EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Exploratory Workshop Scheme

Standing Committee for Life, Earth and Environmental Sciences (LESC)

ESF Exploratory Workshop on

From phenotypes to pathways
Inferring genetic architecture from perturbations maps

Cambridge (UK), 9-11 Sept 2010

Convened by:
Florian Markowetz and Michael Boutros

SCIENTIFIC REPORT

---

1. Executive summary

The ESF exploratory workshop '*From phenotypes to pathways – inferring genetic architecture from perturbation maps*' was held from Sept 9 to 11 at Lucy-Cavendish college in Cambridge, UK.

Scientific objectives. This workshop focussed on novel experimental and computational strategies for perturbation analysis in dissecting cellular regulatory networks and disease mechanisms. Modern high-throughput approaches are key technologies at the forefront of genetic research. They enable the analysis a biological response to thousands of experimental perturbations, however require a tight collaboration experimental and computational scientests. The objective of the workshop was to provide a platform for such exchanges and to initiate interdisciplinary collaborations.

The participants are leading European scientists covering both theoretical and experimental approaches for phenotyping and a wide range of research areas, model organisms, human genetics and experimental techniques. Together we worked on defining the main unanswered questions in the field and to identify the next key challenges: *Which (computational or experimental) developments will drive the field forward in the next years?*

The workshop was organized around five key areas:

1. Designing phenotypes: What are the challenges to utilize quantitative phenotypes?
2. From phenotypes to mechanisms: what do perturbation effects tell us about protein function and cellular networks?
3. Gene-gene and gene-drug interactions: How do epistatic effects relate to cellular networks and pathways?
4. Modeling the cell: how can we predict phenotypes and synthetic interactions?
5. Integrative design and analysis: what other complementary data types and experiments can maximize the information gained from perturbations?

Organisation. Twenty-seven participants from fourteen countries presented their work in designing and analysing gene perturbation screens. Most participants (except for a few living in Cambridge or London) stayed at the college for the complete duration of the workshop.

The meeting was organized into five sessions, each chaired by one of the participants and generally containing six talks. Each talk was limited to 20 minutes to ensure it to be concise and focussed. On the last slide each speaker answered the question: *What is the next big thing in my field?* After each talk there was opportunity for questions. The main discussions happened after each session in a general discussion period of almost one hour. In each session two participants acted as discussion leaders, one with a computational and one with an experimental background. They summarized the talks, put them into a greater perspective, and stated which questions seemed to be still left open. Additionally, there were two evening discussion sessions led by the convenors on Friday and Saturday. The first one led to informal discussion in smaller groups, while the second one was focussed on potential follow-up activities.

Prior to the meeting we had asked participants to choose the two questions (from the five listed above) most related to their own research and to send us a very concise (two sentences) description of their talk topic. This gave us the opportunity to match people with

similar interests and distribute them into sessions, while at the same time allowing the participants great flexibility in arranging their talk.

General atmosphere. The participants very much appreciated that the focus of the workshop was on discussion, not formal presentations. Talks were kept deliberately short and much time was allocated to discussions. All participants engaged actively in discussions. The participants also greatly appreciated the mix of computational and experimental researchers at the workshop. Usually meetings tend to be very homogeneous and almost exclusively contain either computational or experimental scientists. At our workshop, however, the ratio was almost 50/50 and the range of expertise was much wider than in other workshops. A third important point is the tight focus we kept on gene perturbations. Almost all talks directly addressed the issue, while the few others provided necessary background, e.g., in other analysis strategies. In sum, the clear focus together with the mix of expertises and the time allowance lead to very engaging and deep discussions. In general, the atmosphere was very open and lively, even though many of the participants had not met before.

Conclusions. The key question was: *Which (computational or experimental) developments will drive the field forward in the next years?* The participants answered this question in a variety of ways, but certain patterns began to emerge: For example, there is a need for more complex phenotypic descriptions. Many speakers stressed the higher resolution of gene function resulting from exchanging ”simple” by richer phenotypes, e.g. cell morphology or molecular read-outs like global gene expression. Participants stressed the importance of integrated analysis of heterogeneous data. This poses new challenges for data analysis and computational method development.

Participant feedback. During the meeting and afterwards by email we got very positive feedback from participants. They all felt that the focus on gene perturbations, the mix of computational and experimental researchers, as well as the amount and quality of discussions set this meeting apart from many others they have visited.

For example, Tom Michoel (Freiburg/DE, theory) commented «It really was a splendid workshop which couldn't have come at a better moment for me, now that I've just started my group» and Roderick Beijersbergen (NKI/NL, experimental) wrote «I also have to thank you for an excellent meeting. (…) I certainly think we should give some form of continuation to this initiative.»

Overall, the meeting was extremely well received and many participants were interested in a follow-up workshop which we plan to hold in Fall 2011 (intended to be called “Cavendish Workshops on Network Perturbations”).

2. Scientific content of the event

How to link genotypes and phenotypes is a long-standing question in modern biology. To put this systematically in practice, however, poses new challenges for experimental and theoretical approaches. The ESF workshop gathered more than 25 leading scientists working on theory and practice of large-scale perturbation analysis and discussed their approaches over a period of two days.

Large-scale and high-content perturbation screens

In the first session, **Julie Ahringer** (University of Cambridge) talked about her lab's approaches to elucidating mechanisms of cell polarity establishment and transduction. She aims to identify the genes involved and understand their roles and interactions, using the *C. elegans* 1-celled embryo as a model. To this end, her lab has carried out 18 RNAi screens for suppressors of temperature sensitive cell polarity and asymmetric cell division mutants, and are analyzing these data in the context of current knowledge. **Julian Downward** (CRUK London Research Institute) uses genome-wide RNAi screening approaches to investigate RAS oncogene addiction and synthetic lethality to identify ways of achieving optimal differential killing of RAS mutant cancer cells relative to normal cells. In addition, his lab uses similar methods to investigate mechanisms of development of resistance of cancer cells to targeted therapies, and how these might be overcome. **Roderick Beijersbergen** (NKI Amsterdam) presented approaches to tackle the complexity of gene networks that govern breast tumorigenesis to enable the development of predictive models for pathway-targeted therapies. His work focuses on the role of the PPI3K and MAPK pathway in therapy response and novel treatment options for triple negative breast cancer. **Lucas Pelkmans** (ETH Zurich) presentes his lab's recent work on how cell-to-cell variability can be harnessed to reveal the complexity of effects in RNAi phenotypes. He showed that population context modelling approach is crucial for a correct interpretation of RNAi phenotypes in virus infection, and to seperate between indirect effects that act through the population and direct effects that act in single cells independent of population effects. **Jean-Philippe Vert** (Ecole des Mines, Paris) discussed an ongoing project of high-content high-throughput screening of cellular phenotypes in cancer research. In particular, he described how to automatically quantify proliferation and differentiation phenotypes. **Rengul Cetin-Atalay** (Bilkent University) presented a data- and model-driven hybrid approach to evaluate quantitatively biological activity of a specific cellular process and to identify significant paths leading to this process. This approach is exemplified on PI3K/AKT pathway by microarray data obtained with specific inhibitors of this pathway as well as publicly available array and ChIP-Seq data.

In the general discussion, participants critically evaluated the reliability of current technologies and the impact of possible off-target effects. While some participants found that improvements in screening design and better statistical analysis have resolved many artefacts in the data, others underlined that there are examples of 'real' off-target effects which often limit the interpretation of results.

Synthetic genetic interaction networks

On second day, in particular in the first session, speakers focused on synthetic genetic interaction networks, both from the standpoint of the experiment and the computational analysis. **Michael Boutros** (Heidelberg) discussed approaches how to map synthetic genetic interactions at a large scale in metazoan cells by multiparametric phenotyping. He showed

examples from how to perform synthetic genetic interaction screens by RNAi in Drosophila cells using multiple independent RNAi sequence. **Sebastian Nijman** (IMP Vienna) talked about genetic interactions, particularly those involving small compounds, which provide promising angles for new cancer therapies. His lab employs functional genetic screens with drugs and RNAi to identify such interactions in human cells. **Balazs Papp** (Hungarian Academy of Sciences, Szeged) combined a cell-scale metabolic network model of yeast with large-scale genetic interaction maps to i) evaluate the model's performance to capture epistasis, and ii) develop algorithms to refine the metabolic model based on epistasis data. **Chad Myers** (University of Minesota) discussed efforts to map the first complete digenic genetic interaction network for any cell using Synthetic Genetic Arrays in yeast. He described insights about genetic interaction hubs, the structure of genetic interaction networks, and potential applications in cancer therapeutics. **Wolfgang Huber** (EMBL Heidelberg) presented analysis strategies for combinatorial RNAi to measure matrices of (vector-valued) gene-gene interactions and to use interaction profiles to infer functional modules and assign genes into 'pathways'. **Blaz Zupan** (University of Ljubljana) first reviewed some methods to infer gene-gene interactions from mutant-based experimental data. Then, he presented his group's recent experiments in epistasis analysis of data from Synthetic Gene Array methodology (S. Cerevisiae, data from Charlie Boone's Lab) and shoed how well 4the computationally-inferred hypothesis match known networks. participants discussed the successes of function prediction from different data sources (co-expression, phenotype similarity, genetic interactions). These methods can e.g. predict the *membership* of a protein to a particular pathway, but they do not allow to infer the internal organization or *structure* of the pathway. Participants also discussed if inferring details of pathway structure (instead of a less well resolved general *membership* of genes/proteins to pathways) is maybe a too ambitious goal from available high-throughput data sources and should be at the center of medium- or small-scale follow-up studies.

Identification of signalling networks

Thomas Meyer (MPI for Infection Biology, Berlin) showed how RNAi screens determine crucial host cell factors and signalling networks governing the infection process of a variety of pathogens. HIs work has implications on the identification of host susceptibility determinants and also novel anti-infective targets. **Buzz Baum** (UCL) spoke about intrinsic and extrinsic control of cell shape. Using a combination of classical genetics, RNAi and cell biological techniques in Drosophila his lab has identified genes that control actin dynamics at specific sites within cell. **Cecile Arrieumerlou** (ETH Zurich) described approaches to dissect the signaling pathways that control inflammation during infection of epithelial cells by the pathogenic bacterium *Shigella flexneri*. In particular, she showed results on the identification and characterization of cross-talks between the NF-kB, JNK, p38 and ERK pathways. **Niko Beerenwinkel** (ETH Zurich) talked about computational methods for gene ranking from perturbation experiments. His approach is based on a stability analysis of rankings. **Florian Markowetz** (CRUK Cambridge Research Institute) presented approaches to network analysis of gene perturbation screens, in particular (i) how to interpret single reporter screens in the light of other available data sources and (ii) how to reconstruct pathways from high-dimensional screens. **Jasmin Fisher** (Microsoft Research Cambridge) On executable strategies for cellular decision-making. She described how computational modeling of cellular functions led to experimentally verifiable predictions that extended our knowledge of the mechanisms by which these functions act. Participants discussed how to interpret the results of RNAi screens. How much follow-up do we need to do on the results of large screens?

How do we use cellular networks in this context? Participants discussed different approaches, from function prediction, to prioritizing candidate genes and as maps between genotype and phenotype.

Integrating perturbation maps with other types of data

Data integration is key for interpreting screening results. Open questions are: how to amplify the signal and not the noise? How to measure the relevance of a particular data type/set? **Jussi Taipale** (University of Helsinki) talked about transcription factor - DNA interactions, and computational and experimental dissection of the transcriptional mechanisms that control cell growth and division. **Anna Gambin** (University of Warsaw) modelled the kinetics of the proteolytic degradation in human blood serum. Using mass spectrometry she worked on refining the role of proteolytic processing in cancer development and progression. **Artemis Hatzigeorgiou** (BSRC 'Alexander Fleming') talked about integrating computational and experimental data in modeling the function of non-coding RNA's from high throughput data and integrating miRNA's in pathways. **Tom Michoel** (FRIAS) *presented* computational methods to reconstruct regulatory modules and pathways by integrating perturbational expression data and large-scale networks of protein-DNA, protein-RNA, protein-protein and phosphorylation interactions. **Yves Moreau** (KU Leuven) talked about approaches to prioritize candidate genes for genetic disorders by identifying those that are located among the most heavily perturbed subnetworks when comparing expression data from healthy vs. affected individuals on a network basis. **Johan Bjoerkegren** (Karolinska Institute) talked *about* understanding risk for coronary artery disease using intermediate phenotypes from multiple organs. Integrating gene expression profiles with genome-wide DNA variation to uncover gene networks of coronary artery disease and atherosclerosis. It was discussed that many algorithms rely on training sets of positive examples to guide the search. This allows to find genes that look similar to already known ones, but may miss important novel features of the data. Participants discussed the different time-scales on which experimentation and computation operate: Once there is a good algorithm for a particular data type, a new technology producing different data has appeared (computation lags behind experimentation); but on the other hand even the best computational predictions take years to validate (experimentation lags behind computation).

Diego Di Bernardo (Telethon Institute) discussed recent results on the use of gene expression data to elucidate the function of disease genes and to characterise the mode of action of a drug. He presented (i) a large co-expression network predictive of gene function and (ii) mode-of-action analysis using the *connectivity map* dataset. **Rainer Spang** (University of Regensburg*)* showed what does the nesting of perturbation effects and the time gaps between perturbation and observation of effects can tell us about cellular networks. **Julio Saez-Rodriguez** (EMBL-EBI HInxton) discussed how generic network information encoded in pathway literature and databases can be integrated with functional biochemical data of signal transduction to construct cell-specific pathway models. The presentations were followed by an intense discussion about the status of *networks* and *models* in biological and medical research. Are networks more than just visualizations of data? Participants generally agreed that networks should be *predictive*, *integrate* different data types and *weight* them by relevance and consistency. In this way, networks make the step from visualizations to models. But models *of what*? Participants' views ranged from models of *biochemical reactions* to models of *(unspecified) functional relationships*. It seemed to us that these questions are at the heart of the questions what advantages systems approaches (in contrast to 'single gene approaches') have to offer.

3. Assessment of the results, contribution to the future direction of the field, outcome

Our key question was: **Which (computational or experimental) developments will drive the field forward in the next years?** The participants' individual answers were very diverse and differed according to their scientifc background (computational or experimental), model organism used and research field (basic biology or medical research). However, in the discussion periods some common patterns began to emerge:

- **We need richer phenotypes**. Many speakers stressed the higher resolution of gene function resulting from exchanging single-reporter read-outs by richer phenotypes, e.g. cell morphology or molecular read-outs like global gene expression.
- **We need more medium-scale studies**. Genome-wide screens allow a global overview of cellular phenotypes, while single-gene studies allow in-depth analysis of individual phenomena. Between these extremes lie many so far mostly unexplored opportunities for medium scaled studies, which are more focussed than genome-wide screens but still provide a broader view than single-gene studies.
- **We need to integrate many different data types**. Participants stressed the importance of integrated analysis of heterogeneous data. Each individual phenotype can only explain parts of the cellular system. Only by integrating screening results with, e.g. protein interaction or transcription factor binding data can the mechanisms underlying the phenotypes be uncovered.
- **We need better algorithms**. An integrated analysis poses challenges for data analysis and computational method development: To amplify the signal and not the noise by data integration it is necessary to automatically judge the relevance and information content of individual data sets. While first approaches addressing this issue exist, they have so far not been effectively leveraged for the analysis of large gene perturbation screens

Concrete actions. We are pursuing several concrete actions as an outcome of this workshop:

- We got very positive feedback from participants at the workshop and later by email. They all felt that the scientific focus, the wide range of expertises and the quality of discussions set this workshop apart from other conferences. As a result, we are now exploring ways to **repeat the workshop in Sept 2011.** We plan to invite again 25-30 participants but to also include contributions from selected postdocs or students.
- Several participants plan to jointly propose a **FP7 Initial Training Network** with a focus on systems genetics.
- **Cambridge University Press** has agreed to publish a volume edited by the convenors, possibly in the newly established Cambridge Series on Systems Genetics. Contributions will come from workshop participants and additional international experts.
- The convenors plan to publish a short **workshop summary in a scientific journal** (like e.g. Genome Biology or the EMBO reports).

- **4. Final programme**

Thursday, 9 September 2010

Morning	*Arrival*
12.00-13.30	*buffet lunch*
13.30-14.00	**Welcome by Convenors** **Florian Markowetz** and **Michael Boutros**
14.00-17.00	**Afternoon session** Chair: **Michael Boutros**
14.00-15.10	Julie Ahringer, Julian Downward, Roderick Beijersbergen
15.10-15.30	*coffee break*
15.30-16.40	Lucas Pelkmans, Jean-Philippe Vert, Rengul Cetin-Atalay
16.40-17.30	General discussion Leaders: **Wolfgang Huber** and **Jussi Taipale**
18.00-21.00	**Get-together in The Anchor**

Friday, 10 September 2010

09.00-12.00	**Morning session** Chair: **Florian Markowetz**
09.00-10.10	Michael Boutros, Sebastian Nijman, Balazs Papp
10.10-10.30	*coffee break*
10.30-11.40	Chad Myers, Wolfgang Huber, Blaz Zupan
11.40-12.15	General discussion Leaders: **Tom Michoel** and **Johan Björkegren**
12.30-14.00	*lunch break*
14.00-17.00	**Afternoon session** Chair: **Rainer Spang**
14.00-15.10	Thomas Meyer, Buzz Baum, Cecile Arrieumerlou
15.10-15.30	*coffee break*
15.30-16.40	Niko Beerenwinkel, Florian Markowetz, Jasmin Fisher
16.40-17.30	General discussion Leaders: **Chad Myers** and **Sebastian Nijman**
18.00-19.00	*dinner*
19.00-...	**At the bar: How to integrate experiments and computation?** Chair: **Michael Boutros** and **Florian Markowetz**

Saturday, 11 September 2010

09.00-12.00	**Morning session** Chair: **Thomas Meyer**
09.00-10.10	Jussi Taipale, Anna Gambin, Artemis Hatzigeorgiou
10.10-10.30	*coffee break*
10.30-11.40	Tom Michoel, Yves Moreau, Johan Bjoerkegren
11.40-12.30	General discussion led by **Jean-Philippe Vert** and **Julie Ahringer**
12.30-14.00	*lunch break*
14.00-17.00	**Afternoon session** Chair: **Yves Moreau**
14.00-15.10	Diego Di Bernardo, Rainer Spang, Julio Saez-Rodriguez
15.10-15.30	*coffee break*
15.30-16.40	General discussion led by **Niko Beerenwinkel** and **Roderick Beijersbergen**
16.40-17.30	**Final discussion: plans for follow-up research activities** Chair: **Michael Boutros** and **Florian Markowetz**
18.00-19.00	*dinner*
End of workshop	

Sunday, 12 September 2010

morning	*Departure*

5. Final list of participants

1	Ahringer, Julie	University of Cambridge	UK	Exp
2	Arrieumerlou, Cecile	Biozentrum Basel	CH	Exp
3	Baum, Buzz	LMCB UCL	UK	Exp
4	Beerenwinkel, Niko	ETH Zurich	CH	Comp
5	Beijersbergen, Roderick	NKI Amsterdam	NL	Exp
6	Bjoerkegren, Johan	Karolinska Insitute Stockholm	SE	Exp
7	Boutros, Michael	DKFZ Heidelberg	DE	Exp
8	Cetin-Atalay, Rengul	Bilkent University	TR	Exp
9	Di Bernardo, Diego	Telethon Institute, Naples	IT	Comp
10	Downward, Julian	CR-UK London Research Institute	UK	Exp
11	Fisher, Jasmin	Microsoft Research Cambridge	UK	Comp
12	Gambin, Anna	University of Warsaw	PL	Comp
13	Hatzigeorgiou, Artemis	BSRC Alexander Fleming	GR	Comp
14	Huber, Wolfgang	EMBL Heidelberg	DE	Comp
15	Markowetz, Florian	CR-UK Cambridge Research Institute	UK	Comp
16	Meyer, Thomas	MPI for Infection Biology, Berlin	DE	Exp
17	Michoel, Tom	Freiburg Institute for Advanced Studies	DE	Comp
18	Moreau, Yves	KU Leuven	BE	Comp
19	Myers, Chad	University of Minnesota	US	Comp
20	Nijman, Sebastian	Research Center for Molecular Medicine	AT	Exp
21	Papp, Balazs	Hung Acad of Sciences, Szeged	HU	Comp
22	Pelkmans, Lucas	ETH Zurich	CH	Exp
23	Saez-Rodriguez, Julio	EMBL-EBI	UK	Comp
24	Spang, Rainer	University of Regensburg	DE	Comp
25	Taipale, Jussi	Helsinki University	FI	Exp
26	Vert, Jean-Philippe	Ecole de Mines, Paris	FR	Comp
27	Zupan, Blaz	University of Ljubljana	SI	Comp

6. Statistical information on participants

- The participants were almost equally split between computational (15) and experimental scientists (12).
- Six of the participants were female.
- Participants came from fourteen countries (AT, BE, CH, DE, FR, GR, IT, NL, PL, SE, SI, TI, TR, UK, US).
- Most of the participants are leading very established labs, but some of them have just started their first independent position in the last couple of years (Beerenwinkel, Gambin, Markowetz, Michoel, Nijman, Saez-Rodriguez) – this provided a good mix of different experiences and ideas.