Visualising Complex Omics Information

Scientific Visualisation for Genomics and Systems Biology

Scientific visualisation of complex information is a fundamental means to reinforce the heuristic reasoning of research scientists, in the process of data interpretation and hypothesis generation. This is especially critical for emerging disciplines where central theories are yet to be structured; therefore, scientific visualisation for molecular biology in the post-genome era, which is rapidly reshaped into data-driven science, is an area of growing importance in computational biology.

Kazuharu Arakawa,
PhD, Keio University

Nozomu Yachie, Keio
University

Masaru Tomita, Professor, Keio University

Data Presentation Premised on User Interaction

An aspect that makes clear distinction of data in molecular biology from other large-scale subjects is the highly multi-layer nature of information, and the fact that these layers are also highly multi-scale. We look at a cell from different layers of omics (i.e. genome, transcriptome, proteome, and metabolome) where each layer has unique set of molecular entities and interrelations, and it is also essential to simultaneously study the specifics as well as the workings of the whole in each of these layers. In the E-Cell project (http://www.e-cell.org), an international collaboration aiming to realise precise dynamic cell simulation at whole-cell level, our current approach to this challenge is a software system that allows multi-timescale (e.g. micro-seconds for diffusion, milli-seconds for enzymatic reactions, minutes for gene expression, hours for cell division) and multi-algorithm (e.g. ordinary differential equations for enzymatic reactions, stochastic calculation for gene expression) simulation [1–2].

In scientific visualisation, utilisation of interactive user interface is a conventional solution in order to display multi-layer/multi-scale data, transforming the graphical presentation upon user interaction. With the success of Google Maps (http://maps.google.com/) and Google Earth (http://earth.google.com/) for displaying geographical information,

Zoomable User Interface (ZUI) has been rapidly gaining attention as a feasible technique for the visualisation of large-scale data, including those from molecular biology. The use of ZUI is recently being explored in several pathway database projects, such as the KEGG Atlas [3] and Reactome [4]. Coupling ZUI with tabbed browsing for multiple views, we have been developing a freely accessible web-based genome browser designated Genome Projector (http://www.g-language. org/GenomeProjector) [5]. Genome Projector is a searchable database browser with four zoomable views (circular genome map, genome map, Roche biochemical pathway map, and DNA walk map) currently containing 320 circular bacterial genomes, developed using Asynchronous JavaScript and XML (AJAX) technology with Google Maps API

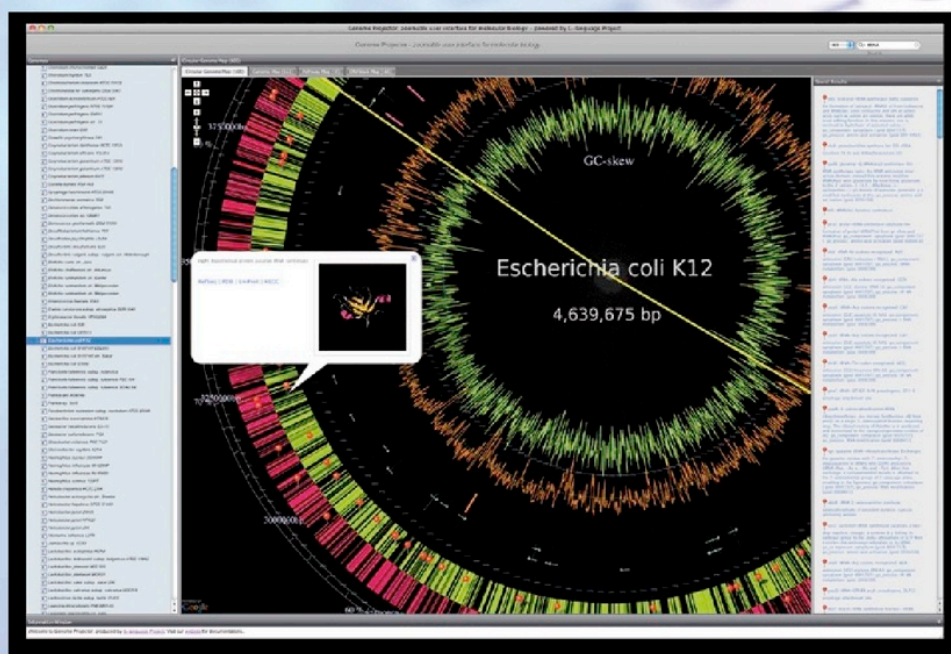

Fig. 1: Screenshot of Genome Projector. Circular genome map of Escherichia coli is searched with the keyword "tRNA", and detailed annotation is shown for a putative tRNA synthetase gene with 3D structure image obtained from PDB.

(fig. 1). Each view provides a biological context for the search query; therefore, users can quickly find out how the genes of interest are located within the chromosome, how they are expressed, what role they play in the metabolic pathway, and what is the nucleotide composition around them, through rapid tabbed-browsing and zooming. Information can be further overlaid on top of each view, and currently Genome Projector has overlays of gene expression data on genome map, and highlighting of organism specific information on pathway map (fig. 2). Unlike traditional genome browsers, user interaction through responsive interface allows researchers to quickly search through multitude of information in numerous contexts and scales.

3D Visualisation of Cell Simulation

Cell simulation in silico is typically monitored with graphs tracing the time-course progression of reaction velocities or molecular concentrations. Although this approach is acceptable for the observation of several specific entities of interest, it is not scalable for the purpose of understanding the workings of the whole, requiring the observation of tens or sometimes hundreds of graphs at once for complex multi-scale models. Modelling process is a current bottleneck in systems biology researches, and effective visualisation that assists the heuristics of systems biologists can help expedite this process [6–7]. In order to solve this task, we adopted a 3D visualisation approach in a software named E-Cell 3D (http://ecell3d.iab.keio.ac.jp/) [8]. E-Cell 3D takes advantage of the latest 3D graphics API such as Quartz and OpenGL available on MacOS X, and visualises the components and the reaction networks among the components in 3-dimensional spaces (fig. 3). Here any model in Systems Biology Markup Language (SBML) or E-Cell format is automatically laid out in 3D-space, where molecular concentrations are represented by the size and brightness (changing from blue to yellow) of the nodes, and the flux is represented by the speed of particles travelling along the edges that connect the molecular entities. Since every component of a model is present in this visualisation, system biologists can capture the
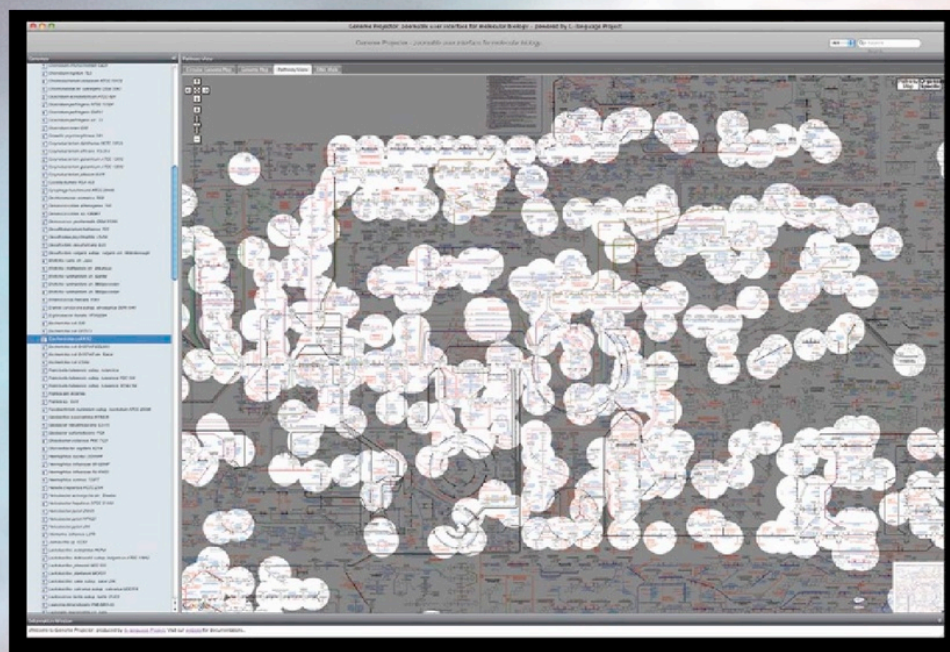

Fig. 2: Example of overlay in Genome Projector. Here the organism specific pathway is highlighted in the Roche biochemical pathway map. Similarly, gene expression level can be overlaid on the genome map.
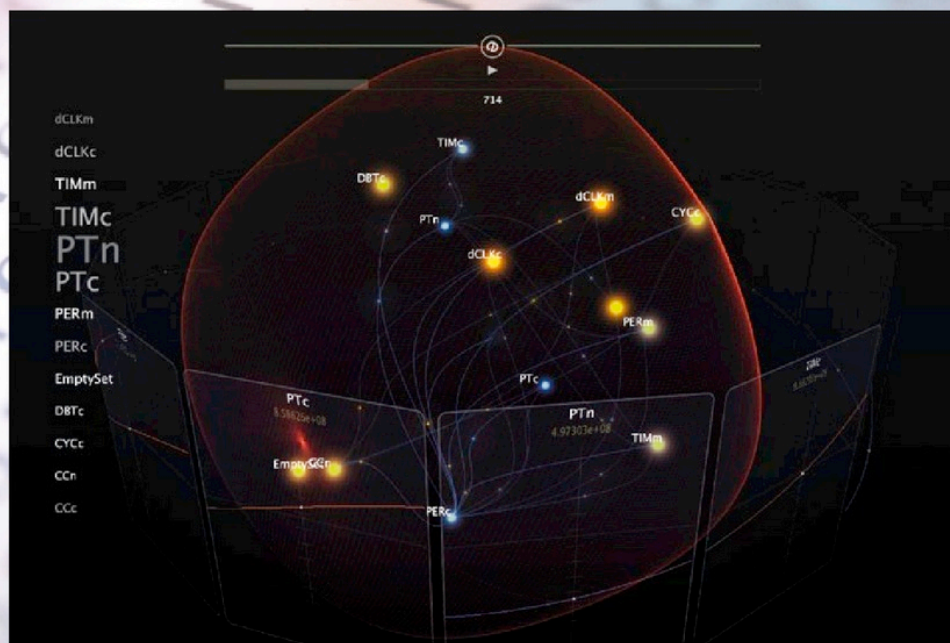

Fig. 3: E-Cell 3D giving visualisation for a simulation of Drosophila circadian clock model [9].

dynamic workings of the entire system at a glance. Moreover, when there is large amount of flux in or out of a node, the coordinates of the node are altered, allowing rapid identification of controlling factors within the cellular network. In highly dynamical models such as the circadian clock, we can clearly observe the synchronisation of oscillations within the network (see screen capture videos available at http://ecell3d.iab.keio.ac.jp/

features.html#section5), and identify the controlling molecules of feedback oscillators (fig. 4). Here again, user interaction is a key factor for visualisation. Using mouse and keyboard or motion sensor remote controllers such as the Nintendo Wii Remote with 3-axis accelerometers, users can take full control of the 3D graphics by rotating, zooming, and showing traditional graphs for the molecular concentrations.
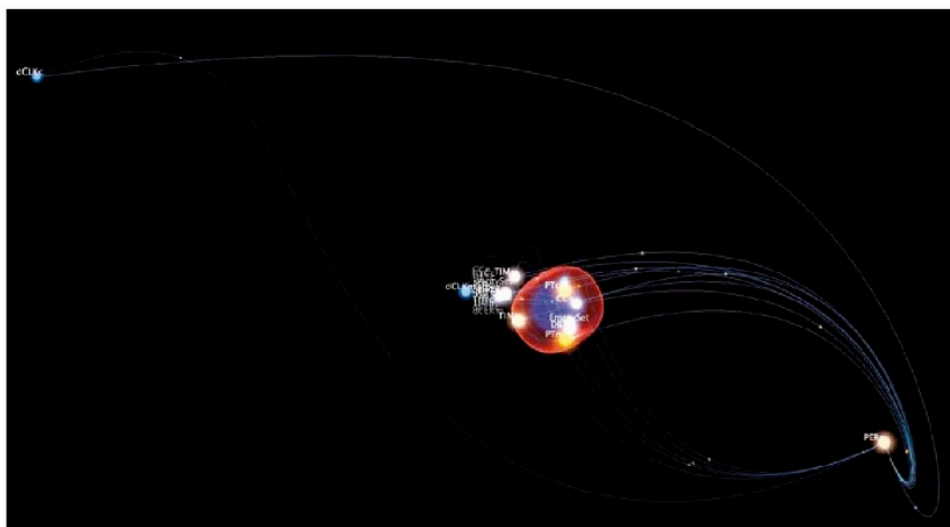
Fig. 4: By changing the coordinates of molecules with large amount of influx or efflux, rate controlling entities can be identified. The two interlocked feedback loops of Drosophila circadian clock are correctly visualised.

[4] Vastrik I. *et al.:* Genome Biology 8, R39 (2007)

[5] Arakawa K.: Bioinformatics Open Source Conference (2007)

[6] Arakawa K. *et al.:* BMC Bioinformatics 7, 168 (2006)

[7] Arakawa K. *et al.:* Journal of Pesticide Science 31, 282–288 (2006)

[8] Arakawa K. *et al.:* Intelligent Systems for Molecular Biology, (2007)

[9] Ogawa Y. *et al.:* Artificial Life 14, 29–48 (2008)

For further information on the research referred to in this article, publications, detailed references, and activities of the research groups, see www.iab.keio.ac.jp.

▶ www.eMagazineBIOforum.com

Data explosion in biology seem to continue at an accelerated rate in the near future, with the introduction of next-generation sequencers and other numerous high-throughput measurement technologies. Combined with analytical methods of bioinformatics for hypothesis generation by data mining and hypothesis testing through computer simulation, scientific visualisation is an effective and necessary tool to understand the complex biological information.

References

[1] Takahashi K. *et al.:* Bioinformatics 20, 538–546 (2004)

[2] Tomita M. *et al.:* Bioinformatics 15, 72–84 (1999)

[3] Kanehisa M. *et al.:* Nucleic Acids Research 36, D480–484 (2008)

CONTACT:

Kazuharu Arakawa, PhD
Institute for Advanced Biosciences
Keio University, Japan
Tel.: +81 466 47 5099
Fax: +81 466 47 5099
gaou@sfc.keio.ac.jp
www.iab.keio.ac.jp

RESEARCH & DEVELOPMENT

Newly Released Study of Human Genetic Structural Variation

Agencourt Bioscience, a Beckman Coulter company, announced that its Genomic Services operation performed the genome sequencing in a groundbreaking new study on human genetic variation. The work is published in an article, "Mapping and Sequencing of Structural Variation from Eight Human Genomes," in the May 1 issue of the journal *Nature*.

The study used a novel sequence-based approach to achieve higher accuracy than previous work that used signal intensity differences to predict regions of variation. Agencourt worked closely with Evan Eichler of the Howard Hughes Medical Institute and University of Washington Department of Genome Sciences to pioneer this sequence-based approach. Ten Agencourt scientists are among the lead authors of the study, which was funded by the National Human Genome Research Institute.

www.beckmancoulter.com
www.agencourt.com

Heparin Contaminant Can Cause Reactions

A biological link between a contaminant recently found in some lots of heparin and allergy-like reactions including deaths has been clarified in articles in *The New England Journal of Medicine* and the journal *Nature Biotechnology*, both of which are available online. The study in swine reported in *The New England Journal* found a means by which the contaminant, oversulfated chondroitin sulfate (OSCS), can cause the severe allergy-like reactions. The reactions in humans have included a sudden drop in blood pressure, difficulty breathing and abdominal distress. So far, the contaminant has been reported in 12 countries worldwide.

www.nigms.nih.gov
www.fda.gov

Reconstructing Oetzi's Family Tree with 454 Sequencing

Since Oetzi's sensational discovery, in 1991 and his extraction from the glacier in Alto Adige in Italy, Europe's most famous mummy has been the subject of intensive scientific study. Currently, Dr. Gianluca De Bellis of the CNR Institute of Biomedical Technology is taking part in new investigations promoted by Prof. Franco Rollo of the University of Camerino. De Bellis is analysing Oetzi's mitochondrial DNA by means of ultrafast 454 sequencing. In using the Genome Sequencer FLX System, it will now be possible to shed more light on the subject of Oetzi's descendents, and establish his place in the genetic scenario of the present European race. Furthermore, 454 Sequencing data will enable researchers to trace Oetzi's ancestors; and reconstruct an accurate phylogenetic family tree, obtaining an authentic global view of human evolution.

www.roche-applied-science.com/sis/sequencing

Nurture Over Nature

Score one for the nurture side of the nature *vs.* nurture debate, as North Carolina State University geneticists have shown that environmental factors such as lifestyle and geography play a large role in whether certain genes are turned on or off. By studying gene expression of white blood cells in 46 Moroccan Amazighs, or Berbers – including desert nomads, mountain agrarians and coastal urban dwellers – the NC State researchers and collaborators in Morocco and the United States showed that up to one-third of genes are differentially expressed due to where and how the Moroccan Amazighs live. The research is published in the journal PLoS Genetics.

www.news.ncsu.edu

Trojan Horse of Viruses Revealed – a Step Forward in Virology

Viruses use various tricks and disguises to invade cells. ETH Zurich researchers have now discovered yet another strategy used by viruses: the vaccinia virus disguises itself as cell waste, triggers the formation of evaginations in cells and is suspected to enter the cell interior before the immune defense even notices. The research results have been published in Science. The vaccinia virus has a problem: it is a giant among viruses and needs a special strategy in order to infiltrate a cell and reproduce. Prof. Ari Helenius and Postdoc Jason Mercer from ETH Zurich's Institute for Biochemistry have now discovered what this strategy is. In the process, they stumbled upon new and surprising findings.

www.fm-eth.ethz.ch