Metabolomics, Vol. 2, No. 3, September 2006 (© 2006)
DOI: 10.1007/s11306-006-0030-6

Genomes to systems 3

Stephen J. Wilkinson,[a,b] Irena Spasić,[a,b] and David I. Ellis[a,b,*]

[a]*School of Chemistry, University of Manchester, Sackville Street, Manchester, M60, UK*
[b]*1QD, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester, M1 7ND, UK*

Received 18 May 2006; accepted 22 May 2006

A report on the 3rd Conference of the Consortium for Post-Genome Science 'Genomes to Systems' at the Manchester International Convention Centre, Manchester, UK, 22–24 March 2006 on the latest developments in post-genome science. The meeting contained over 60 presentations on a diverse range of subjects including; genomics, proteomics, transcriptomics, metabolomics, informatics, and systems biology, and included plenary lectures by some of the world's most distinguished speakers.

KEY WORDS: meeting report; systems biology; metabolomics.

This was the third consortium conference which portrayed the breadth of the post-genome sciences including Genomics, Transcriptomics, Proteomics, Metabolomics, Informatics, and integrative Systems Biology. The conference was attended by over 800 delegates from 40 countries and included 56 presentations and 5 plenary lectures on a diverse range of subjects including; structural genomics of disease, metabolomics in health and disease, advances in biosensors and instrumentation, disease proteomics, high-throughput functional genomics and gene manipulation, public population projects in medicine, pharmacogenomics, standards for reporting 'omics data, nanoscale technologies, transcriptomics and expression microarrays, and two sessions on systems biology.

Bernhard Palsson's (San Diego, http://www.gcrg.ucsd.edu/) opening plenary lecture was the first public display of his groups work on the reconstruction of the ''The Human Metabolic Network'' in which he discussed his bottom up approach to systems biology. For Palsson, ''Genomes to Systems'' equates to ''1-D to 2-D''. The genome-scale reconstructions being painstakingly developed by Palsson's lab are 2-D representations of the complete genome of an organism that build upon the 1-D representations such as that delivered by the human genome sequencing project. These models facilitate a constraint based analysis of the metabolic network based on topological information i.e., the set of admissible biochemical transformations (Palsson, 2002, 2004). The steady state fluxes of metabolites must lie within the feasible constraint space as defined by the stoichiometric matrix. Palsson gave a useful analogy: the network stoichiometry can be compared to the road transportation system which is fixed; whereas the metabolite fluxes are functional states like the traffic flows that vary depending on the time of day etc. The state of the metabolic network in the liver, for example, is quite different depending on whether one is hungry or has just been fed.

As well as the stoichiometric constraints, there are also upper and lower bounds which constrain the permissible flux values. It is these bounds that can be manipulated to simulate pathological states. Diabetes, for example, is modelled using a reduced value for the maximal glucose uptake rate (75% of the normal physiological value). Different dietary regimes can be investigated in a similar manner. A model for a high fat diet, for example, has an increased lower bound on cellular fatty acid import. In addition, efficient mathematical optimization algorithms enable the space of possible network states to be rapidly explored. Successful applications of this approach for model organisms such as yeast have explained substrate preferences, the consequences of gene knock-outs and optimal growth rates. Since Build 35 of the human genome sequence was released in October 2004, Palsson's group have been reconstructing the human metabolic network. Validation of their model (currently 3280 reactions and 2757 metabolites) has already proved valuable. It has revealed that inter-organelle transporters are poorly characterized and has also shown up serious gaps in current knowledge e.g. how is Vitamin C metabolized?

The model is also demonstrating that well established phenomena could be explained solely on stoichiometric grounds rather than by more complex regulatory mechanisms. The TCA cycle compound acetyl-coEnzyme A, for example, represents a convergence point for different catabolic pathways such as glycolysis, oxidation of fatty acids etc. As a consequence, these pathways are stoichiometrically linked. In steady state, the fluxes must

* To whom correspondence should be addressed.
E-mail: D.Ellis@manchester.ac.uk

balance and it is this simple fact (rather than enzymatic inhibition) that may be responsible for the reduced apparent activity of pyruvate dehydrogenase in diabetic cardiomyocytes.

The opening presentation of the metabolomics in health and disease session by Douglas Kell (Manchester, http://www.dbkgroup.org/) covered a range of his group's activities, such as the HUSERMET programme, one of the many aims of which is to determine exactly what the 'normal' and diseased human serum metabolome are composed of in (bio)chemical terms (http://www.metabolomics.co.uk/). A philosophy of the cycle of knowledge was then introduced, whereby it was forwarded that computational methods of data analysis, which may be automated, provide the means of generating novel hypotheses, especially in the post-genomic era (Kell and Oliver, 2004). This was then followed by several examples, such as the robot scientist which is a physically implemented robotic system that applies techniques from artificial intelligence (in this case inductive logic programming) to undertake cycles of experimentation. The system was applied to the determination of gene function of *Saccharomyces cerevisae* and auxotrophic growth experiments (King *et al*., 2004). It operated by automatically generating original hypotheses to explain observations, devising experiments to test these hypotheses, physically ran the experiments via a lab robot, and finally, interpreted results to falsify hypotheses inconsistent with the data before repeating the cycle.

Kell then went on to explain how similar computational methods were used for the closed-loop multi-objective optimization of analytical instrumentation, in this case gas chromatography time-of-flight mass spectrometry, of the metabolomes of human serum and yeast fermentations (O'Hagan *et al*., 2005). Using evolutionary computation, entirely automated and without any human intervention, this system was able to optimize the number of chromatographic separations (in this case >1200 peaks from a total of 200,000,000 combinations). Whilst in many cases preserving or reducing run times and preserving excellent signal-to-noise ratios. This methodology has been further developed and successfully applied to a GC×GC tof MS system, in the search for biomarkers of several diseases including pre-eclampsia (Kenny *et al*., 2005). It was stated that <3500 peaks are now being detected due to optimization but these also include peaks from derivatives and not just metabolites. It was concluded that closed-loop optimization can increase substantially the number of metabolites observed in serum.

Jan van der Greef (Leiden, http://www.research.leidenuniv.nl/index.php3?c=150) gave a stimulating presentation on the impact of metabolomics on health and disease, and connectivity mapping and systems response profiling. He first introduced his concepts of systems thinking (van der Greef and McBurney, 2005), inspired in part by Fritjof Capra (Capra, 1996). The connectivity hypothesis was illustrated and van der Greef said that an embryonic paradigm shift was already under way this century, which he termed 'quantum biology'. Systems science, van der Greef said, was the science of organized complexity and the challenge in human systems biology was to unravel this complexity and connectivity and answer questions and challenges such as; how are biomarkers related to systems knowledge, what is the relationship of molecules in health and disease, and how are patterns of self-organisation changing from the healthy to the disease state?

An overview of systems biology methodology was given in terms of parallel analyses of mRNA, proteins and metabolites from complex samples, through to discovery of biomarkers of disease and drug/nutritional effects. This accumulation of systems knowledge is finally translated into systems pathology and systems pharmacology/toxicology. Illustrated examples were given including the effects of drugs on lipid metabolism in a mouse model of type 2 diabetes and optimising drugs in terms of reducing liver toxicity. Van der Greef demonstrated that metabolomics is an excellent tool for optimization of pharmaceutical research and development based on a systems perspective. He demonstrated that the use of systems-based strategies could show us how the combination of different drugs could lead to beneficial synergistic effects, as well as the potential for drug 'rescue'. He also spoke of health promotion, more recently termed disease management, and the requirement for focusing on prevention and sub-health, concluding that so-called personalized medicine will fundamentally change current approaches in medicine (van der Greef, 2005).

Chris Newgard's (Duke, http://www.pharmacology.mc.duke.edu/faculty/newgard.htm) presentation centred on metabolic profiling in obesity and diabetes using GC-MS and MS-MS for targeted analysis and UPLC-MS for 'unbiased' analysis. Newgard also collected metabolic flux measurements via NMR analysis of stable isotopes followed by informatics. He termed his group's approach "more a metabolic physiology platform than a metabolomics one". Newgard's aim is to gain mechanistic metabolic regulatory knowledge and he illustrated his research with a series of examples including; molecular and biochemical mechanisms of fuel-mediated insulin secretion and their impairment in obesity and diabetes, mechanisms of pancreatic beta-cell growth and survival, and a series of two-stage models of the regulation of glucose metabolism by flux analysis and metabolic profiling (Lin *et al*., 2005; Schisler *et al*., 2005).

One of the few plant science presentations at the conference was given by Luis Mur (Aberystwyth, http://www.users.aber.ac.uk/lum/) on the metabolomic analysis of rice blast disease. The two species of rice provide

more than one-fifth of the calorific intake of humans and it is therefore, an extremely important and significant crop. The causal agent of the devastating disease rice blast is the pathogenic fungus *Magnaporthe grisea* (Dean *et al*., 2005) which is responsible for a 11–30% world-wide annual loss in rice crop (enough to feed over 60 million people) and is known to occur in 85 countries worldwide. Mur and colleagues used metabolomic approaches to elucidate the key metabolite changes which occur during the interactions of *M. grisea* with the alternative host *Brachypodium distachyon* (a model system for grasses (Routledge *et al*., 2004)). To date, few metabolomics studies have investigated the effects of two interacting organisms. In this study, both metabolic fingerprinting (FT-IR) and metabolic profiling (ESI-MS) approaches were applied to analyse single resistance gene-mediated resistance to *M. grisea* and disease system development in *B. distachyon* which were very similar to those occurring in rice (*Oryza sativa*).

Metabolic fingerprinting of *B. distachyon* challenged by *M. grisea* after 3 and 5 days was undertaken by Fourier transform infrared (FT-IR) spectroscopy. Chemometrics was then applied to the resultant data to assess whether the spectra contained enough (bio)chemical information to allow differentiation of infected plants and controls. It was observed that both accessions infected with *M. grisea* produced more distinctive metabolic fingerprints, indicating a greater host response. More in-depth chemometric analysis indicated that metabolites within the amide and polysaccharide regions contributed to differential and common ecoptypic responses to pathogenic challenge, and in addition, a distinct group of fatty acid metabolites were observed. Following screening by FT-IR, a more targeted analysis was required and it was decided that the fatty acid group was technically easier to target using non-polar extraction procedures. Further, it was also known that this group includes chemicals that are emerging as important defence signals. Targeted metabolic profiling was undertaken by injecting non-polar extracts of *M. grisea* challenged *B. distachyon* directly into an ESI-MS. Results revealed discriminatory analytes between each interaction and seven metabolites were subsequently identified as phospholipids by ESI-MS-MS. It was observed that phosphatidyl glycerol phospholipids were suppressed during both resistant and susceptible responses whilst different phosphatidic acid phospholipids either increased, or were reduced during resistance or during disease development. Mur suggested that phospholipid processing of membrane lipids during each interaction may be associated with the elaboration/suppression of defence mechanisms or developing disease symptoms (Allwood *et al*., 2006).

The enormity of the post-genomic challenge was further illustrated by Mike Snyder (Yale, http://www.yale.edu/ snyder/). In his plenary lecture he described research by his group to assemble gene regulatory networks for humans using novel high-throughput screening techniques. They have found that NF-κB has 200 distinct binding sites on human Chromosome 22 (Urban *et al*., 2006). A network mapping assists in the task of identifiying master transcriptional regulators and target 'hubs' – DNA sites that bind to several distinct transcription factors. Work continues on key regulators involved in other signalling pathways such as MAPK and cAMP. Progress will be dramatically improved if we can answer the inverse problem in transcriptional regulation: given a piece of DNA, how can we find the protein that binds to it?

In terms of the proteome, proteins interact by catalysing post-translational modifications of themselves such as the phosphorylation of kinases by other kinases in the MAPK signalling cascade. Snyder believes that a step change in the speed of knowledge discovery is needed given the comparatively slow progress up to now, even in such genetically well characterized organisms as yeast. At least 30% of yeast proteins are phosphorylated and there are 122 protein kinase homologues, yet fewer than 160 kinase-substrate interactions have been characterized. Snyder's lab has pioneered protein microarray technology to probe the 'phosphorylome' of yeast. A total of 4200 phosphorylation events have been cataloged for 13,250 proteins. On average, each kinase phosphorylates 47 proteins on the chip but even closely related kinases (e.g., 85% homologous) can quite have different substrate target lists. One of the key challenges ahead is to probe the interactions between the networks of regulating transcription and phosphorylation to identify common modules and motifs.

One interesting and very important session, when one considers the flood of data being generated by post-genomic technologies, was a discussion workshop on the progress in standards for reporting 'omics data. It is well recognised that modern high-throughput technologies have resulted in an exponential increase in the volume and the complexity of the 'omics data. Obviously, no significant progress in this type of biological research can be achieved without extensive use of computational tools. Therefore, 'omics data need to be rendered in a computationally tractable form. Moreover, standards are needed to support the sharing of data within the community.

Charlotte Capener (BBSRC) reported on the BBSRC data-sharing policy (http://www.bbsrc.ac.uk/society/consult/data%5Fsharing%5Fpolicy/). Data sharing must be driven by scientific need and cost effective. The areas where there is most scientific need for data sharing are high throughput 'omics technologies. Therefore, research proposals should include data sharing plans and standards are fundamental for effective data sharing. Where standards do not currently exist or are not widely accepted, community development is encouraged.

Jason Snape (AstraZeneca) emphasised the importance of data standards to facilitate the uptake of genomic science, where data sharing is essential for scientific progress. John Quackenbush's vision (Quackenbush, 2004) that different types of 'omics data should not be considered independently since they provide us with complementary views of the fundamental biological processes we are studying is reliant upon on quality capture, management and archiving of data and the associated metadata. Data alone have no future unless they are properly annotated and stored, locatable and accessible. Not surprisingly, genomic science is increasingly viewed as being data rich and knowledge poor. Therefore, there is a need to collect the data in an added value format. However, initial standards were too domain specific and not inclusive. In order to support common data sharing platforms, data standards need to be community wide and not community specific. Open community-driven standards are required to instil confidence, address scepticism, facilitate/maximise value of data and minimise non-compliant data and legacy data.

Chris Taylor (EMBL-EBI) provided a coordinated multi-omics view of data standards. Facing a myriad of data formats, databases and varying requirements, the role of standards is to facilitate data exchange, define minimal reporting requirements and provide controlled vocabularies and ontologies, but not to specify a database structure and recommend experimental protocols. Inclusivity is the key to developing an acceptable standard relying on active outreach, where silence implying assent represents a practical way of working.

The transcriptomics community took a lead in providing standards for the 'omics data. The Microarray Gene Expression Data Society was founded in 1999 to facilitate the exchange of microarray data. They proposed a standard for the Minimum Information About a Microarray Experiment (MIAME) and the Microarray Gene Expression Markup Language. Subsequently, the Human Proteome organisation founded the Protein Standards Initiative in 2002 to standardise the Minimum Information to describe a Proteomics Experiment (MIAPE). Finally, the Metabolomics Society (http://www.metabolomicssociety.org/) was established in 2004 and amongst other activities to coordinate the efforts in standardizing reporting structures of metabolomics experiments. These efforts will result in a draft document on 'Reporting standards in Metabolomics' aiming at reaching a consensus to be used as a guideline for researchers, journals, funding and scientific organizations, vendors and regulatory bodies. Five working groups have been founded to cover the key areas for describing metabolomic experiments: biological sample context, chemical analysis, data analysis, ontology and data exchange.

At the start of the conference Bernard Palsson described progress in building on '1-D' DNA sequence data in order to fully characterise the networks of interacting biomolecules. Mike Snyder (Yale, http://www.yale.edu/snyder/) showed that the challenge of writing down these static or steady-state '2-D' networks still remains. The last day of the conference demonstrated that research is well underway on '3-D' systems modelling; the added dimension being that of time. Dynamic signalling models are essential for unravelling the mechanisms by living systems respond to their environment. The temporal aspects of systems biology were perfectly illustrated in lectures by Tom Kirkwood on understanding the ageing process and by Dennis Noble on his whole organ model of heart rhythm.

Tom Kirkwood (Newcastle, http://www.ncl.ac.uk/medi/research/gerontology/) defines aging as the accumulation of cellular defects caused by random molecular damage in cells. Every cell in the human body suffers an average of 10,000 hits by free radicals every day causing oxidative stress and DNA damage. Fortunately, we possess excellent repair functions to restore our cells to their healthy state. But maintenance of somatic cells does not come cheap in terms of ATP and takes a significant amount of energy away from other activities such as growth and reproduction. The ''disposable soma theory'' (Kirkwood, 1977) says that effort should be only devoted to cell maintenance for as long as an organism is expected to live. This is backed up by data; higher levels of poly ADP-ribose polymerase 1 (ARP-1) – a key enzyme involved in repair – correlate with longer life spans across species and even within the human population.

The disposable soma theory provides a powerful evolutionary argument for why aging occurs but there is still the challenge of building models to tell us how it proceeds, how it is regulated and how the rate of aging can be altered. It is well known that aging has a strong genetic component but also that individual organisms can delay their own aging in response to reduced levels of food (dietary restriction). Kirkwood believes a ''network'' approach is required include all the different mechanisms that are thought to contribute to the overall aging process. The balance between the various constituents of aging is well illustrated at the cellular level during mitosis. Synthesis of new cellular material in cell division dilutes any accumulated trash (metabolic waste, aberrant proteins etc.) and will therefore, tend to slow the rate of molecular damage in rapidly dividing cells. On the other hand, these same cells will be far more prone to DNA mutations than their slower growing counterparts.

Aging and death is not pre-programmed into us but merely represents our failure to keep going says Kirkwood. Understanding the intrinsic complexity of how this comes about represents a big challenge for systems biology to tackle in the future.

The closing plenary lecture was delivered by Denis Noble (Oxford, http://www.noble.physiol.ox.ac.uk/)

whose career has spanned the development of computer-aided physiological modelling. The audience enjoyed his unique historical perspective of systems biology, and indeed, he was introduced as one of the pioneers of systems biology. He developed the first computer model of the heart in 1960 building on the seminal work of Hodgkin and Huxley (Noble, 1960). At this time computers were still in their infancy – the simulation ran with the help of punched cards. This was genuine 'systems biology' long before the name was invented.

Noble showed the inputs and outputs of his early model in graphical form. A heart beat involves the co-ordinated depolarisation and repolarisation of heart muscle cells to create cyclic mechanical contractions. Stimulated cardiac cells vary the electrical potential across their membranes by the opening and closing trans-membrane ion channels. Originally, Noble assumed two types of potassium channel and one type of sodium channel. He ignored calcium channels completely – reason: they had not been discovered yet! Despite this the model successfully predicted the existence of a plateau in the action potential – a key feature that enables the cycle to operate with high energy efficiency.

Subsequent efforts were fuelled by new experimental findings as Noble and others increased the scope and accuracy of their models adding, for example, ion pumps and transporters. For Noble, a model that is subsequently proved to be 'wrong' by experiments is a not a failure but an important step on the road to greater understanding. Another important modelling success for Noble was his prediction of a 3:1 stoichiometry for Na:Ca exchange rather than the previously accepted electrically neutral ratio of 2:1 (DiFrancesco and Noble, 1985).

Activity and progress in heart modelling went through something of a lull until 1990 but since then has undergone a considerable renaissance. Incorporation of single cell models into tissue models was delayed until the development of sufficiently powerful computers. First steps in this direction used a uniform of grid of cells – even this unrealistic arrangement of cells verified that a small ischaemic focus could trigger a wave of excitation through the rest of the tissue.

The contrast of the early days with today's state-of-the-art in computational modelling was striking as the audience was fast-forwarded to the virtual heart project and some stunning animations. The virtual heart is an anatomically detailed, whole organ model which can simulate the conditions for initiating arrhythmia (Noble, 2006). Such studies are hugely important for the drug industry since many prospective pharmaceutical compounds can trigger fatal arrhythmias by binding to one of the channel proteins in cardiac cells. The virtual heart model is also helping us to understand the mechanisms by which electrical defibrillation might work.

Despite the successes of the virtual heart in integrating different levels from genes up to the whole organ, Noble accepts that there will still be a need for different models depending on the particular questions they are trying to answer in detail. In this respect the trade-off between model accuracy and computational tractability will remain for the foreseeable future. Even today, calculations are often restricted to a 2-D slice rather than the whole ventricle. But if we extrapolate from the progress of the previous 40 years, we can realistically imagine a complete set of organ models inspired by the virtual heart in the not too distant future. The ultimate fulfillment of Professor Noble's work on building and integrating models – the virtual human – seems now to be within our grasp. The conference was excellently concluded with this *tour de force* by Noble, with the next Genomes to Systems conference scheduled for Manchester in 2008 (http://www.genomestosystems.org/).

References

Allwood, J.W., Ellis, D.I., Heald, J.K., Goodacre, R. and Mur, L.A.J. (2006). Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory non-polar metabolites in responses by *Brachypodium distachyon* to challenge by *Magnaporthe grisea*. *Plant J.* **46**, 351–368.

Capra, F. (1996). *The Web of Life: A New Scientific Understanding of Living Systems*. Doubleday, New York.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H.Q., Read, N.D., Lee, Y.H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W.X., Harding, M., Kim, S., Lebrun, M.H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E. and Birren, B.W. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**, 980–986.

DiFrancesco, D. and Noble, D. (1985). A model of cardiac electrical activity incorporating ionic pumps and concentration changes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **307**, 353–398.

Kell, D.B. and Oliver, S.G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* **26**, 99–105.

Kenny, L.C., Dunn, W.B., Ellis, D.I., Myers, J., Baker, P.N., Consortium, G. and Kell, D.B. (2005). Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning. *Metabolomics* **1**, 227–234.

King, R.D., Whelan, K.E., Jones, F.M., Reiser, P.G.K., Bryant, C.H., Muggleton, S.H., Kell, D.B. and Oliver, S.G. (2004). Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252.

Kirkwood, T.B. (1977). Evolution of ageing. *Nature* **270**, 301–304.

Lin, J.D., Yang, R.J., Tarr, P.T., Wu, P.H., Handschin, C., Li, S.M., Yang, W.L., Pei, L.M., Uldry, M., Tontonoz, P., Newgard, C.B. and Spiegelman, B.M. (2005). Hyperlipidemic effects of dietary saturated fats mediated through PGC-1 beta cactivation of SREBP. *Cell* **120**, 261–273.

Noble, D. (1960). Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations. *Nature* **188**, 495–497.

Noble, D. (2006). Systems biology and the heart. *Biosystems* **83**, 75–80.

O'Hagan, S., Dunn, W.B., Brown, M., Knowles, J.D. and Kell, D.B. (2005). Closed-loop, multiobjective optimization of analytical instrumentation: gas chromatography/time-of-flight mass spectrometry of the metabolomes of human serum and of yeast fermentations. *Anal. Chem.* **77**, 290–303.

Palsson, B. (2002). In silico biology through ''omics''. *Nat. Biotechnol.* **20**, 649–650.

Palsson, B. (2004). Two-dimensional annotation of genomes. *Nat. Biotechnol.* **22**, 1218–1219.

Quackenbush, J. (2004). Data standards for 'omic' science. *Nat. Biotechnol.* **22**, 613–614.

Routledge, A.P.M., Shelley, G., Smith, J.V., Talbot, N.J., Draper, J. and Mur, L.A.J. (2004). *Magnaporthe grisea* interactions with the model grass *Brachypodium distachyon*closely resemble those with rice (*Oryza sativa*). *Mol. Plant Pathol.* **5**, 253–265.

Schisler, J.C., Jensen, P.B., Taylor, D.G., Becker, T.C., Knop, F.K., Takekawa, S., German, M., Weir, G.C., Lu, D.H., Mirmira, R.G. and Newgard, C.B. (2005). The Nkx6.1 homeodomain transcription factor suppresses glucagon expression and regulates glucose-stimulated insulin secretion in islet beta cells. *Proc. Natl. Acad. Sci. USA* **102**, 7297–7302.

Urban, A.E., Korbel, J.O., Selzer, R., Richmond, T., Hacker, A., Popescu, G.V., Clubells, J.F., Green, R., Emanuel, B.S., Gerstein, M.B., Weissman, S.M. and Snyder, M. (2006). High-resolution mapping of DNA copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA* **103**, 4534–4539.

van der Greef, J. (2005). Systems biology, connectivity and the future of medicine. *IEE Proc. Syst. Biol.* **152**, 174–178.

van der Greef, J. and McBurney, R.N. (2005). Innovation – rescuing drug discovery: in vivo systems pathology and systems pharmacology. *Nat. Rev. Drug Discov.* **4**, 961–967.