Perhaps the most interesting aspect of these studies is the finding that interchromosomal interactions are not found in just any cell type, but specifically in naive T cells. This is important because naive helper T cells have the potential to transcribe both IFNg and the Th2 cytokine gene locus upon activation. Subsequently, helper T cells differentiate into highly specialised cells with distinct functions in the immune system, which express either IFNg or the Th2 cytokine gene locus, but not both.[10] These differentiated Th1 and Th2 cells no longer have the potential to coexpress both loci, and the interchromosomal interactions between IFNg and the Th2 cytokine gene locus are lost to be replaced by new intrachromosomal contacts (Figure 1). The authors point out the exciting possibility that genes may colocalise in the nucleus in readiness for their coordinated expression. In support of this idea, Flavell and colleagues demonstrate that removing the RHS-7 sequence from the Th2 cytokine gene locus compromises its interaction with the IFNg locus and – most significantly – the expression of the interaction partner, IFNg.

The new findings raise several burning questions. Those of us who work on developmentally regulated genes will want to find out how general the new findings are. Will they apply to other genes poised for coordinated expression? If so, will interchromosomal interactions change when one or the other of the interacting partners is excluded from expression during development? Will the authors'

interpretation that colocalisation prepares genes for coordinated expression hold up for other interacting genes, or will interchromosomal interactions turn out to have other functions? There are suggestions that they will. For example, human cells contain several hundred ribosomal (rDNA) genes distributed over five chromosome pairs, and rDNA genes from several chromosomes merge to form a single nucleolus.[11] Inactive genes from different chromosomal addresses have also been found together, in association with polycomb proteins[12] or in the vicinity of centromeric heterochromatin.[13] The most important question concerns the mechanisms that set up and disperse these interchromosomal interactions. If we can define the molecules involved, we will be in a great position to understand more about the way in which patterns of gene expression are established and progressively modified during development ∎

V Parelho and M Merkenschlager are at the Lymphocyte Development Group, MRC Clinical Sciences Centre, Du Cane Road, London W12 ONN, UK.
Tel: +44 (0) 20 8383 8239;
Fax: +44 (0) 20 8383 8338;
E-mail: matthias.merkenschlager@ csc.mrc.ac.uk

References

1 Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA: Interchromosomal associations between alternatively expressed loci. *Nature* 2005; **435**: 637–645.

2 Cook PR: The organization of replication and transcription. *Science* 1999; **284**: 1790–1795.

3 Haaf T, Schmid M: Chromosome topology in mammalian interphase nuclei. *Exp Cell Res* 1991; **192**: 325–332.

4 Dillon N, Sabbattini P: Functional gene expression domains: defining the functional unit of eukaryotic gene regulation. *BioEssays* 2000; **22**: 657–665.

5 West AG, Fraser P: Remote control of gene transcription. *Hum Mol Genet* 2005; **14**: (Spec. No. 1): R101–R111.

6 Dekker J, Rippe K, Dekker M, Kleckner N: Capturing chromosome conformation. *Science* 2002; **295**: 1306–1311.

7 Lee GR, Spilianakis CG, Flavell RA: Hypersensitive site 7 of the TH2 locus control region is essential for expressing TH2 cytokine genes and for long-range intrachromosomal interactions. *Nat Immunol* 2005; **6**: 42–48.

8 Vakoc CR, Letting DL, Gheldof N *et al*: Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 2005; **17**: 453–462.

9 Osborne CS, Chakalova L, Brown KE *et al*: Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 2004; **36**: 1065–1071.

10 Murphy KM, Reiner SL: The lineage decisions of helper T cells. *Nat Rev Immunol* 2002; **2**: 933–944.

11 Leung AK, Lamond AI: The dynamics of the nucleolus. *Crit Rev Eukaryot Gene Expr* 2003; **13**: 39–54.

12 Bantignies F, Grimaud C, Lavrov S, Gabut M, Cavalli G: Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. *Genes Dev* 2003; **17**: 2406–2420.

13 Brown KE, Guest SS, Smale ST, Hahm K, Merkenschlager M, Fisher AG: Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin. *Cell* 1997; **91**: 845–854.

Research Networks

● ●

BioSapiens: a European network for integrated genome annotation

The BioSapiens Network of Excellence
....

The BIOSAPIENS Network of Excellence exists to provide infrastructure to support laboratories across Europe in annotating genome data using both bioinformatics tools and experimental data. The annotations generated by the network will be made available in the public domain and easily accessible on the web.

The first draft of the human sequence, published in 2001, was followed by other related genomes and the detailed resequencing of the human genome. This

explosion in genomic information was achieved in a remarkably short time period. However, a DNA sequence must be interpreted in terms of the RNA and proteins that it encodes, and the promoters and regulatory regions that control transcription and translation. A genomic sequence also provides a convenient 'coordinate' reference onto which functional information can be mapped.

Annotation can be described as the process of 'defining the biological role of a molecule in all its complexity' and mapping this knowledge onto the relevant gene products encoded by genomes (see Table 1). This involves both experimental and computational approaches and, indeed, absolutely requires their integration. As such, this effort will occupy the majority of biologists (experimental and theoretical) for the foreseeable future. The network provides the necessary expertise and infrastructure to allow distributed annotation. These expert annotations will be made available to everyone on the web.

Currently, we are far from solving even simple annotation tasks *in silico*. There are two key problems in annotation. Firstly, many of the current computational tools need further development and careful validation against experimental data. Secondly, the integrated results from all this work need to be made available to experimenters in ways that will guide their future work.

In this very dynamic scientific area scientists will provide annotation with their best methods while they continue with the development of the underlying databases and algorithms that are to be incorporated in future annotation pipelines.

Together with this systematic effort to develop general annotations we have selected important biological areas of application in which in-depth computational analysis might be required. Our goal is to provide a focus for the network and offer a wide variety of Bioinformatics methods in order to advance the understanding of diseases. The objective of this network is to extensively and thoroughly analyze the genomic information related to various diseases, and to create a publicly available resource for each of them. This resource will combine existing information with data produced within the network, and allow discussion with experimental scientists working in the corresponding areas.

The first two selected topics are Down's syndrome (chromosome 21) and HCV and HIV viruses, and they will later be followed by other diseases.

In the case of HCV, we will use all the available methods developed by the network partners to try to assign a function to the less well-characterized proteins. There are at least three proteins that are potential antiviral targets for which no structural or functional information is available, and the network will strongly focus on their analysis. In the case of HIV, the aim is both to correlate the variability of the virus with its ability to escape drug therapy, and to use predictive methods to identify immunogenic regions that are potential candidates for vaccine development.

For the study of the human chromosome 21 (Hsa21) we will reannotate the full chromosome with the integrated BioSapiens battery of methods, but also experimentally assay the function of conserved nongenic sequences and contribute to the sequencing of 21p, which is part ($\sim$6%) of the human genome that remains to be sequenced. In addition, we will analyze the HapMap results for this chromosome, as well as the results of the 'Perlegen' resequencing effort, generate 21p CGH BAC arrays covering the long arm of human chromosome 21, and analyze transcriptome results in human and mouse models.

How to distribute the annotations generated by BioSapiens?

For the success of the project it is essential to make the complex annotations generated by BioSapiens accessible to the experimental biologist. For this task we have adopted the DAS technology. DAS was first created for the visualization of the annotations of the human genome, using the sophisticated Ensembl system for the representation of the results (www.emsembl.org), which is the entry point most commonly used by experimental biologist to access the human genome information. In the context of BioSapiens we also have other variants to view the annotations related with protein sequences (DASTY), and three-dimensional protein structures (SPICE, extending the version originally developed by the Sanger Institute) (see Figure 1). The annotations generated by the DAS servers can be viewed from any site worldwide using any of these three very easy to use systems (DAS clients), with the additional possibility for the users to select the annotations appropriate to their biological questions. In this first year of the network our goal has been to establish the DAS infrastructure, and to test the DAS access systems (clients) before releasing them to the public domain. We now intend to focus on the development of new tools, and very importantly, on setting up the evaluation, scoring and consensus systems to provide not only the best annotations but also a scientifically sound evaluation of their reliability that can help

Table 1 BioSapiens genome annotation levels

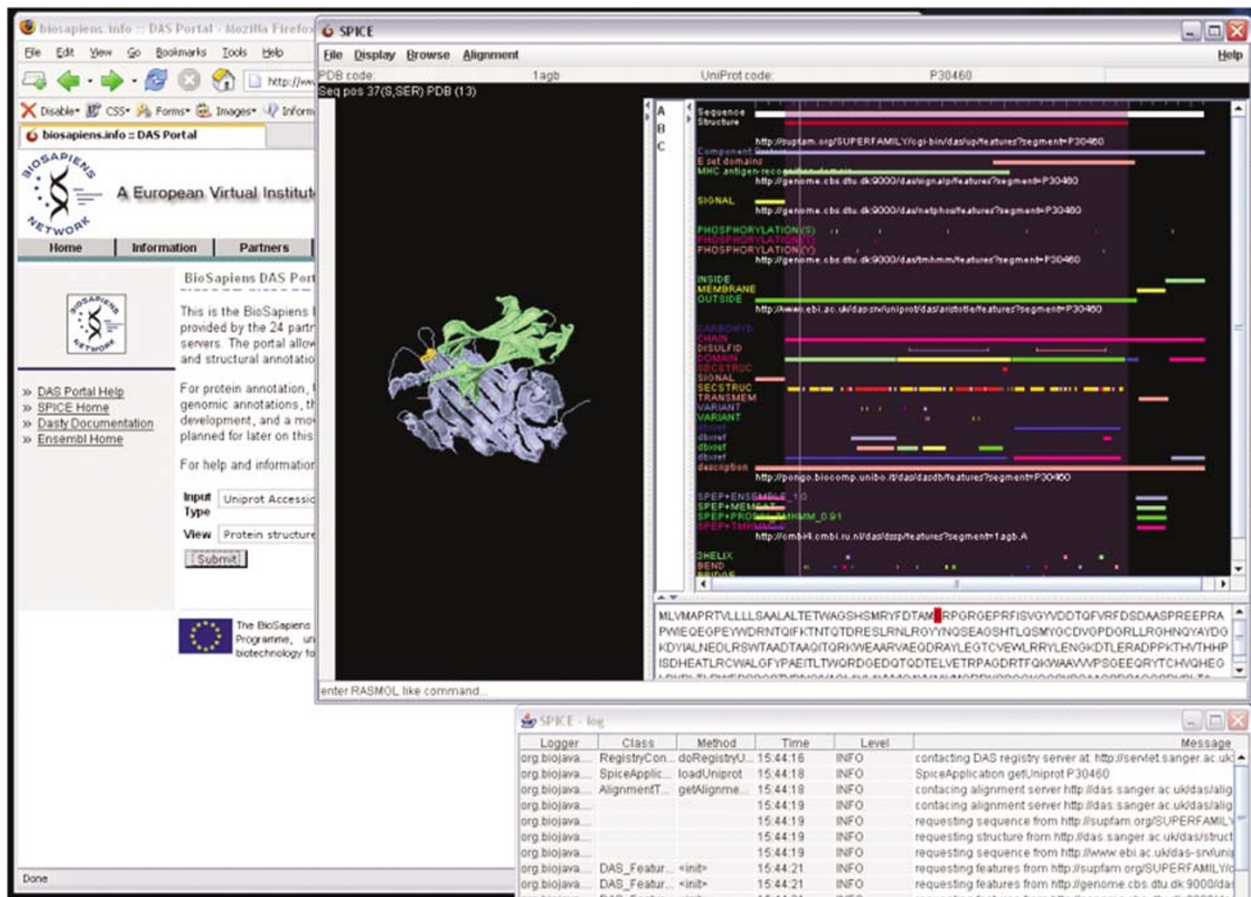| DNA annotation⟶	Proteome annotation⟶	Functional annotation
• Gene definition (alternative splicing)	• Protein families	• Sequence and structure to function
• Regulators and promoters	• Protein structure and modelling	• Protein–protein complexes
• Expression	• Membrane proteins and ligands	• Pathways and networks
• Variation (haplotypes and SNP's)	• Post translation modification and localization	

Figure 1 The BioSapiens DAS portal. Screen shot of the BioSapiens web portal showing functional annotations for a given protein sequence collected from the predictions provided the various teams of the network. The annotations are simultaneously displayed along the sequence of the protein and in their position in the corresponding three-dimensional structure of that protein. Using this type of systems biologist will be able to analyze genomes at different levels of complexity by freely combining the annotations provided by the DAS servers developed by the BioSapiens network. Figure kindly provided by Brendan Vaughan (EBI-EMBL).

users to assess the significance of the predictions.

The network's Scientific Advisor Committee is organized to stimulate participation in biological projects and to help with their critical views to better organize the activities of the network.

Barry Honig, Columbia University (chair); Siv Andersson, University of Uppsala; Janan Eppig, The Jackson Laboratory; Ian Harrow Pfizer, Veronica van Heyningen, MRC Human Genetics Unit, Edinburgh; Minoru Kanehisa, Kyoto University; Jonathan Knowles, Hoffman-La Roche; Carlos Martinez-A, CSIC, Madrid; Iain Mattaj, EMBL; Trudy McKay, North Carolina State University; Gert-Jan van Ommen, University of Leiden; Kai Simons, Max Planck Institute of Molecular Cell Biology and Genetics; Mathew Woodwark, AstraZeneca.

Collaboration with the community
The Network has established a permanent 'European School of Bioinformatics' (see http://www.BioSapiens.info), to train bioinformaticians and to encourage best practice in the exploitation of genome annotation data for biologists. We have organized two basic training workshops, one in Verona, Italy followed by the related Advanced Workshop on 'Molecular Interactions' (Sponsored by ESF), and in Nijmegen, Netherlands, followed by a workshop on 7TM Receptors.

As in other networks, BioSapiens is designed to create a focus of collaboration and development in Bioinformatics in Europe; we have supported a number of sectorial meetings, such as CAPRI: Critical Assessment of Predictions on Interactions; CASP6:

Critical Assessment of Techniques for Protein Structure Prediction; Bologna Winter School in Bioinformatics; meeting on 'Genome Annotation' at ISMB/ECCB2004. This year we will contribute to the organization of: the annual European Conference for Computational Biology (next meeting at Madrid in September 2005, www.eccb05. org); II meeting on 'Genome Annotation'; future editions of the Bologna Winter School in Bioinformatics that will host a permanent 'BioSapiens corner'; the BioSapiens workshop at 2nd ESF Conference: Functional Genomics and Disease (www.functionalgenomics.org.uk). We will also host specific meetings with scientist working on Down's syndrome (chromosome 21) and a Symposium, 'Molecular and computational Biology of HIV and HCV' jointly with the viRgil NoE, which focuses in HIV.

Table 2

The groups participating in the BioSapiens Network are:
Prof Janet Thornton, Director of the European Bioinformatics Institute-EMBL, UK (Coordinator)
Prof Anna Tramontano, U Roma, Italy (Training Coordinator),
Prof Alfonso Valencia , CNB-CSIC, Spain (Outreach Coordinator)
Prof Søren Brunak, Technical University of Denmark, Denmark (Thematic work package coordinator)

Prof Stylianos Antonarakis, University of Geneva, Switzerland
Peer Bork, EMBL Heidelberg, Germany
Prof Rita Casadio, University of Bologna, Italy
Prof. Antoine Danchin, Institut Pasteur, France
Dr Richard Durbin, Wellcome Trust Sanger Institute, UK
Prof. Dmitrij Frishman, German Nat. Centre for Environment and Health, Germany
Dr Roderic Guigo, IMM Laboratory of Computational Genomics, Spain
Prof Gunnar von Heijne, Stockholms Universitet, Sweden
Prof. Jacques van Helden, Université Libre de Bruxelles, Belgium
Prof Dr Thomas Lengauer, Max-Planck Institute für Informatik, Germany
Prof Michal Linial, The Hebrew University of Jerusalem, Israel
Dr Richard Mott, Wellcome Trust for Human Genetics, UK
Prof Christine Orengo and Prof David Jones, University College London, UK
Prof Laszlo Patthy, Institute of Enzymology, Hungary
Dr Leszek Rychlewski, BioInfoBank Institute, Poland
Dr Vincent Schachter, Genoscope-CNS, France
Prof Dietmar Schomburg, University of Cologne, Germany
Prof Esko Ukkonen, University of Helsinki, Finland
Dr Anne-Lise Veuthey, Swiss Institute of Bioinformatics, Switzerland
Prof Martin Vingron, Max-Planck Institute for Molecular Genetics, Germany
Prof Gert Vriend, University of Nijmegen, The Netherlands
The network manager is Dr Kerstin Nyberg (knyberg@ebi.ac.uk)

The web portal

All the network information and activities, together with the various web DAS clients displaying the full range of genome annotations, is progressively incorporated into the BioSapiens Portal (http://www.biosapiens.info) together with the information about how to install the DAS systems locally.

The portal is also the entrance to the activities organized by BioSapiens, and to the interaction with the groups participating in the network.

A Valencia is at the National Centre for Biotechnology, CNB-CSIC, Cantoblanco, Madrid 28049, Spain. E-mail: Valencia@cnb.uam.es, http://www.biosapiens.info