Genomics in 2011: challenges and opportunities

Citation: Adams, David et al. "As We Come to the End of 2011, Several Members of the Genome Biology Editorial Board Give Their Views on the State of Play in Genomics." Genome Biology 12.12 (2011): 137. Web. 3 May 2012. © 2011 BioMed Central Ltd.

As Published: http://dx.doi.org/10.1186/gb-2011-12-12-137

Publisher: Springer (Biomed Central Ltd.)

Persistent URL: http://hdl.handle.net/1721.1/70494

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Genome Biology 2011, **12**:137
http://genomebiology.com/2011/12/12/137

Genome Biology

OPINION

Genomics in 2011: challenges and opportunities

Abstract

As we come to the end of 2011, *Genome Biology* has asked some members of our Editorial Board for their views on the state of play in genomics. What was their favorite paper of 2011? What are the challenges in their particular research area? Who has had the biggest influence on their careers? What advice would they give to young researchers embarking on a career in research?

What in your opinion was the most important paper published in your field in the past year?

DA: I have a relatively broad area of scientific interest encompassing cancer genetics, mouse genetics and genome sequencing. With my cancer hat on, the paper that impressed me most was by Sodir *et al.* [1], which showed that inhibition of Myc can cause cancer regression even in advanced tumors. This work illustrates the critical role that *Myc* plays in tumorigenesis, and clearly defines it as a therapeutic target. In terms of genome sequencing, I'm very impressed by some of the new *de novo* assembly algorithms, such as Li *et al.* [2], and I think our contribution of sequencing mouse genomes, Keane *et al.* [3], is also important and will be of great use to the mouse genetics community. These sequences now make it possible to take a systems biology approach to mouse genetics and to link variants to phenotypes like never before.

BB: I would say the Foldit paper by David Baker's group [4] because it shows a new model of research is possible. Some large-scale problems such as protein folding remain challenging to solve with just computational approaches, and the problem is too difficult for even experts to solve manually. However, by designing tools that can break such problems into puzzles that people, even non-experts, can play with in their spare time can provide cutting-edge solutions [5].

OH: In cancer genetics, two studies struck me this year. The first is the discovery and analysis of chromothripsis by Peter Campbell at the Sanger Institute [6]. This idea that some of the complex chromosomal rearrangements observed in cancer cells can come from a single catastrophic event, where an entire chromosome gets shattered and stuck back together at random, is astounding. This has changed our understanding of cancer genome instability and repair: a truly novel finding and groundbreaking idea [7]. The other cancer genetic study is the work of Inder Verma [8], Rusty Gage and colleagues at the Salk Institute [9] demonstrating a new function for *BRCA1* involved in heterochromatin chromatin formation and repressing satellite expression via the ubiquitylation of histone H2A [10]. This is beautiful work, which might deeply change our understanding of *BRCA1*'s role in cancer. This adds to the numerous and recent evidence of the importance of chromatin-modifying processes in cancer.

CH: One important recent paper is the work on the vaginal microbiome in reproductive-age women by Jacques Ravel and Larry Forney [11]. This is a very well-structured and well-analyzed study that includes a variety of important findings. It establishes that diversity in the human microbiome is personalized and much more than 'noise', and that following individual subjects longitudinally can be extremely informative. It shows that the microbiota can be linked both to host phenotype and to host environment, but that neither of these alone is the whole story. In combination with other work, it shows that beyond community composition, the complete picture of microbial ecological structure, function and dynamics differs widely among humans and human body habitats. And it provides evidence that microbial community structure and its interaction with phenotype have potential links to genotype by way of racial background or ethnicity. Finally, all of these results together suggest ways by which personalized medicine can be influenced by the microbiome. Many of these are general results that have been shown in one or more previous studies, but this paper brings them all together nicely and applies them quite interpretably in the specific microbiological context of the vaginal community.

SL: There have been many interesting papers in the last year. One paper from the Dekker and Young labs [12] demonstrated that mediator can form a complex with cohesin to connect enhancers and core promoters of active genes in mouse embryonic stem cells. Predicting targets from transcription factor ChIP-seq has been a challenging problem, and perhaps mediator and cohesin binding could offer some insights into which bindings are functional with transcriptional outcomes. Another paper, from Howard Chang's group [13], developed a new technique called 'ChIRP-seq' to identify the *in vivo* genome-wide location of long non-coding RNAs and understand their chromatin regulation functions. The Chang lab has shown exceptional ingenuity in both genomic technology development and biological mechanism discovery in recent years. A third paper, from Peter Laird's group [14], showed that many cancer samples have small regions of hypermethylation within bigger domains of hypomethylation, and these domains coincide with LAMIN binding domains. It demonstrates the power of correlating unpublished genomics and epigenomics data with publicly available data to make interesting discoveries.

CM: A recent paper that I think demonstrates an important point about mapping insights from functional genomic data in model organisms is McGary *et al.* [15], in which the authors define the concept of 'phenologs', which are pairs of phenotypes across two species that share a set of orthologous genes. The key observation is that sets of functionally related genes are often conserved as functional modules across species, but that their corresponding phenotypes in the two species might have no obvious connection. Thus, after constructing a phenolog map pairing phenotypes across species, the authors show that one can accurately predict new genes related to phenotypes in higher eukaryotes based on functional genomic data from more genetically tractable model organisms. For example, they identify a phenolog connecting yeast sensitivity to the drug lovastatin to abnormal angiogenesis in mice. This connection then enabled them to predict new genes with a role in angiogenesis based on yeast genes that showed the lovastatin phenotype, and the prediction was confirmed in the frog *Xenopus*. This is an important paper because it provides a model for how we can leverage the wealth of data produced in model organisms that are more genetically tractable to gain insights about higher organisms and disease. The method itself was very simple but based on a powerful idea, and is one that will work in a number of different contexts.

AO: A lot of my recent work has been on the analysis of RNA-seq data. There are many opportunities provided by sequencing the transcriptome that have not been possible with previous technologies, such as detecting all expressed transcripts in a sample and annotating them, determining splicing variants and how isoforms change between samples, looking for allele-specific expression and looking at RNA editing. However, I think there is still a lot of work to be done in order to develop analytical methods to capture all the different types of information that we might want to obtain from an experiment. Each type of question requires different analysis. One of the problems with developing analysis methods is actually showing that the method is doing the right thing and understanding possible biases and limitations of analysis strategies. In this context I believe a recent paper on spike-in standards for RNA-seq experiments will be very influential [16], as these sorts of data sets enable us to identify and study biases in data generation and analysis and allow us to assess and compare methods. Hopefully more of these 'truth' data sets will be generated to enable us to further develop our understanding of the technology.

JR: Two recent papers have shown that a long noncoding RNA termed Coldair plays a key role in regulating plant flower formation [17,18]. What's remarkable is that this RNA 'senses' temperature and indicates to the plant when it is warm again and time to flower. This is an incredible example of sensing the environment and making epigenetic decisions. It also highlights the importance of reading primary plant literature; discoveries in plant science are often a decade ahead of the curve!

MW: There are so many wonderful papers in systems biology, which is by now quite a large discipline. It is hard to nail down only one; my favorites that are destined to become classics include a paper from Frank Holstege's group [19] that identified different network topologies that can ensure robustness through paralog redundancy in yeast genetic and signaling networks. This paper is a beautiful example of true, integrative systems biology that takes advantage of different techniques, available datasets and modeling to gain systems-level insights into signaling network architecture. The second paper is from Karen Oegema's group, in collaboration with Fabio Piano and Kristin Gunsalus [20]. This paper utilizes the power of systems biology in the worm by combining phenotypic profiling with high-resolution imaging of defects in the gonad, and a new clustering method that provides unprecedented specificity. Numerous functional annotations of genes were obtained by delineating more than 100 distinct and highly detailed gonadal phenotypes.

Who has had the most influence in your career so far?

DA: I have worked with many wonderful colleagues. I've fed off the contagious enthusiasm of Jos Jonkers [21] and Jonathan Flint [22] from Amsterdam and Oxford, respectively. Both of them are outstanding scientists and fabulous collaborators who never seem to tire and always make me do better science. At the Sanger Institute, I've been fortunate to work with Louise van der Weyden. No experiment is too difficult or demanding for Louise, and what she doesn't know about the study of cancer genes in mice is not worth knowing! I've also learnt a lot from several people who I've worked with who were incapable of managing people. While I'm sure I don't always get it right, I try hard to motivate people with the carrot rather than the stick and to make the lab a happy place to work.

BB: I would say there were three people who have most influenced my career thus far, one mathematician and two biologists: my postdoctoral mentor Daniel Kleitman [23], who was Professor of Applied Mathematics at MIT, told me 'Proteins, that's what you should do!' when I was looking for interesting areas to apply my algorithms background - he encouraged me to initiate comparative genomics; Jonathan King, also at MIT [24], who taught me much about protein folding and helped guide my group meetings through the first years; last but not least, Peter S Kim, President of Merck Research Laboratories [25], who worked tirelessly with me to define the new brand of research and publications in the burgeoning field of computational biology.

OH: I am happy and fortunate to have worked with several brilliant and enthusiastic scientists, but I acknowledge the wonderful interactions I have had with Dr Kelly Frazer [26] for the past 4 years. Beyond a fully accomplished genomic biologist, she has been an exemplary mentor, concerned both with traditional academic training, such as grant application and scientific communication, and also with more practical mentoring on how to manage successful and highly collaborative genomic projects.

CH: Definitely my PhD and postdoctoral adviser, Olga Troyanskaya [27]. Just like we all grow up to resemble our parents, I've academically inherited much of Olga's lab management style. Although we're working in very different scientific areas now, my approaches to mentoring, organization and presentation remain heavily influenced by hers at Princeton. In terms of scientific content, I give a great deal of credit and respect to Dirk Gevers at the Broad Institute [28], who's been a phenomenal collaborator on many of the projects my group has tackled since starting my own lab. The best collaborations are with folks who both know the science and have fun doing it, and Dirk and his team have been stellar on both counts.

SL: Three people had the biggest influence on my career. The first is my PhD mentor, Jun Liu, who showed me the fun of methodology development in computational biology, which convinced me to adopt an academic career. Jun has continued to help me over the years in research projects and career decisions, and taught me how to be a good mentor. The second person is my first collaborator, Jason Lieb, from whom I learned a lot about writing, leadership, and working as a team in a consortium. Last but most importantly, I had a great collaborator, Myles Brown [29], now at Dana-Farber Cancer Institute. We started collaborating in 2004, and have published 26 papers together. Both of us are passionate about new genomic technologies and gene regulation questions, and it is really fun working together. We brainstorm on how to better adopt new genomic techniques, generate the data, develop the computational method, and apply them to important biological problems. I have learned so much from Myles and his postdocs, and some of our postdocs continue to collaborate with each other after they become independent. Myles is really generous in giving credit and resources to my group. He is also a role model for balancing work and life, and has given me much good career and life advice.

CM: My PhD advisor, Olga Troyanskaya [27]. I started my graduate studies at Princeton with an entirely computational background and with little direction in terms of my research focus, other than being interested in machine learning, signal processing and data mining. Olga introduced me to the exciting genomics revolution that was well underway and was willing to take a chance on me, even though I knew very little biology and had no previous experience in computational biology or bioinformatics. In her mentoring, Olga emphasized establishing a solid foundation in computer science and statistics but also developed our skills to identify relevant and impactful biological questions, one of the biggest challenges of transitioning from a computer scientist to a 'computational biologist' in my opinion. She established a lab environment at the Lewis-Sigler Institute for Integrative Genomics that was centered on collaboration with experimentalists and developing practical bioinformatics solutions to problems faced in these collaborations. This perspective permanently shaped me as a scientist, and I certainly would not be where I am today without Olga's mentoring.

AO: I have been very fortunate to have several great mentors in my career so far who have all been very influential over different aspects of my life and career. It's very hard to single out people but I will just focus on two. Firstly, my PhD supervisor, Professor Rachel Webster [30], has been very supportive of me and my career, giving me advice and guidance for the last 15 years. I did my PhD in astrophysics and when I was considering leaving the field she fully supported that decision. Since then she continues to both share her experiences and give me advice on many issues that are important to a research career, such as leading a research group, conducting collaborations, managing my time, employing people, and effectively combining a family and a successful career. Even though we no longer work in the same discipline, I value highly her experience and advice and make an effort to catch up with her every few months. The second person who has had a major influence on my career is Professor Gordon Smyth [31]. He was the one who first gave me a postdoc position in bioinformatics even though I had no specific background in the area. He taught me the ropes of bioinformatics and statistics and guided me through the initial years in this field. I will always be grateful to him for taking that initial risk on me as I find genomics absolutely fascinating.

JR: It's almost impossible to single out one person. My career has been guided by amazing mentors such as Michael Snyder, Howard Chang and Eric Lander [32-34]. Outside of my mentors, Linus Pauling had a strong influence on me scientifically. I was motivated by him early in my undergraduate chemistry training. His work was an infectious balance of scientific vigor and rigor, analytical yet creative interpretation with pioneering vision. His profound scientific contributions were equally met with his political importance. His humble upbringing, ability to admit when he was wrong, and persistence for what was right has continued to inspire me.

MW: My postdoctoral mentor and friend Marc Vidal [35]. He is very creative and has an open, yet critical, mind. He continues to push the envelope in the field while sticking to his ambitious objective to attain complete protein-protein interaction networks in a variety of model systems.

What advice would you give to young scientists starting off in a research career today, or what advice would you give to your younger self?

DA: Science needs to be your passion and you really need to love it and jump in with both feet. I'm not talking about turning up and doing a few experiments between coffee breaks but making your science one of the most important things you do. I don't know of any successful scientist who periodically doesn't find themselves spending virtually every waking hour in the lab or in their office. Surrounding yourself with people who are smarter than you also helps you raise your game.

BB: Published is better than perfect: it's advisable to communicate intermediate results while you work to polish your system. Also, computational biologists should try to become familiar with wet lab work in their chosen area of research. Lastly, make the algorithms you develop available through tools that others can easily use.

OH: Beyond showing the necessary scientific achievements through regular publication, I would emphasize the importance of communication. Communication is crucial to build a network and learn important lessons for a successful scientific career. Talk to your mentor(s) with an open mind and build a trusted and honest relationship with your colleagues. I have seen too many people frustrated due to communication issues with their mentor or colleagues. Big science is team science. If it does not work out, it is OK to move on. No one can blame you for it, as long as you are honest with yourself and others. Also, be curious of other people's work. When going to a couple meetings a year, make sure you talk to the senior investigators whose work you admire, and do not hesitate to tell them so. The next thing you know they will be writing you support letters and becoming collaborators.

CH: Stereotypes are often true, both the good ones and the bad ones. As a faculty member, there can be a lot of politics, you do spend a lot of time asking for money, and you are even busier than you were before. But on the flip side, you really can 'do anything', both in terms of choosing exciting scientific directions and in shaping your group and your environment. If there's a question you want to work on, you can - there's nothing to stop you except your own motivation to recruit a team and funding for the investigation. And there's no one right way to do it - lab size, teaching/training balance, approach, methods, impact and sales pitch are all up to you. Take advantage of your time as a student or postdoc to do as much of your own work as possible, though, since you'll likely have fewer and fewer chances to perform the investigations with your own hands as time goes on.

SL: In computational biology, you need to appreciate the biology as much as, if not more than, the statistical method or computer science. Computational genomics is an applied science with strong technology components. When starting out, get your hands as 'dirty' as possible and as quickly as possible with data; that is, dive into big datasets. Without seeing enough data in detail and understanding the data characteristics, you won't be able to develop good computational biology algorithms. Find a good and niche research direction that: (i) is important in the long run; (ii) you have a real interest in; (iii) maximizes your existing expertise (for example, previous training) and advantages (for example, timing, location, connections and other available resources).

CM: My advice to young computational biologists is that the key to success in our field is to remain grounded in specific biology questions, particularly ones that are pursued in collaboration with experimentalists. The success of our field is ultimately measured by the impact we can have on our understanding of biology, and how we spend our time should reflect that. In my opinion, too many computational biology researchers are working in isolation on marginally relevant problems or making incremental improvements in areas that have already been well-populated by methods that are already adequate. Meanwhile, there are pressing 'big data' challenges being faced by our experimental colleagues, including everything from collection and normalization of massive datasets to scalable methods for integration of heterogeneous data types. Often such collaborative projects involve some amount of what is often considered service work (for example, processing/normalization of raw data, applying existing tools to new data, and so on), but this requires computational expertise and can have a major impact. Furthermore, contributing our efforts to these service-oriented tasks almost always leads to exciting new issues and methods that generalize to other scenarios. The most exciting and impactful research is being accomplished by synergistic cross-disciplinary teams, so if you want to have an impact get involved in this sort of research. Finally, remember that building collaborative relationships takes time - establishing the specific relevant expertise, effective communication and trust that are necessary for successful collaborations is hard work - but investing time in honing these skills will pay off and can serve as the basis for a successful and rewarding career.

AO: I think one of the most important skills in research is the ability to communicate ideas. My advice would be to spend time practicing both writing skills and oral presentation skills. There is no point making great discoveries if no one knows about them or uses your discoveries. Therefore, your ideas need to be communicated effectively. I think it is very important to understand the audience that you are communicating to and work out effective ways to explain concepts, especially in the current environment where there are many multidisciplinary teams with different background knowledge. What may seem obvious to you may not be obvious to other people that are potentially interested in your work. Some people have a natural talent for communicating their ideas but for the majority of us it's something we need to spend time working on. My way of doing this, at least initially, was to look at papers that I really like reading and work out why I think they are good and try to emulate the style in my own work. With regard to oral presentations, I always practice them out loud several times and try to get feedback from colleagues on which parts worked and which weren't clear or interesting.

JR: I try not to give too much advice as I have seen many diverse paths met with great success in science. If there was one thing, I would say it's the ability to understand *both* the experimental and the computational sciences. Modern biology is becoming a seamless integration of these two disciplines. If trained as an experimentalist, I would learn the key principles in computational biology or at the very least the linguistics, and *vice versa*.

MW: Put yourself 'out there': ask questions at meetings, build a scientific network with colleagues, do not sacrifice quality for quantity of data. Be fearless and ask hard questions - and try to answer them!

What in your opinion are the top three challenges in your field right now (and what progress is being made to address them)?

DA: In cancer genetics the biggest challenge is integrating data from genome sequencing, transcriptomes and the epigenome, so that it makes sense. The problem is no longer acquiring the data but 'embracing the chaos'. Some of the Boolean logic approaches are making inroads in this area. In mouse genetics, it's all about engineering the mouse genome faster. The TAL nucleases, which can be used to tailor the mouse genome with base-pair precision, potentially represent a big advance in this area, but we need to understand if they have off-target activity. In genome sequencing, read length really matters for assembly and while the last few years of short-read sequencing have been amazing we really need long-read technology that is truly scalable and accurate (and cheap).

BB: The mission of computational biology is to answer biological and biomedical questions by using computation in support of or in place of laboratory procedures, with one goal being to get more accurate answers at a greatly reduced cost. Three major emerging challenges are: how to make sense of massively accumulating data, how to develop reasonable gold standards for testing our algorithms, and how best to integrate computational studies with real biological experiments (on both sides).

The past two decades have seen an exponential increase in genomic and biomedical data, which will soon outstrip advances in computing power to perform current methods of analysis. Extracting new science from these massive datasets will require not only faster computers; it will also require smarter algorithms. Moore's Law has been a great friend of computational biologists: the amount of processing you can do per dollar of compute hardware is more or less doubling every year. Back in the 1990s, the growth rate of genomic data was balanced by the growth rate of computing speeds. However, one way this balance is being disrupted is by the advent of next-generation sequencing. The size of genomic databases is going up by a factor of 10 every year, far outstripping the growth in our computational capacity. It's tempting to think that cloud computing is going to solve this problem, but that's not the case. It doesn't change the problem that the data are increasing exponentially faster than computing power per dollar. The only solution is to discover fundamentally better algorithms for processing these databases. Better algorithms can make an enormous difference. In fact, you've got to devise algorithms that are so fast that, in some cases, they can't even grow linearly in the size of the databases.

Another big challenge in computational biology is the determination of gold-standard datasets for training computational techniques. For example, consider the problem of determining orthology relationships across species. What we really want to identify are functional orthologs (that is, genes that perform the same functions across various species). Direct experimental data about this are scarce. The most commonly available datasets capture this only indirectly by looking at, say, sequence similarity between the genes. There are many computational approaches that use this indirect data to predict such orthology relationships, but determining which one works best is difficult. One direction we have been exploring is using protein and genetic interaction data to improve orthology prediction by better capturing function correspondence. It would really help if we had even a limited set of proteins for which gold-standard orthology information was available. All the computational techniques can then be better trained. However, we are still some distance away from having any gold-standard orthology sets. There are many other problem domains where being able to generate good gold-standard datasets would significantly improve our ability to use computational methods.

The final challenge is the need to improve the integration of biological and computational methods. In some domains, algorithmic thinking is already very tightly integrated into the process of experiment design, execution and interpretation. Genome sequencing is a great example of where such integration has yielded great success. In other cases, however, biological methods use computational analysis only as an afterthought; for example, many studies of cell signaling could benefit greatly from having knowledge of innovative computational techniques applied early in the design stage, so that the right data are available to enable the full power of these methods to be applied. The converse of this criticism also applies to computer scientists. We need to have a better understanding of the subtleties of various biological experiments. Far too often, enough biological details are abstracted away so that the solution loses its biological relevance.

OH: Access to adequate clinical research samples in cancer genetics is one of the most important challenges in cancer genetics. Collection of samples by biopsy or surgical resection has been traditionally performed for clinical care only. It is currently extremely hard to use the same samples for research for various reasons, from preparation methods, to logistic or consent. Several institutions like the University of California, San Diego are developing master protocols to systematically consent a majority of oncology patients and collect samples from surgery or biopsy for investigational purposes. The resistance is high, in general legitimated by the patient's protection, but people start to understand that it is the only way to eventually deliver the promises of personalized diagnostics and care.

Another challenge is to educate people about genomics and to tone down the natural hype of the genomics field. Investigators involved in clinical and translational projects are in some way victims of the hype created by the fantastic and recent technological advances. I frequently talk to clinicians who are enthusiastic about sequencing their samples, but many projects often fall short due to the ignorance of the requirement of sample number or quality. For example, there is no good rationale to sequence the whole genome of thousands of samples except to make it a general resource for the community. Biological questions might be better addressed with a more focused approach, such as sequencing exons or candidate regions in properly selected patients. The hype of the sequencing field is a wonderful catalyzer of novel ideas and provides much needed public exposure of our field, but we have to regularly educate prospective collaborators on basic notions of genetics or the reality of the sample preparation or data analysis. At our institution, my function in the University of California, San Diego, Clinical and Translational Research Institute (a Clinical and Translational Science Awards funded entity) [36] is to do just that: consult with people and help them with the design, preparation and analysis of their translational genomic experiments.

Finally the last challenge is to transform the academic review system in our institutions. Traditional institutions expect faculty to lead independent projects typically funded through the R01 NIH grants. However, genomics has traditionally functioned differently, following the principle of team science, where multiple principle investigators contribute to a large endeavor. This was the case for the Human Genome Project and The HapMap project, and today the 1000 Genome Consortium and The Cancer Genome Atlas, for example. This 'big science' is usually financed through alternative sources of funding requiring collaborations and multiple principle investigators, and the results do not always lead to first or last author publications for the majority of the participants despite their essential roles. Traditional institutions that promote faculty based on R01 awards and last author publications do not always recognize this aspect. This divergence does not favor the retention of brilliant researchers in academic genomic research. Some institutions, such as Harvard or The Ontario Institute for Cancer Research, have established alternative academic review criteria that recognize participation in team science and allow investigators to successfully grow in this environment. At the time when funding is becoming scarce and more directed to specific projects, let's hope that more institutions will follow these examples.

CH: (i) Understanding the systems-level ecological rules governing microbial community structure, (ii) relating the human microbiome to health and disease, and (iii) streamlining methods for turning next-generation data into actionable biology. Addressing the combination of the first two challenges will help us realize some of the human microbiome's potential as a means of diagnosis and therapeutic intervention. Investigating the first challenge in particular should let us leverage systems biology's successes in molecular biology during studies of microbial communities. The second will likewise feed back into the broader metagenomics community by identifying 'interesting' microbiome properties, environments and phenotypes on which to focus. Finally, the third challenge includes finding ways to collaborate on biological 'big science' projects, to collectively analyze sequence data (of all sorts, not just metagenomic), and to leverage shared computing resources. All of these continue to be necessary to solve the considerable data management and interpretation challenges brought about by next-generation sequencing technologies. These technologies will continue to accelerate biological discovery - but there are still many opportunities for computational methods to accelerate that acceleration.

SL: Now that it's possible to profile transcription factor binding using ChIP-seq, the ability to predict the target genes and the direction of their expression changes upon factor activation or inactivation is still an important challenge. For factor binding, there are often thousands of genes nearby binding, but only a minority of the nearby genes really show differential expression and we don't know why. Also, for transcription factors with multiple functions such as CTCF (for example, transcriptional repressor and insulator) the challenge is whether we can differentiate their functions from ChIP-seq of other factors or histone marks. Approaches such as HiC and ChIA-PET can identify genome-wide higher-order chromatin interactions, which have the potential to answer this question, although there are still technical and cost challenges for these techniques to be widely adopted.

Performing ChIP-seq or DNase-seq with a small amount of starting material is a challenge. Currently one needs 100,00 to 500,000 cells to do a histone mark ChIP-seq, and 1 million to 2 million cells for transcription factor ChIP-seq. To make ChIP-seq or DNase-seq work well on tissues or tumors, it is important to start from smaller numbers of cells. The laboratories of Peggy Farnham [37] and Brad Bernstein [38], and many other laboratories, have explored this issue. Recently the Gronemeyer group published a new method to linear amplify picogram DNA [39]. Commercial companies like Illumina are developing kits for library construction from <1 ng of DNA, and third-generation sequencing techniques promise to offer a better solution to working with small amounts of starting material.

Finally, there are many transcription factors, chromatin-modifying enzymes and histone marks functioning together to regulate gene expression. The specificity (for example, which transcription factors specifically recruit which histone marks or histone modifying enzymes) and the cooperativity (for example, which factors are pioneering factors for the binding of other factors) of these factors in different cells or conditions are still poorly understood. Without understanding this, the effect of epigenetic drugs could be hard to interpret. As sequencing technologies increase throughput, multiplex ChIP-seq would allow us to investigate many more conditions in combination and we might have a better answer for this question.

CM: In my specific area of interest, genetic interaction networks, there are a few challenges we face as a community. (i) Scalable technology for mapping genetic interactions for other phenotypes, conditions and organisms, especially higher eukaryotes. The yeast community has been very successful in the past several years at developing technology for rapid construction of combinatorial mutants to map genetic interactions. Specifically, the typical approach is to look for combinations of mutations that result in a surprising phenotype (usually fitness defect) given the phenotypes of the mutations introduced independently. These efforts have produced global interaction maps covering millions of combinatorial mutants, which have proven to be quite useful for understanding gene function and general organization of the cell. In yeast, efforts are underway to expand these maps to other phenotypes and other conditions; this requires new scalable technologies given the space of possible experiments. Such maps in higher eukaryotes will be important for understanding the genetic basis for complex phenotypes and disease and developing new therapeutic approaches, but the technology for mapping these interactions is still relatively limited in throughput. Several exciting efforts are underway to address this challenge, most of them leveraging RNA interference technology. The past year has produced new successes in *Drosophila* and human cell lines but continued focus on improving and scaling the technology will be fruitful. (ii) Translating insights about genetic interactions from perturbation studies to questions in population genomics. The focus of the genetic interaction community has largely been on precise combinatorial genetic perturbation in single individuals (for example, standard lab strains). This approach is attractive because the effects of perturbations can be studied in a controlled genetic background. However, we would ultimately like to leverage this knowledge about how genetic variations combine to influence phenotype to understand the link between genotypic and phenotypic variation across individuals in a population. The latter challenge is of course the main goal of genome-wide association studies in humans; to date these have struggled to explain large portions of the heritable phenotypic variation. Applying insights derived from large-scale perturbation studies to the population genomics questions will be an interesting direction, especially as the mapping technologies become more feasible in higher eukaryotes. There are new opportunities and the necessary data to make progress on this front in yeast with the recent sequencing and phenotyping of several *Saccharomyces cerevisiae* strains. The combination of this information on genomic and phenotypic variation, combined with extensive functional studies on the reference genome, will provide a good testing ground for new methods in this area. (iii) Leveraging functional genomic data across species. As I noted above, a more general challenge is the problem of leveraging functional genomic data across species to speed the process of functional characterization. Even the most basic question in systems biology, 'What are all of the genetic components related to biological process X?', has not been answered comprehensively in most species, particularly in higher eukaryotes. Enormous resources have been spent generating functional data in model organisms, but these data are relatively underutilized for mapping functions in other species. The paper from McGary *et al.* [15] provides a nice demonstration of how insights from relatively data-rich model systems can be used to direct experimental investigation of genes related to specific phenotypes in more complex organisms, and I suspect similar approaches can be developed in other settings. Accomplishing this will require new computational infrastructure and tools to support integration and comparative analysis of functional genomic data.

AO: Getting a handle on the propensity and type of RNA editing that is occurring is a fascinating area which, as yet, has not been fully resolved [40]. It has been documented that the sequence of RNA can be modified post-transcriptionally, resulting in an RNA sequence that is different from the DNA from which it was derived. High-throughput sequencing technologies give us the opportunity to study RNA editing on a genome-wide scale and there have been several publications recently on this topic [41,42]. However, there is quite a debate about how frequently this actually occurs. In my view, results are probably influenced by biases in mapping procedures (see Joe Pickrell's blog [43] and the recent paper by Schrider *et al.* [44]) and it will be fascinating to see how this debate gets resolved in the near future and what the results mean for the diversity of the transcriptome.

There are many projects producing massive amounts of sequencing data. One of the major scientific challenges right now is the integration of different types of data to explain a biological phenomenon. For example, the ENCODE project is producing genome-wide expression, trascription factor and epigenetic data on many different cell types [45]. Making sense of even just a small fraction of these data sets is extremely challenging and will require major breakthroughs in analysis and interpretation. In particular, I believe the integration of epigenetic and expression data will be a major challenge over the next few years and there are many specific questions that are unresolved. There are two questions that I think are particularly interesting in the area of data integration. (i) How can we describe the epigenetic landscape and how is it related to development and disease? There are over 100 epigentic histone modifications known to date and more are being discovered all the time. Therefore, there are millions of possible epigenetic combinations that could be predictive of expression and function, and most probably only a small fraction of these are important, however. Recently there has been some excellent work published on combining epigenetic marks to annotate the genome [46,47]. (ii) How is alternative spicing controlled and what is the role of epigenetics? Next-generation sequencing has shown that many genes in the genome have multiple isoforms; however, the mechanisms that control the switching between alternative transcripts are not well understood. Recently there have been extremely fascinating observations that show an important role for epigenetics in controlling splicing events (for example, [48,49]). There is still a long way to go in order to try and integrate epigenetic and expression data on a genome-wide scale.

JR: (i) What properties of large non-coding RNA genes would identify subfamilies and classes? Imagine the text book had already been written for non-coding RNAs and someone recently discovered protein genes. One of the first things to do is identify functional domains (for example, helix-loop-helix, DNA-binding domains, and so on) that could be extrapolated to families related by functional properties. With RNA it's a bit trickier but initial progress is being made for large non-coding RNAs using co-expression with proteins, a process termed 'guilt by association'. We recently got a glimpse of some of the first emerging global properties after mapping and characterizing 8,000 long non-coding RNAs [50]. They are strikingly more tissue-specific than protein-coding genes, an interesting feature that we could potentially use in medical diagnostics. (ii) Why is there so much non-coding RNA? It's clear that there are numerous functional large non-coding RNAs but almost the entire genome is transcribed. Progress is being made by more global loss-of-function and gain-of-function experiments. (iii) What do these non-coding RNAs do and how do they do it? We have identified an emerging theme of non-coding RNA interacting with proteins and modulating their function. These RNA-protein complexes are important for maintaining cellular identity. We need to further understand the structural and functional elements that drive these interactions. If we could learn how these RNAs work, we could envision engineering them to guide stem cells into distinct cell types.

MW: (i) Single cell systems biology: many labs are making good progress toward this, from cell biological imaging screens, to measuring parameters in single cells. This will result in high resolution of biological information and will provide important insights into cell-to-cell variability in cellular networks. (ii) Dynamic networks: currently, many available networks collapse all measured interactions into a single graph. However, it is clear that only parts of the network are active in different cells or under particular conditions. Therefore, we need to start including spatiotemporal components of networks and their activity. Visualization is an important component of this, as is measuring reaction kinetics and the concentration of different biomolecules. (iii) Developing integrative networks that combine metabolism, protein-protein interactions, genetics and regulatory networks. So far, most studies focus on a single type of network. However, it is clear that many biological processes are controlled by a flow of information through different types of molecules, and thus networks, and often result in differences in cellular and organismal metabolism. To better capture the events important to a biological process, it will be important to combine all relevant, active networks into a single graph

If money were no object, what study would you love to perform?

DA: If money were no object, I'd sequence all the athletes in the Olympic village and all the dogs at Crufts!

BB: I would do much larger-scale genome-wide association studies to have sensitivities at the levels of small multifactorial effects (that is, leverage statistical power to be able to detect a signal in millions of human genomes that isn't currently distinguishable from noise). Assuming one can do the analysis efficiently with my lab's compressive genomics paradigm, whole genome sequences offer the ability to directly infer causal variants, as opposed to single-nucleotide polymorphisms, which currently just map marker regions (using linkage disequilibrium) to phenotypes. I would also like to do a genome-wide epigenome scan for discordant twins, to find out why they are different despite near identical genomes.

OH: I would study cancer genetics at a much higher resolution: at the single-cell level to study the process of clonal evolution and selection; at the environment or niche level to study the effect of the stroma and surrounding cellular environment on cancer progression; and finally, at the diploid level to determine the differential role of each allele in cancer progression. I'd start by addressing the technical challenges of single-cell genomics.

CH: There are two directions in which I'd like to go with such a grant, one computational and the other translational. The former is a bit prosaic, in that the field has now completed enough human microbiome projects to have a good idea of what data and metadata are useful. We need to develop a platform to standardize and automate the process of boiling microbial sequence down to functional information, then comparing that to subject phenotype, much as has been done successfully for environmental metagenomes up to this point. Translationally, this would enable a sustainable effort to track the human microbiome longitudinally, in a large prospective cohort. Epidemiology in such cohorts has been uniquely successful in characterizing dietary and environmental influences on health; the microbiome is a big component of our everyday environment that's so far not been assessed in such a manner. Building and maintaining such a cohort over time would take more than just my group and a few million dollars, but it's a project I'd be excited to see happen soon!

SL: I would examine the epigenetic status at regulatory sequences (especially enhancers) and see how this changes between conditions (for example, development, stimulation and drug treatment) and between individuals (for example, tumor tissues).

CM: I'd invest in tools for perturbing and characterizing mutants derived from a reference 'normal' human genome, so that we could enable true systems biology in humans. To do this, we'd need to improve technology for single and combinatorial knockdown of transcripts genome-wide, quantitative read-outs for various cell states (for example, transcription levels, protein levels, protein/chromatin modifications, protein-protein interactions) and single-cell quantitative phenotyping. In addition to improving the technology on all of these fronts, I would invest significant resources in applying them systematically and globally on a set of carefully chosen reference cell lines. I think the only way we will understand how the genotype determines phenotype - the basis of human disease - is through systematic functional genomics.

AO: As I'm interested in epigenetics and gene regulation, I would love to spend some money building up interesting epigenetic and expression data sets using ChIP-seq, RNA-seq and DNA-seq to look at histone modifications, transcription factors, RNA expression, microRNAs and RNA-protein interactions, all with a large number of controls. However, I would probably spend most money on employing and training people to develop methods to analyze and visualize the data in new and imaginative ways and to collaborate on exciting projects. There are so many different possibilities for exploring the vast amounts of genomic data that are now being produced. However, I believe the biggest bottleneck is the bioinformatics and the shortage of researchers in the field. There needs to be a big investment to address this shortage.

JR: I would study ants and bees. Social insects fascinate me, and I have a hunch that maternal and/or paternal RNA storage might be involved in establishing which larva becomes a worker or a queen. More practically, I would use the money to perform combinatorial genetic engineering of stem cells, aiming to guide and tweak their metamorphoses into other types of cells.

MW: If money (and time!) were no object, I would measure all metabolites in a system of interest (*Caenorhabditis elegans* in my case), and repeat this upon perturbation of gene regulatory networks by transcription factor RNA interference and under a variety of environmental and nutritional conditions. Simultaneously, I would measure the transcriptome under each condition. The data would then be used to create comprehensive and integrative network models that link metabolism to gene expression.

Published: 28 December 2011

References

1. Sodir NM, Swigart LB, Karnezis AN, Hanahan D, Evan GI, Soucek L: **Endogenous Myc maintains the tumor microenvironment.** *Genes Dev* 2011, **25:**907-1016.
2. Li Y, Zheng H, Luo R, Wu H, Zhu H, Li R, Cao H, Wu B, Huang S, Shao H, Ma H, Zhang F, Feng S, Zhang W, Du H, Tian G, Li J, Zhang X, Li S, Bolund L, Kristiansen K, de Smith AJ, Blakemore AI, Coin LJ, Yang H, Wang J, Wang J: **Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome *de novo* assembly.** *Nat Biotechnol* 2011, **29:**723-730.
3. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellåker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, McIntyre RE, Bhomra A, Nicod J, Gan X, Yuan W, van der Weyden L, Steward CA, Bala S, Stalker J, Mott R, *et al.*: **Mouse genomic variation and its effect on phenotypes and gene regulation.** *Nature* 2011, **477:**289-294.
4. Khatib F, DiMaio F; Foldit Contenders Group; Foldit Void Crushers Group, Cooper S, Kazmierczyk M, Gilski M, Krzywda S, Zabranska H, Pichova I, Thompson J, Popović Z, Jaskolski M, Baker D: **Crystal structure of a monomeric retroviral protease solved by protein folding game players.** *Nat Struct Mol Biol* 2011, **18:**1175-1177.
5. BM Good, AI Su: **Games with a scientific purpose.** *Genome Biol* 2011, **12:**135.
6. **Wellcome Trust Sanger Institute - Dr Peter Campbell** [http://www.sanger.ac.uk/research/faculty/pcampbell/]
7. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW, Quail MA, Burton J, Swerdlow H, Carter NP, Morsberger LA, Iacobuzio-Donahue C, Follows GA, Green AR, Flanagan AM, Stratton MR, *et al.*: **Massive genomic rearrangement acquired in a single catastrophic event during cancer development.** *Cell* 2011, **144:**27-40.
8. **UC San Diego - Inder M Verma** [http://biology.ucsd.edu/faculty/verma.html]
9. **Salk Institute - Fred H Gage** [http://www.salk.edu/faculty/gage.html]
10. Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH, Verma IM: ***BRCA1* tumour suppression occurs via heterochromatin-mediated silencing.** *Nature* 2011, **477:**179-184.
11. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ: **Vaginal microbiome of reproductive-age women.** *Proc Natl Acad Sci U S A* 2011, **108(Suppl 1):**4680-4687.
12. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA: **Mediator and cohesin connect gene expression and chromatin architecture.** *Nature* 2010, **467:**430-5.
13. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY: **Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions.** *Mol Cell* 2011, **44:**667-678.
14. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA, Van Den Berg D, Laird PW: **Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.** *Nat Genet* 2011 [Epub ahead of print].
15. McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM: **Systematic discovery of nonobvious human disease models through orthologous phenotypes.** *Proc Natl Acad Sci U S A* 2010, **107:**6544-6549.
16. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B: **Synthetic spike-in standards for RNA-seq experiments.** *Genome Res* 2011, **21:**1543-1551.
17. Heo JB, Sung S: **Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA.** *Science* 2011, **331:**76-79.
18. Angel A, Song J, Dean C, Howard M: **A polycomb-based switch underlying quantitative epigenetic memory.** *Nature* 2011, **476:**105-108.
19. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, van Leenen D, Groot Koerkamp MJ, Ko CW, Miles AJ, Brabers N, Brok MO, Lenstra TL, Fiedler D, Fokkens L, Aldecoa R, Apweiler E, Taliadouros V, Sameith K, van de Pasch LA, van Hooff SR, Bakker LV, Krogan NJ, Snel B, Holstege FC: **Functional overlap and regulatory links shape genetic interactions between signaling pathways.** *Cell* 2010, **143:**991-1004.
20. Green RA, Kao HL, Audhya A, Arur S, Mayers JR, Fridolfsson HN, Schulman M, Schloissnig S, Niessen S, Laband K, Wang S, Starr DA, Hyman AA, Schedl T, Desai A, Piano F, Gunsalus KC, Oegema K. **A high-resolution *C. elegans* essential gene network based on phenotypic profiling of a complex tissue.** *Cell* 2011, **145:**470-482.
21. **The Netherlands Cancer Institute - Jonkers, dr. J.M.M. (Jos)** [http://www.nki.nl/Research/Faculty+and+Research/Divisions/Molecular+Biology/Jonkers.htm]
22. **Oxford Neuroscience - Jonathan Flint** [http://www.neuroscience.ox.ac.uk/directory/jonathan-flint]
23. **MIT Mathematics - Daniel Kleitman** [http://www-math.mit.edu/people/profile?pid=135]
24. **Jonathan A King** [http://web.mit.edu/king-lab/www/people/JKing/JKing.html]
25. **Merck: Executive Committee** [http://www.merck.com/about/leadership/executive-committee/home.html]
26. **UC SanDiego - people** [http://frazer.ucsd.edu/people.html]
27. **Princeton University - Olga Troyanskaya** [http://reducio.princeton.edu/cm/ogt]
28. **Broad Institute - Dirk Gevers** [http://www.broadinstitute.org/~dgevers/]
29. **Myles Brown** [http://research4.dfci.harvard.edu/brownlab]
30. **The University of Melbourne - Prof Rachel Webster** [http://www.findanexpert.unimelb.edu.au/researcher/person14490.html]
31. **The Walter and Eliza Hall Institute of Medical Research - Professor Gordon Smyth** [http://www.wehi.edu.au/faculty_members/professor_gordon_smyth]
32. **Michael Snyder** [http://snyderlab.stanford.edu/]
33. **The Chang Lab** [http://changlab.stanford.edu/]
34. **MIT Department of Biology - Eric S Lander** [http://www.mit.edu/~biology/facultyareas/facresearch/lander.html]
35. **CCSB** [http://ccsb.dfci.harvard.edu/web/www/ccsb/]
36. **UC San Diego Bio-Computational Center (BCC)** [http://ctri.ucsd.edu/Informatics/Pages/Genome-Study.aspx]
37. O'Geen H, Nicolet CM, Blahnik K, Green R, Farnham PJ: **Comparison of sample preparation methods for ChIP-chip assays.** *Biotechniques* 2006, **41:**577-580.
38. Adli M, Zhu J, Bernstein BE: **Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors.** *Nat Methods* 2010, **7:**615-618.
39. Shankaranarayanan P, Mendoza-Parra MA, Walia M, Wang L, Li N, Trindade LM, Gronemeyer H: **Single-tube linear DNA amplification (LinDA) for robust ChIP-seq.** *Nat Methods* 2011, **8:**565-567.
40. Check Hayden E: **Evidence of altered RNA stirs debate** [http://www.nature.com/news/2011/110525/full/473432a.html]
41. Li M, Wang IX, Li Y, Bruzel A, Richards AL, Toung JM, Cheung VG: **Widespread RNA and DNA sequence differences in the human transcriptome.** *Science* 2011, **333:**53-58.
42. Bahn JH, Lee JH, Li G, Greer C, Peng G, Xiao X: **Accurate identification of A-to-I RNA editing in human by transcriptome sequencing.** *Genome Res* 2011 [Epub ahead of print].
43. Pickrell J: **Notes on the evidence for extensive RNA editing in humans** [http://www.genomesunzipped.org/2011/05/notes-on-the-evidence-for-extensive-rna-editing-in-humans.php]
44. Schrider DR, Gout J-F, Hahn MW: **Very few RNA and DNA sequence differences in the human transcriptome.** *PLoS One* 2011, **6:**e25842.
45. The ENCODE Project Consortium: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9:**e1001046. [http://www.genome.gov/Pages/Research/ENCODE/ENCODE_UsersGuide.pdf]
46. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE: **Mapping and analysis of chromatin state dynamics in nine human cell types.** *Nature* 2011, **473:**43-49.
47. Filion GJ, van Bemmel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, van Steensel B: **Systematic protein location mapping reveals five principal chromatin types in Drosophila cells.** *Cell* 2010, **143:**212-124.
48. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T: **Regulation of alternative splicing by histone modifications.** *Science* 2010, **327:**996-1000.
49. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S: **CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.** *Nature* 2011, **479:**74-79.
50. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL: **Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses.** *Genes Dev* 2011, **25:**1915-1927.

doi:10.1186/gb-2011-12-12-137
Cite this article as: Genomics in 2011: challenges and opportunities. *Genome Biology* 2011, **12**:137.

About the contributors

David J Adams (DA) is a faculty member at the Wellcome Trust Sanger Institute. He performs forward genetic screens to uncover cancer genes and cancer pathways and is also leading a program to use new-technology sequencing to decode the genomes of several mouse strains that form the basis of mouse experimental genetics.

Bonnie Berger (BB) is Professor of Applied Math and Computer Science at Massachusetts Institute of Technology. Her recent work focuses on designing algorithms to gain biological insights from advances in automated data collection and the subsequent large data sets drawn from them. She works on a diverse set of problems, including network inference, protein folding, comparative genomics and medical genomics.

Olivier Harismendy (OH) is an Assistant Professor of Pediatrics at the University of California, San Diego and Member of the Moores UCSD Cancer Center; he works on several projects, including investigating the function of regulatory DNA variants in cardiovascular diseases and cancer, the identification of DNA markers of cancer progression and the development and implementation of genomic assays and analysis in the clinic to guide personalized cancer care.

Curtis Huttenhower (CH) is Assistant Professor of Computational Biology and Bioinformatics at Harvard University. His research is concerned with the discovery of useful biological knowledge in large collections of genomic data. This requires the development of computational methodology that is efficient enough to deal with billions of data points while remaining biologically rich enough to capture the complexities of molecular biology.

X Shirley Liu (SL) is an Associate Professor at the Department of Biostatistics and Computational Biology at the Dana-Farber Cancer Institute and Harvard School of Public Health. A computational biologist, with expertise on the integrative modeling of transcription and epigenetic regulation, she and her colleagues have developed a number of widely used algorithms for transcription factor motif discovery.

Chad L Myers (CM) is a Principal Investigator at the University of Minnesota. The research in his laboratory focuses on machine-learning approaches for integrating diverse genomic data to make inferences about biological networks, the main purpose of which is to further understanding of gene function and how genes or proteins interact to carry out cellular processes.

Alicia Oshlack (AO) is the head of the bioinformatics research group at the Murdoch Childrens Research Institute. She is an expert in developing analysis methods for high-throughput genetic technologies and works on many collaborative projects throughout the institute.

John L Rinn (JR) is an Assistant Professor of Stem Cell and Regenerative Biology at Harvard University and Medical School, and Senior Associate Member of the Broad Institute. His research aims to understand the role of long intergenic non-coding RNAs in establishing the distinct epigenetic states of adult and embryonic cells and their misregulation in diseases such as cancer.

Albertha J M Walhout (MW) is Professor at the University of Massachusetts. Her laboratory uses a variety of experimental and computational systems biology approaches to map and characterize gene regulatory networks and to understand how regulatory circuitry controls animal development, function and homeostasis. Ultimately, her aim is to understand how dysfunctional networks affect or cause diseases such as diabetes, obesity and cancer.