developed for the analysis of CGH-arrays of tumoral samples. As VAMP is actually an interface, it is assumed that the microarray data have already been normalized, and also, for CGH data, that breakpoints have been established and

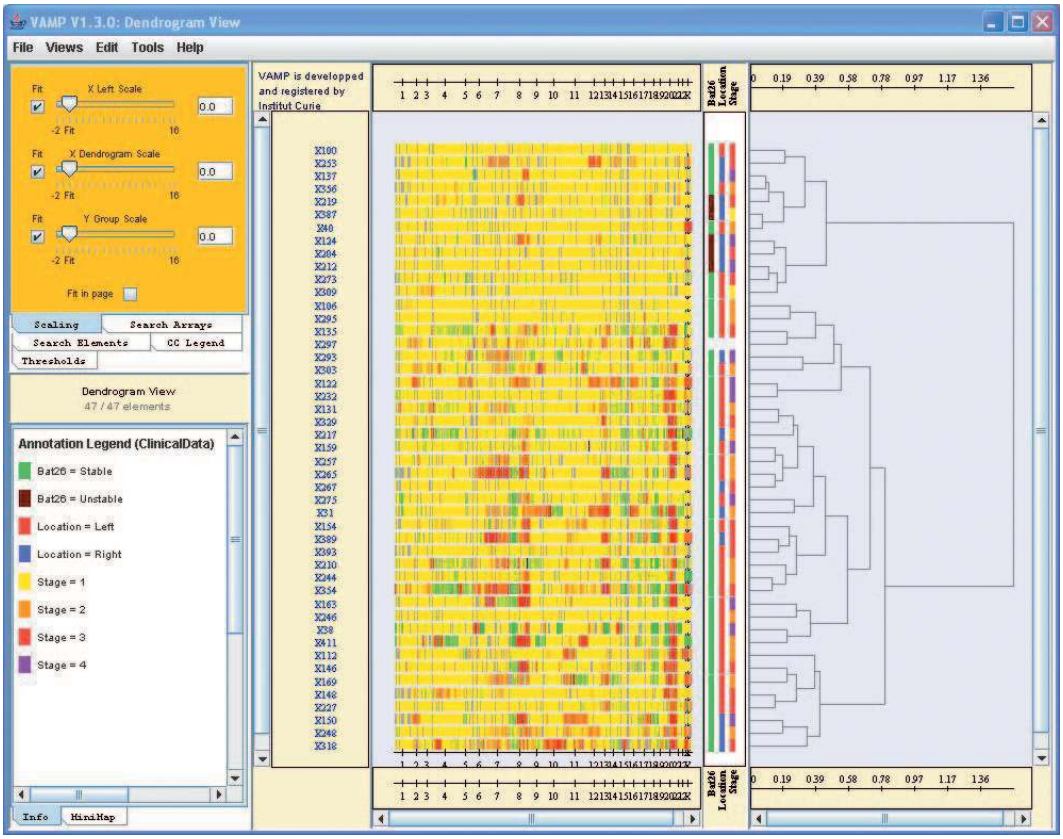

Fig. 4. VAMP interface, dotplot view of array-CGH profiles (middle panel), and dendrogram resulting from a hierarchical clustering (right panel). In between, color-coded clinical information about the samples, with a legend (bottom left). Data from Nakao *et al*. (2004).

regions of DNA loss or gain inferred. VAMP can then display in the profile frame (Fig. 2) the breakpoint positions, the status of each region (by default, green for loss, yellow for normal, red for gain, blue for amplicons), and the estimation of the signal value in each region, which is computed, for example, using smoothing techniques (Hupé *et al*., 2004). VAMP also allows the defining of the gain and loss regions by simply applying a threshold to the signal ratios. Examples of data analyses available within VAMP are given below and are described in more detail in the software documentation (http://bioinfo.curie.fr/vamp/doc).

Finding common alterations among a collection of CGH- array profiles. CGH array analysis principally consists in finding common regions of alterations, i.e. regions that are lost in many tumors. It is essential in these studies to distinguish between recurrent and random alterations. Recurrent alterations pinpoint regions involved in tumoral progression, whereas random alterations are simply the consequence of the general instability that affects the genome of a tumor. Among the recurrent alterations we distinguish the minimal regions and the recurrent regions. Minimal regions are extracted by intersecting the profiles of many tumors and looking for a sufficient number of alterations in the tumors (this parameter is set by the user) over the smallest possible region of the profile (Fig. 2). Tumoral progression obeys a selection principle, and it would be expected that the genes that need to be altered for a cell to become tumoral must be located in the smallest possible intersection of all alterations of a region. Recurrent regions are defined differently: in a given tumor, an alteration is bounded by two extremities, which can be a breakpoint or a chromosome end; when a sufficient number of tumors have the same extremities, these extremities define a recurrent region. We have implemented a linear algorithm that detects such minimal and recurrent regions, which is described in (Rouveirol *et al*., 2006). Gained regions appear in red in the main frame, and lost regions appear in green (Fig. 2). Amplicons (defined as gained regions with signal-ratio above a threshold typically equal to two) are colored in blue. The tumors that support a region of alteration may be optionally shadowed in the region, and for each region the user can sort these tumors.

Clustering profiles. Clustering is a general technique for unsupervised data classification widely used in microarray data analysis. A VAMP function offers the possibility to perform a hierarchical clustering (Kaufman and Rousseuw, 1990) on the profiles in the dot plot view. This can cluster genes and tumors from transcriptome arrays, or tumors from a CGH profile. In a CGH profile, the clustering uses the smoothed values of the CGH profile as variables and the Euclidean distance and Ward method for group distance computation. VAMP displays the results as a cluster view including a heat map and the trees resulting from the clustering algorithm (Fig. 4).

Comparing profiles. The Menu proposes several different data manipulation procedures for the profiles such as loading any type of

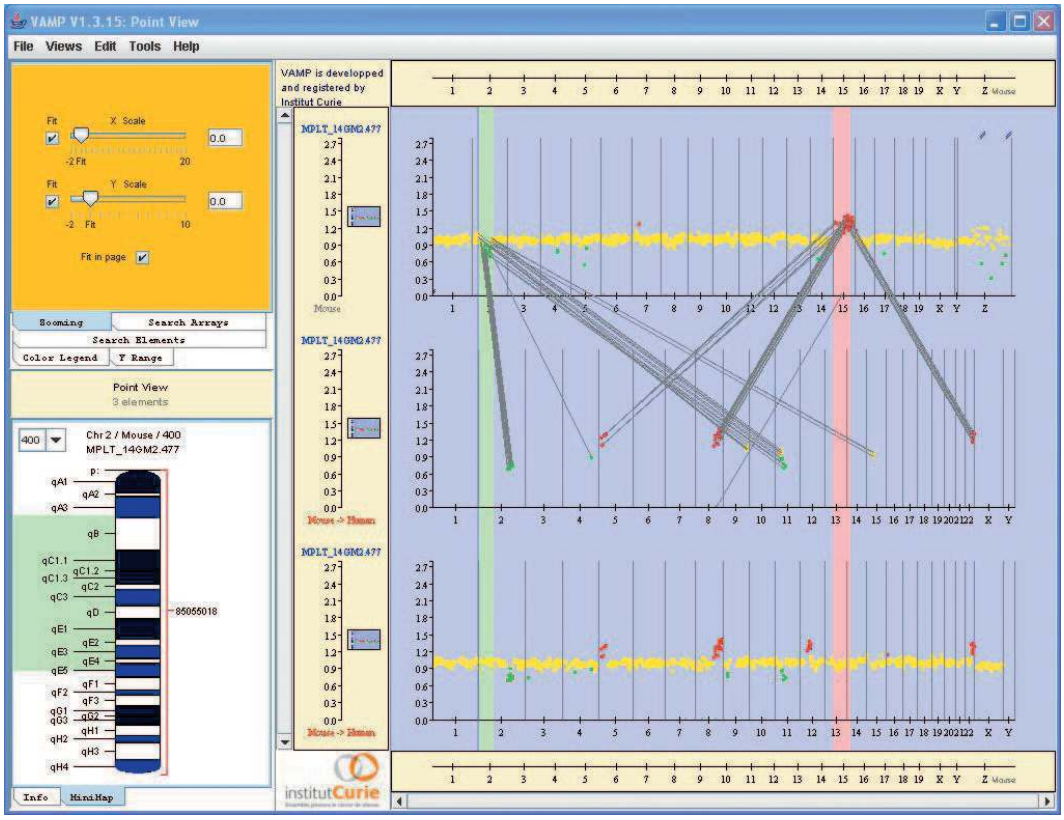

Fig. 5. Array-CGH profile for a mouse tumor (top) and its syntenic projection, i.e. a humanized array-CGH profile after mapping each mouse clone onto the human genome (bottom) and projection for two regions (middle profile) with resulting synteny relationships. Mapping is done from each clone of the mouse profile onto the location of the most similar sequence of the human genome. Mouse clones with ambiguous syntenic locations have not been mapped onto the human genome.

profile (CGH, expression, LOH, ChIP chip—an icon at the left of each profile shows the type of loaded profile) for a given sample (e.g. a typical application of VAMP is the simultaneous visualization of the DNA alterations and gene under- and over-expression in a region, Fig. 1); defining a profile as a reference and calculating the ratio of a profile to the reference (useful for one-color microarrays such as Affymetrix); averaging profiles; drawing marks (vertical bars) or regions (such as the green regions in Fig. 2) across all profiles (and simultaneously on the MiniMap); and many others.

Confrontation with sample annotation. Clinical data, or any other sample annotations, present in the XML files can be used for filtering tumors or for sorting them. This data can be visualized as color-coded bars in an annotation frame on the left of the profiles, and can be easily compared with a clustering result (Fig. 4).

Synteny analysis. VAMP can display the syntenic projection of a profile onto the genome of another species, in which that genome serves as a reference; a typical application is the projection of a mouse array-CGH profile onto the human genome (Fig. 5). In our case if an unambiguous syntenic locus was found, the mapping was done from each clone of the mouse profile onto the location of the most similar sequence of the human genome. The synteny relationships can be shown, for a selection of regions of the genome, as links from each clone of the profile to the location of the most similar sequence of the reference genome.

Other functions. The right mouse button brings up a menu with several actions associated to the clone/probe currently under the mouse pointer. These include: centering the profile around the current position; drawing of a vertical bar through all the profiles (to define a locus or a region); and linking to external web pages from NCBI clone or MapViewer (http://www.ncbi.nlm.nih.gov/mapview and Wheeler *et al*., 2005), UCSC Genome Browser (http://genome.ucsc.edu and Kent *et al*., 2002), Ensembl Contig View or CytoView (http://www.ensembl.org and Hubbard *et al*., 2005), *Saccharomyces* Genome Database (http://www.yeastgenome.org). New links are defined in a XML configuration file and adding them is straightforward. Most data and results (profiles, minimal regions, etc.) can be exported and saved in full text, csv (comma separated values) or HTML format. We refer the reader to the user manual for a description of the other functions.

2.4 Software architecture and requirements

The software architecture is shown in Figure 6. The core of the interface consists of a Java applet, and was developed using the Swing library. It runs on any operating system supporting Java 1.4.2

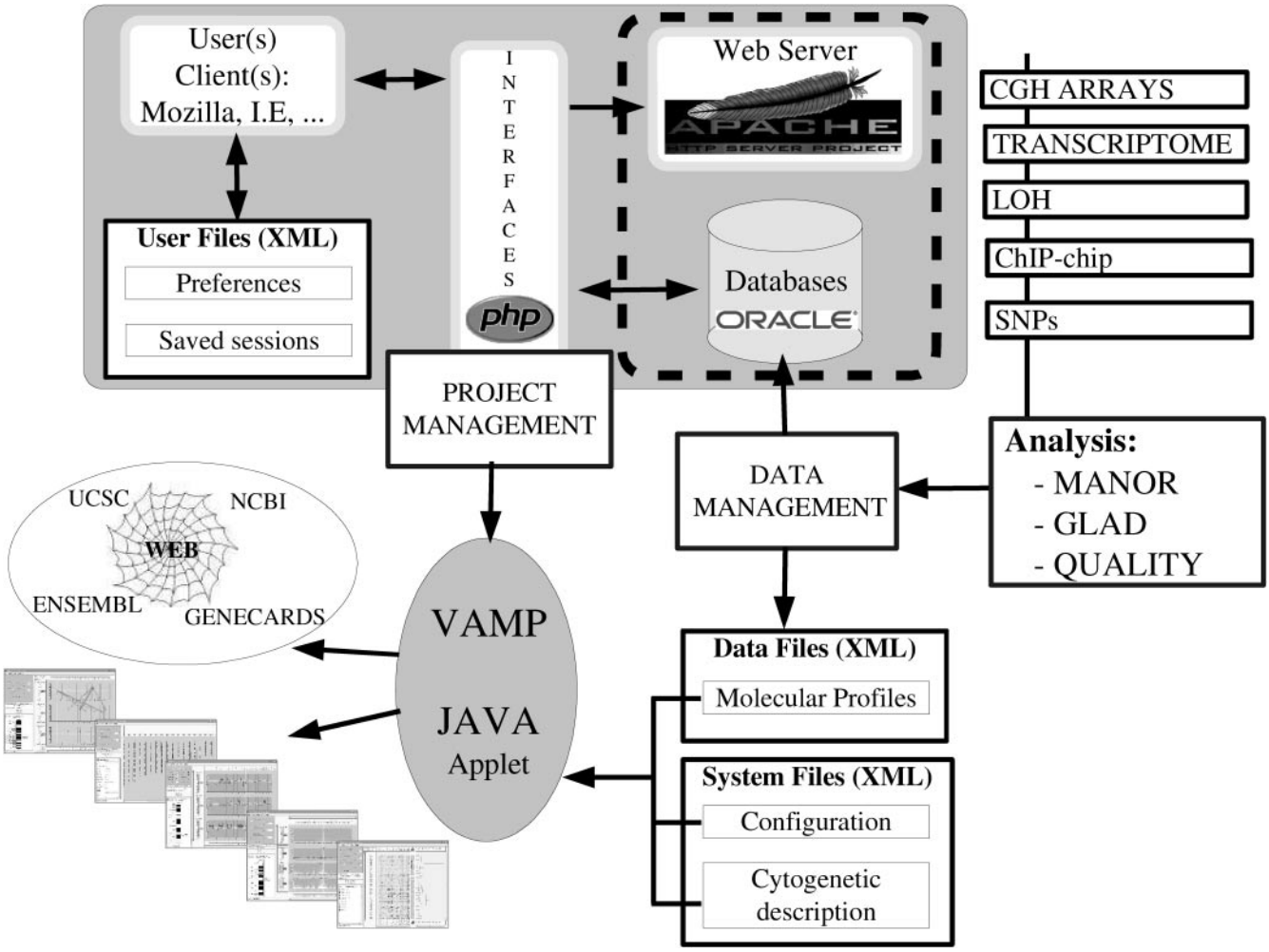

Fig. 6. Software architecture of a microarray environment based on VAMP. VAMP can also be used as a local application.

(we recommend computers with a minimum of 1 Gb memory, although 256 Mb is enough for small projects). The data used by the program are of several types:

- The genome profile information, which are retrieved either from a relational database management server (currently Oracle™) or from XML data files. These include the signal value for each clone/probe and its genomic location.
- The system files (also in XML), which includes the cytogenetic description of the genome under study and the configuration parameters (environment variables for file and URL management). Cytogenetic banding files for human ISCN 400, 550 and 850 descriptions, as well as mouse and yeast genome descriptions are also available. The user files, which consist of the user visualization preferences and saved sessions.

VAMP can be used either as a local application, with all data and configuration files directly accessible to the client, or as an applet, with all data and configuration files installed on a server. In this mode, only the user configuration file is stored locally on the client machine.

VAMP can be easily installed on any platform running Java 1.4.2. All that is needed is to convert the microarray data into XML files, with a specific syntax described in a DTD (XML Document Type Definition). The use of a database management server is not mandatory, although it is recommended for large-scale projects. Arbitrary complementary profile information can be added to the XML files, and this information can be displayed by the interface.

3 DISCUSSION

We have developed a graphical user interface for the visualization and analysis of any type of genomic profile, with an emphasis on array-CGH. VAMP is currently used in cancer genomic projects on human and mouse samples and in studying the proteins involved in the reparation, recombination and replication of DNA in yeast. It is used in Institut Curie and many labs in Europe and the United States. Several publications describing data analysis with VAMP are coming soon. Janoueix-Lerosey *et al*., (2005) describe the use of VAMP for replication timing data analysis (http://microarrays.curie.fr/publications/U509/reptiming). In Institut Curie, ∼3600 microarray profiles have been interfaced with VAMP to date.

VAMP aids greatly in finding genes of clinical and biological importance from CGH, transcriptome, LOH, ChIP chip profiles and SNP arrays. VAMP improves upon existing solutions such as SeeGH (Chi *et al*., 2004), CGHPRO (Chen *et al*., 2005), CGH-Analyzer (Margolin *et al*., 2005) or general purpose spreadsheet software, because it offers many different modes of visualization, allows the display of several samples and of several types of profiles simultaneously, and offers many data analysis functions. VAMP can be compared with other general-purpose genomic browsers such MapView (NCBI), Genome Browser of UCSC or Ensembl. VAMP is well suited to handle sample profiles and to analyse this type of data, which the other genomic browsers are not designed to do. Therefore, in cancer research it addresses a real need and is a useful tool for biologists and clinicians. Our software is fully portable and only requires a computer running Java 1.4.2 and data in XML format.

VAMP can be run on public datasets at http://bioinfo.curie.fr/vamp. The array-CGH data from Snijders *et al*. (2001, 2005), Pollack *et al*. (2002), Veltman *et al*. (2003), Nakao *et al*. (2004), Douglas *et al*. (2004), de Leeuw *et al*. (2004), Gysin *et al*. (2005), Patil *et al*. (2005) and Bredel *et al*. (2005) are currently browsable. Expression profiles are also available for the samples from Pollack *et al*. (2002).

ACKNOWLEDGEMENTS

This work was supported by the Institut Curie, the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the CNRG and the Ligue contre le Cancer.

Conflict of Interest: none declared.

REFERENCES

Albertson,D.G. *et al*. (2003) Chromosome aberrations in solid tumors. *Nat. Genet*., **34**, 369–76.

Bignell,G.R. *et al*. (2004) High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res*., **14**, 287–295.

Bilke,S. *et al*. (2005) Detection of low level genomic alterations by comparative genomic hybridization based on cDNA micro-arrays. *Bioinformatics*, **21**, 1138–1145.

Bredel,M. *et al*. (2005) High-resolution genome-wide mapping of genetic alterations in human glial brain tumors. *Cancer Res*., **65**, 4088–4096.

Buck,M.J. and Lieb,J.D. (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*, **83**, 349–360.

Chen,W. *et al*. (2005) CGHPRO—a comprehensive data analysis tool for array CGH. *BMC Bioinformatics*, **6**, 85.

Chi,B. *et al*. (2004) SeeGH—a software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics*, **5**, 13.

de Leeuw,R.J. *et al*. (2004) Comprehensive whole genome array CGH profiling of mantle cell lymphoma model genomes. *Hum. Mol. Genet*., **13**, 1827–1837.

Douglas,E.J. *et al*. (2004) Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res*., **64**, 4817–4825.

Eilers,P.H.C. and de Menezes,R.X. (2005) Quantile smoothing of array CGH data. *Bioinformatics*, **21**, 1146–1153.

Fridlyand,J. *et al*. (2004) Application of hidden markov models to the analysis of the array CGH data. *J. Multivari. Anal*. (Special Issue on Multivariate Methods in Genomic Data Analysis), **90**, 132–153.

Gysin,S. *et al*. (2005) Analysis of genomic DNA alterations and mRNA expression patterns in a panel of human pancreatic cancer cell lines. *Genes Chromosomes Cancer*, **44**, 37–51.

Huang,J. *et al*. (2004) Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics*, **1**, 287–299.

Hubbard,T. *et al*. (2005) Ensembl 2005. *Nucleic Acids Res*., **33**, 447–453.

Hupé,P. *et al*. (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*, **20**, 3413–3422.

Ishkanian,A.S. *et al*. (2004) A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet*., **36**, 299–303.

Janoueix-Lerosey,I. *et al*. (2005) Preferential occurrence of chromosome breakpoints within early replicating regions in neuroblastoma. *Cell Cycle*, **4**, 1842–1846.

Jong,K. *et al*. (2003) Chromosomal breakpoint detection in human cancer. In Raidl,G.R., Cagnoni,S., Cardalda,J.J.R., Corne,D.W., Gottlieb,J., Guillot,A., Hart,E., Johnson,C.G., Marchiori,E., Meyer,J.-A. and Middendorf,M. (eds), *Applications of Evolutionary Computing, EvoWorkshops2003: EvoBIO, EvoCOP, EvoIASP, EvoMUSART, EvoROB, EvoSTIM,* vol. 2611 of *LNCS*. Springer-Verlag, University of Essex, England, UK.

Kaufman,L. and Rousseuw,P. (1990) *Finding Groups in Data—An Introduction to Cluster Analysis*, Wiley Series in Probability and Mathematical Sciences. John Wiley & Sons.

Kent,W.J. *et al*. (2002) The human genome browser at UCSC. *Genome Res*., **12**, 996–1006.

Margolin,A. *et al*. (2005) CGHAnalyzer: a stand-alone software package for cancer genome analysis using array-based DNA copy number data. *Bioinformatics*, **21**, 3308–3311.

Nakao,K. *et al*. (2004) High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis*, **25**, 1345–1357.

Neuvial,P., Hupé,P., Brito,I., Liva,S., Manié,E., Brennetot,C., Radvanyi,F., Aurias,A. and Barillot,E. (2005) Spatial normalization of array-CGH data. *BMC Bioinformatics*, **7**, 264.

Olshen,A.B. *et al*. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, **5**, 557–572.

Patil,M.A. *et al*. (2005) Array-based comparative genomic hybridization reveals recurrent chromosomal aberrations and Jab1 as a potential target for 8q gain in hepatocellular carcinoma. *Carcinogenesis*, **26**, 2050–2057.

Picard,F. *et al*. (2005) A statistical approach for array CGH data analysis. *BMC Bioinformatics*, **6**, 27.

Pinkel,D. and Albertson,D.G. (2005) Array comparative genomic hybridization and its applications in cancer. *Nat. Genet*., **37** (Suppl.1), 11–17.

Pinkel,D. *et al*. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet*., **20**, 207–211.

Pollack,J.R. *et al*. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl Acad. Sci. USA*, **99**, 12963–12968.

Rouveirol,C. *et al*. (2005) Computation of recurrent minimal genomic alterations from CGH data. *Bioinformatics*, **22**, 849–856.

Snijders,A.M. *et al*. (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet*., **29**, 263–4.

Snijders,A.M. *et al*. (2005) Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. *Oncogene*, **24**, 4232–4242.

Solinas-Toldo,S. *et al*. (1997) Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, **20**, 399–407.

Veltman,J.A. *et al*. (2003) Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res*., **63**, 2872–2880.

Vogelstein,B. *et al*. (1989) Allelotype of colorectal carcinomas. *Science*, **244**, 207–11.

Wheeler,D.L. *et al*. (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*., **33**, 39–45.