FEBS Letters 583 (2009) 1668–1673

journal homepage: www.FEBSLetters.org

Minireview

Next-generation sequencing approaches in genetic rodent model systems to study functional effects of human genetic variation

Victor Guryev, Edwin Cuppen *

Hubrecht Institute and University Medical Center Utrecht, Cancer Genomics Center, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 18 February 2009
Revised 8 April 2009
Accepted 14 April 2009
Available online 18 April 2009

Edited by Miguel De la Rosa

Keywords:
Next-generation sequencing
Rodent models
Functional genomics
Genotype–phenotype relationship

ABSTRACT

Rapid advances in DNA sequencing improve existing techniques and enable new approaches in genetics and functional genomics, bringing about unprecedented coverage, resolution and sensitivity. Enhanced toolsets can facilitate the untangling of connections between genomic variation, environmental factors and phenotypic effects, providing novel opportunities, but may also pose challenges in data interpretation, especially in highly heterogeneous human populations. Laboratory rodent strains, however, offer a variety of tailored model systems with controlled genetic backgrounds, facilitating complex genotype/phenotype relationship studies. In this review we discuss the advent of massively parallel sequencing, its methodological advantage for molecular analysis in model organisms and the expectation of increased understanding of biologically relevant consequences of human genetic variation.

1. Introduction

With the advent of massively parallel sequencing technologies [1,2], we are approaching the possibility of resequencing a human genome for only $1000. As this price tag is comparable to that of advanced clinical tests, genome sequencing may become a common diagnostic procedure. As a consequence, personal genomes will mark a new era in biomedicine as medical records of patients (or even those of healthy individuals) will expand to incorporate data about millions of DNA variants. Still, our knowledge of how these variants account for disease or disease susceptibility or determine the success of different treatment strategies is fragmentary.

A common way of obtaining the information linking genotypes to specific phenotypes in humans is to perform genome-wide association studies (GWAS) in which whole-genome genetic information, typically more than half a million polymorphisms, is collected for thousands of patients and control cases, with the goal of finding more than randomly expected correlations between a genomic interval and the disease parameter being studied. Currently, over 150 such relationships have been found and their number is growing rapidly [3]. Although many monogenetic disorders have been successfully mapped and their causal variants identified, a more typical GWAS analysis for a complex trait results in the discovery of a number of genomic loci with modest effects that collectively explain only a small part of the observed phenotypic variation. Clearly, the detection of the complete genetic component of a complex trait is hampered by the many rare genetic variants that are present in the rapidly growing human population [4,5]. Although some of these variants may have deleterious effects with respect to certain phenotypes, the majority are expected to be non-functional. Finally, multiple variants may interact with each other in an additive, neutralizing or complex manner. It is thus a major challenge in human genetics to identify and functionally characterize relevant genetic variants and their contributions to phenotypes.

Given the limits of clinical research in humans, cellular and animal models can greatly aid in deciphering the mechanistic effects of genetic variants. To this end, murine models of human diseases such as mouse and rat laboratory strains represent convenient tools for the investigation of the functional effects of genetic variation. These models are characterized by reduced genetic complexity in a variety of experimental setups. In addition, access to genetic and physiological manipulations makes them an ideal platform for functional genomics research to study disease susceptibility and etiology, as well as testing of intervention strategies.

While there are extensive discussions on the advance of sequencing technologies [6] and how they may revolutionize studies in humans [7], we want to focus this review on the benefits of next-generation sequencing technologies in mammalian model systems for functional genomics research.

* Corresponding author.
E-mail address: e.cuppen@niob.knaw.nl (E. Cuppen).

0014-5793/$36.00
doi:10.1016/j.febslet.2009.04.020

1.1. The power of mammalian model systems

The translation of a DNA variant into phenotypic differences between individuals is one of the central questions of the post-genome era. Several obstacles impede the progress of understanding genotype–phenotype links in humans. First, the extensive genetic heterogeneity of and heterozygosity in the human population, which is intrinsic to its outbred nature, makes the investigation of the effect(s) of a single isolated DNA variant extremely difficult. Given that the contribution of a single variant to a phenotype can be rather small, these effects are likely to be missed in heterogeneous samples. Furthermore, invasive experiments are not common in human subjects and bioethical aspects often prevent the disclosure of all of the patients' data, which could be relevant for statistical interpretation and classification. Finally, the phenotypic data (both of patients and control subjects) is often incomplete and should be updated constantly and extended as people acquire diseases with age.

Experimental genetic model systems such as laboratory mice and rats may greatly facilitate the uncovering of general principles that guide the translation of DNA variants into phenotypic differences seen among individuals [8]. For over 100 years, rodent models have provided a renewable source of laboratory animals that model a wide variety of common human diseases [9]. They serve as indispensable tools for genetic, physiological, behavioral, and toxicological studies. The key advantages of these model systems are:

(i) The possibility to perform experiments in a controlled environment using individuals with a known genetic background. Inbred individuals exhibit homozygosity at every locus, facilitating studies of recessive alleles. In addition, animals can be reared and studied under controlled laboratory conditions.
(ii) *Reproducibility of experiments*. Since individual animals from the same inbred line are genetically identical, experimental results can be reproduced in completely independent experiments, improving experiments in terms of both robustness of results and sensitivity for discovery.
(iii) *Cumulative nature of phenotype and genotype data*. In the case of model systems, any tissue is accessible and can be collected at any developmental stage. In turn, the obtained genotype and phenotype readouts should be invariant, given the inbred nature of the strain.
(iv) *Genome manipulation*. Gene knock-out and knock-in technologies enable disrupting or altering gene structure and expression in a targeted way. It is even possible to 'humanize' a model by introducing genetic variants that are similar to the genetic lesions observed in human patients.
(v) *Flexible modeling of human diseases and intervention strategies*. Since specific strains are selected for their susceptibility to certain common diseases, these models can not only be used for increasing our understanding of disease etiology and biology, but are also well-suited for systematic testing of physiological or pharmacological interventions.

1.2. The variety of murine models

There has been a historical split in murine models: the laboratory rat has been a primary model for physiological and pharmacological studies while the laboratory mouse has served as the most popular mammalian genetic model, especially due to the availability of targeted gene knockout technology. However, thanks to technological developments, experimental limitations in the rat are rapidly disappearing. Current genetic model systems in mice and rats include a variety of experimental platforms tailored to the identification of genomic loci and genetic variants that contribute to the disease of interest or influence disease susceptibility. Examples are classical inbred strains, recombinant inbred strains, outbred strains, heterogeneous stocks, consomic and congenic strains, and gene-knockout models (Fig. 1). These strains include a wide variety of physiological and phenotypic parameters, either collected in small-scale dedicated experiments or large-scale systematic characterization studies, presenting a range of possibilities for follow-up experiments.

Much effort has been put into the development and characterization of these genetic models. A good example of a popular system in rat genetics is the HXB/BXH recombinant inbred (RI) panel that represents a thoroughly characterized set of 30 rat strains that resulted from the inbreeding of F2 animals from a BN-Lx × SHR cross (reviewed in [10]). The development of this panel commenced in 1982 and currently represents the largest set of recombinant inbred rat strains. The RI panel as a model platform simplifies association and linkage analysis while providing direct access to any of the individual strains for experimental follow-up. Multiple animals of each of the RI strains have been characterized for a wide variety of phenotypic parameters, including cardiovascular, metabolic, behavioral, developmental and toxicological traits, as well as molecular markers such as gene-expression levels. Currently, there are hundreds of phenotypic (pQTLs) and thousands of expression quantitative trait loci (eQTLs) described for this panel.

Another type of model platform that can effectively mimic human heterozygosity is the heterogeneous stock (HS, Fig. 1) [11,12]. Heterogeneous stocks represent a defined pool of genetic variation derived from a limited number of inbred progenitor strains that are outbred via pseudorandom breeding. As a result, heterogeneous stocks provide diverse combination of alleles with defined genetic background and exploit historic recombinations that have accumulated in a genetically heterogeneous population. Quantitative trait loci (QTLs) can be mapped using HS strains, resulting in a high resolution with intervals of less than 1 cM. With eight inbred strains that typically contribute to an HS panel, one can expect to find many more QTLs than in an RI panel. The success of QTL mapping in HS mice [11,13] has inspired the development of similar resources for other model organisms [12].

QTL mapping approaches in rodents have shown that for complex human diseases such as hypertension, multiple QTLs can be found throughout the genome. To dissect the contribution of a single locus, multiple consomic and congenic strains have been constructed on the same genetic backgrounds as the RI or HS panels (reviewed in [14]). These designer strains can be used for further crosses to narrow down the QTL region and to help uncover gene–gene interactions in complex diseases. Interestingly, QTL regions identified in rodent studies often overlap with syntenic regions in the human genomes that were identified in relevant human QTL and GWAS studies.

2. Prospects for next-generation sequencing

Recent advances in parallelization of DNA sequencing technology and extensive increases in throughput are dramatically reducing the price and speed of sequencing. Sequencing of a mammalian genome does not cost millions anymore and it is expected that costs will drop below $5000 per genome by the end of 2009. Where consortium efforts were required before, a single laboratory can now afford to sequence their organism or strain of interest. The advantage of cheap sequencing does not end here; many other existing technologies are expected to adopt massively parallel sequencing to attain better performance, resolution and specificity. In this review, we aim to highlight the benefits of next generation

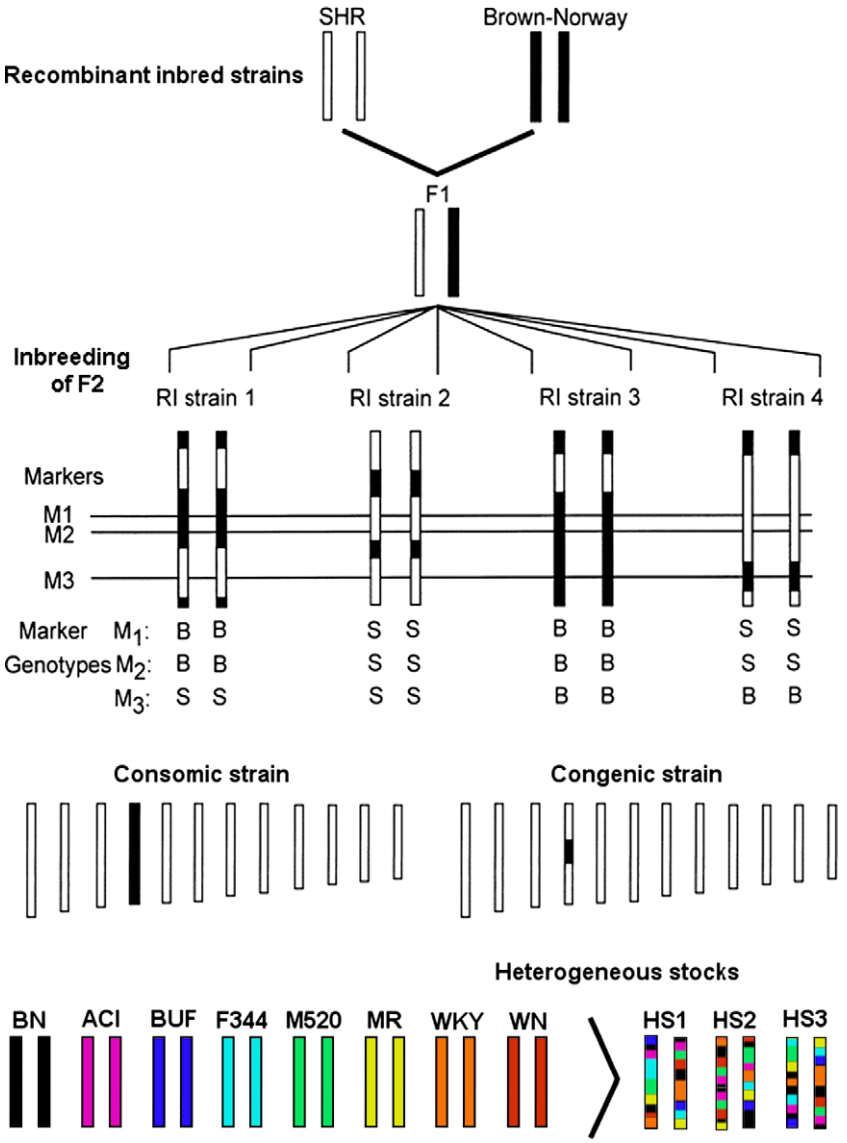

Fig. 1. Various rodent model systems. The top panel illustrates the development of the HXB/BXH recombinant inbred panel in rats. The inbred strains BN-Lx and SHR were crossed and hybrid F1 animals were intercrossed to generate mosaic F2 animals. Individual RI lines were derived by backcrosses for over 80 generations. The middle panel outlines consomic and congenic strains where a chromosome or a single locus from one strain was replaced by the same segment taken from another strain. The bottom panel explains the composition of heterogeneous stocks using the rat HS panel as an example.

sequencing (NGS), keeping in mind specific aspects of mammalian model systems.

2.1. Model genomes: more genomes with less effort

The availability of multiple genomes can greatly advance the field of comparative and functional genomics. A good example is the effort to generate genome assemblies of twelve different *Drosophila* species [15] that concluded in 2007 and inspired many follow-up studies. At present, with several high-throughput sequencing runs it is already possible to perform whole-genome resequencing and polymorphism discovery for model organisms with relatively small genomes like *Caenorhabditis elegans* [16]. As the instruments improve, the read coverage obtained in a single run will be sufficient to decipher complete mammalian genomes.

Complete resequencing of model organisms has a better potential for early adoption than large scale resequencing of human genomes as it can deliver more information with less money and effort. Since the breeding history of rodent genetic model systems is known, it is sufficient to resequence only a handful of parental strains to capture all variation in the complete model system. Because of the inbred nature of parental strains, haplotypes should not interfere with the phasing or genome assembly process. Once parental genomes are well-characterized, relatively simple and inexpensive genotyping can be used to determine the genetic composition of all the derived strains.

Another advantage of using model systems is that sequencing of several strains will result in a complete inventory of all sequence variants. This is in contrast to human studies where, because of rare alleles, obtaining all sequence variants requires full resequencing of all samples. NGS approaches provide the opportunity to rapidly characterize the complete 'variome' in mammalian model systems, thereby ensuring that no variants with potential phenotypic effects are left behind.

2.2. DNA variation discovery and contribution to genetic studies

But what would these additional complete genomes mean for model organisms' genetics? First of all, they will provide a detailed, high-resolution genetic map for forward genetics approaches. With

the current marker spacing of over 100 kb in laboratory rats [17], mutation mapping is, to a great degree, limited by the density of known markers. Even mild resequencing of different strains can significantly improve their genetic maps and increase the precision of QTL or mutation mapping. The new sequencing technologies may enhance the mutation mapping procedure in yet another way: regions of interest can be captured on microarray slides and resequenced to perform fine mapping or mutation discovery [18–20].

Further, with even more sequencing throughput, it will be possible to resequence a full genome of a specific strain to identify the causal variant or variants. Although identification of a 'causal' mutation solely by sequencing is not feasible yet [21], further improvements in the accuracy and length of sequencing reads are expected to improve this situation, although some degree of outcrossing or mapping will remain needed.

Beyond sequencing different strains or isolates of the same organism, closely related taxa can be sampled to study phylogenomics, or the evolutionary history of model organisms. Additional genomes will not only highlight recent selection events, but should also facilitate better alignments, gene predictions and identification and analysis of functionally important noncoding regions such as promoters [22].

2.3. Structural genome variation

Another class of genomic variation that has only recently been appreciated is structural variation [23]. Change in copy-number of genomic segments is implicated in many human genetic disorders and is estimated to account for about 20% of gene expression differences between individuals [24]. In contrast to variation due to copy-number, current genome-wide methods of choice, such as array-based comparative genome hybridization, cannot cope with detection of copy-neutral structural variants such as inversions and translocations, and cannot distinguish tandem from non-tandem duplications. However, these types of variants can effectively be discovered and detected using paired-end mapping (PEM) technology, which is available for all three commonly used next-generation sequencing platforms present on the market. With paired-end mapping facilitated by next-generation sequencing, it becomes in principle possible to characterize copy-number and copy-neutral structural variations at the base-pair level [25]. Nevertheless, the combination of PEM and NGS requires sophisticated data analysis for interpretation of genome structure alterations. As the effect of structural genome variation on phenotypic diversity remains largely unexplored, genetic model systems may play an important role in untangling these relationships in a systematic way.

2.4. Assembling genomes de novo

Affordable sequencing will not only change our approach to model systems, it will also dramatically change the model systems as we know them. Many organisms have unique features, but lack the genomic characterization that stems from advances such as genome assembly. Last year, the feasibility of *de novo* genome assembly using short reads was demonstrated for bacterial genomes [26,27] and we can expect that *de novo* assemblies of more complex genomes will follow. While good reference genomes are available for both mouse and rat, resequencing additional strains will require some degree of *de novo* assembly as well, as genomic segments that are lacking in a reference genome are not automatically included in *de novo* assemblies based on short read-based shotgun resequencing. Reads that cannot be mapped to the reference genome could represent strain-specific genomic sequence. *De novo* assembly of mammalian genome sequences using short reads remains bioinformatically challenging. In addition, such efforts are further complicated by the fact that the pool of unmappable reads includes poor quality reads. However, with increased read lengths and quality due to technological improvements, these complications are expected to disappear.

Finally, next-generation sequencing is also expected to contribute to the functional annotation of the available or newly assembled genomes. Resequencing of the transcribed component of a genome, as discussed below, will improve our understanding of functionally relevant regions in genomes.

2.5. Complete transcriptomes

Undoubtedly, one of the most anticipated applications of NGS is high-throughput sequencing of RNA samples, known as RNA-Seq. RNA-Seq has been shown to be an exhaustive and reproducible method of mRNA expression profiling [28–30]. Direct and truly quantitative, it detects not only unknown transcripts, but also transcript isoforms, providing unprecedented experimental evidence for the expression of predicted genes and non-coding transcripts as well as previously unknown splice forms of known transcripts. The complexity, dynamics and sequence context of transcriptomes can now in principle be studied in a near-complete fashion in a single experimental approach.

Early studies in this area have revealed many unexpected transcript structures, and it is evident that, even in carefully designed microarray-based studies, a substantial fraction of RNA molecules have escaped detection. RNA-Seq effectively combines the discovery of transcript structures and profiling of their expression and can be employed for the *ab initio* construction of complex mammalian transcriptomes [31].

RNA-Seq opens a wide range of scientific questions that can be explored in model systems. First, this complete and robust expression profiling can be done for multiple tissues and at different (including preclinical) stages of disease progression. Secondly, due to the complexity of the regulation network, the effect of genetic variation in a single locus can be addressed in a simplified system such as a congenic strain. Furthermore, it is of interest to investigate what major factors are driving differential gene expression in mammals. To reveal the genetic components in mRNA expression, a survey of allele-specific expression of mammalian transcriptomes can be employed. By looking at allelic imbalance in an F1 intercross of two genetically dissimilar strains, cis-regulation of genes can be studied [32]. Scaling this method to use massively parallel sequencing in combination with model system such as heterogeneous stocks may become a unique and powerful approach. While genome-wide application of this method requires at least one polymorphism in every transcript, a recent study showed that intronic SNPs might be informative for assessing allele imbalance [33].

Finally, the potential to manipulate these genetic model systems can be used to explore the influence of tissue-, sex-, age-, and environment-specific factors on transcription, and holds the key to the comprehensive understanding of expression regulation under variable conditions.

The comparative analysis of genotypic data and gene expression results can be used for the discovery of eQTLs that link expression changes to genomic segments with correlated strain distribution patterns (SDPs). Cis-eQTLs, genomic segments localized to the transcripts, are thought to represent promoter regulatory variants or alleles that affect transcript processing or stability. At the same time, far more trans-eQTLs are observed, which typically have a smaller effect on transcript expression. Expression profiling using microarrays (e.g. Affymetrix RAE230 for BXH/HXB rats) has already revealed hundreds of robust links between DNA polymorphisms and expression changes [34]. Another challenge for eQTL studies

is an increased rate of false-positives, due to a variety of different SDPs throughout the genome, some of which may correlate with expression when the number of strains is limited [34]. RNA sequencing is likely to boost the number and specificity of such discoveries as increased sensitivity improves the detection of low-level expression, enabling better monitoring of more subtle expression changes.

While traditional eQTL studies compare the quantitative effect of genetic variation on gene expression, RNA-Seq provides means to perform these studies in a qualitative manner. In other words, it enables the discovery of genetic variants that affect alternative transcript initiation, splicing and termination, so called splice QTLs (sQTLs) [35]. The correlation between an expressed splice isoform and its genomic locus in RI and HS panels can reveal polymorphisms critical for splicing: cis-sQTLs and polymorphic genes that encode splice-factors, trans-sQTLs.

2.6. Mapping the epigenome and protein–DNA interactions

Massive sequencing is rapidly emerging as a powerful tool in sequence census methods [36]. In recent years, extensive chromatin modification and protein binding data have been generated by chromatin immunoprecipitation coupled with microarray hybridization (ChIP-chip). The use of sequencing as a substitute for microarrays (ChIP-Seq) has improved the method in several ways [37–40]. First, direct sequencing is data 'agnostic' and can address a larger part of the genome (though not the entire genome) in virtual absence of a cross-hybridization background bias. Secondly, it can achieve a level of base-pair resolution that allows more precise mapping of protein binding or DNA/chromatin modification sites. A recent study was able to determine the binding sites of the NRSF (neuron-restrictive silencer factor) and GABP (growth associated binding protein) transcription factors with a resolution of 20 bp and a mean distance between peak call and putative motif of just several base pairs [37]. The unprecedented sensitivity of the method also enables researchers to reveal binding sites of cofactor proteins. Another advantage of the new method is digital profiling of binding/modification events. Since this method counts the number of reads, it is not biased by the hybridization dynamics of microarray-based experiments. ChIP-Seq does not have the problem of continuous hybridization signal (and saturation) and thus has a greater dynamic resolution compared to ChIP-chip (Fig. 2). Furthermore, the direct reading of DNA expands regional information to include allele-specific coverage, enabling the study of the effect of genetic variation on factor binding or epigenetic modification [37]. Finally, since in most cases a given factor of interest occupies only small parts of the genome, sequencing saturation is already relatively rapidly achieved, making ChIP-Seq a financially attractive alternative to ChIP-chip.

These technological benefits of sequencing analysis increase the diversity of substrates that can be detected to include transcription factors, nucleosome cores, DNA methylation and histone modifications. A recent comprehensive analysis of 20 different histone modifications has revealed their genome-wide distribution pattern [38]. These epigenetic markers differentially mark active promoters, transcription start sites, and gene bodies, information that is highly complementary to RNA-Seq results. By employing this method, it has become possible to discern epigenetic changes that distinguish pluripotent from lineage-committed cells [39].

ChIP-Seq is highly complementary to expression profiling experiments, and is a superior method for untangling the mechanistic changes that lead to differential gene expression, for example when normal and diseased tissues are compared. However, the testing of interventions that can reverse epigenetic and regulatory changes back to a healthy state (e.g. using inducible systems) is limited to model animal or cell systems. In these systems, the transition between different regulatory states can be monitored in any tissue type and followed throughout all developmental stages. Although we are aware of the interplay between both genetic and environmental factors affecting the epigenetic state of chromatin, the use of inbred animals, where the genetic component can be fixed, can provide the proper tools to study the influence of isolated environmental factors such as diet, age and disease progression on gene switching.

2.7. Spatial organization of the genome

It is known that the organization of chromatin in the nucleus is highly nonrandom and that the relative location of regulatory modules is critical for gene expression. Several techniques have been developed to discover the proximity of DNA segments at both the local and genome-wide levels, reviewed in [40]. Although, to our knowledge, no studies have been published using chromosome conformation capture with massive sequencing, the analysis of genome organization can hugely benefit from massive-scale sequencing. This combination of technologies may ultimately develop into a method that reconstructs the complete spatial organization of chromosomes in the nucleus [41]. Again, we envision that model systems will provide major contributions, from allowing simplified initial experiments to enabling more flexible and reproducible studies of conformational changes in disease models.

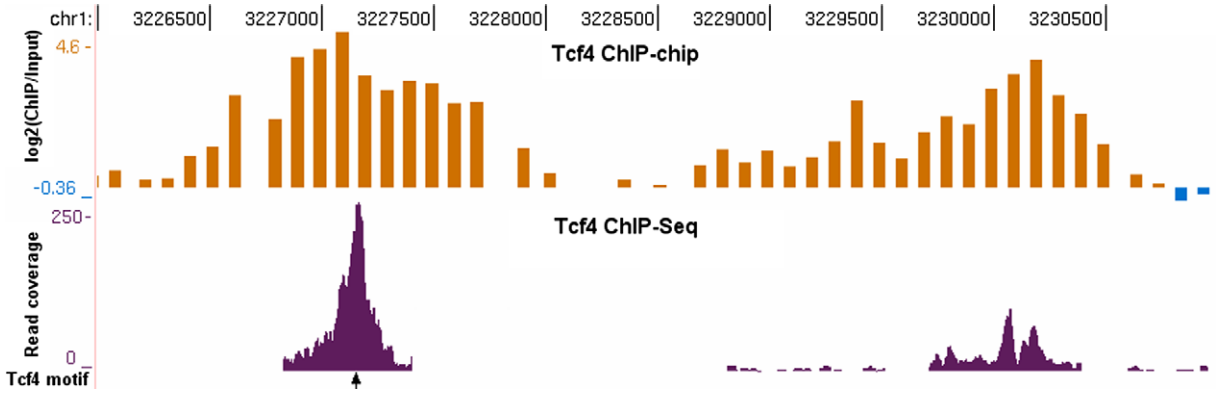

Fig. 2. Comparison of ChIP-chip and Chip-Seq sensitivity and resolution. The panels show a comparison of technologies used for Tcf4 transcription factor binding. The logarithmic ratio (base 2) for the hybridization intensity ratio between ChIP and input DNA for the ChIP-chip experiment (top) and read coverage for the ChIP-Seq experiment (bottom) are shown. The arrow below the ChIP-Seq panel shows the position of the Tcf4 binding site (WTCAAAG). The ChIP-Seq method, supplemented by massively parallel SOLiD sequencing, shows sharper and more structured peaks with greater dynamic range and little background signal.

3. Combined advantage of NGS for model organism functional studies and understanding of human genomic variation

This review has not described all of the potential utilizations of ultra-high-throughput sequencing; many potential utilizations are currently under development, including massive bisulfite sequencing, mapping of DNase hypersensitive sites, and more. It is apparent, though, that affordable sequencing is rapidly changing our experimental approaches and improving the completeness, resolution and sensitivity of their results.

Although the primary goal in biomedical research is to understand human biology and disease etiology and susceptibility, one should also realize that initially applying these large scale technologies to the genetically complex human system may generate a puzzle that consists of so many pieces and potential solutions that current statistical analysis approaches do not suffice to find the correct and causal relation between a genetic event (or more likely several events) and a phenotypic consequence. Applying the next-generation sequencing-based tools mentioned above in a systematic and comprehensive way in a controlled genetic system, such as a rodent RI or HS model system, enables the development of both statistical linkage and association models at the nucleotide level and dynamic computational systems biological approaches to describe gene-regulatory models and their genetic effects. Integration of these models with human linkage data, GWAS results, and, in the near future, personal genomics information is expected to provide a powerful approach towards a better understanding of clinically relevant genotype–phenotype relationships. Ultimately, this approach is also expected to provide novel strategies for pharmacological interference.

As biology is often found to be much more complex than anticipated, devoting sufficient next-generation sequencing resources to model organism applications is likely to pay off quickly by increasing our understanding of the data that soon will be flooding in from thousands of human genome sequencing projects.

References

[1] Margulies, M. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.
[2] Bentley, D.R. (2006) Whole-genome re-sequencing. Curr. Opin. Genet. Dev. 16, 545–552.
[3] Altshuler, D., Daly, M.J. and Lander, E.S. (2008) Genetic mapping in human disease. Science 322, 881–888.
[4] Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. 69, 124–137.
[5] Kryukov, G.V., Pennacchio, L.A. and Sunyaev, S.R. (2007) Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am. J. Hum. Genet. 80, 727–739.
[6] Mardis, E.R. (2008) Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 9, 387–402.
[7] ten Bosch, J.R. and Grody, W.W. (2008) Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. J. Mol. Diagn. 10, 484–492.
[8] Aitman, T.J. et al. (2008) Progress and prospects in rat genetics: a community view. Nat. Genet. 40, 516–522.
[9] Moreno, C., Lazar, J., Jacob, H.J. and Kwitek, A.E. (2008) Comparative genomics for detecting human disease genes. Adv. Genet. 60, 655–697.
[10] Printz, M.P., Jirout, M., Jaworski, R., Alemayehu, A. and Kren, V. (2003) Genetic models in applied physiology. HXB/BXH rat recombinant inbred strain platform: a newly enhanced tool for cardiovascular, behavioral, and developmental genetics and genomics. J. Appl. Physiol. 94, 2510–2522.
[11] Valdar, W. et al. (2006) Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat. Genet. 38, 879–887.
[12] Johannesson, M. et al. (2009) A resource for the simultaneous high-resolution mapping of multiple quantitative trait loci in rats: the NIH heterogeneous stock. Genome Res. 19, 150–158.
[13] Mott, R., Talbot, C.J., Turri, M.G., Collins, A.C. and Flint, J. (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. Proc. Natl. Acad. Sci. USA 97, 12649–12654.
[14] Kwitek-Black, A.E. and Jacob, H.J. (2001) The use of designer rats in the genetic dissection of hypertension. Curr. Hypertens. Rep. 3, 12–18.
[15] Clark, A.G. et al. (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203–218.
[16] Hillier, L.W. et al. (2008) Whole-genome sequencing and variant discovery in *C. elegans*. Nat. Meth. 5, 183–188.
[17] Saar, K. et al. (2008) SNP and haplotype mapping for genetic analysis in the rat. Nat. Genet. 40, 560–566.
[18] Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J. and Zwick, M.E. (2007) Microarray-based genomic selection for high-throughput resequencing. Nat. Meth. 4, 907–909.
[19] Albert, T.J. et al. (2007) Direct selection of human genomic loci by microarray hybridization. Nat. Meth. 4, 903–905.
[20] Porreca, G.J. et al. (2007) Multiplex amplification of large sets of human exons. Nat. Meth. 4, 931–936.
[21] Shen, Y., Sarin, S., Liu, Y., Hobert, O. and Pe'er, I. (2008) Comparing platforms for *C. elegans* mutant identification using high-throughput whole-genome sequencing. PLoS ONE 3, e4012.
[22] Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K.D., Ovcharenko, I., Pachter, L. and Rubin, E.M. (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science 299, 1391–1394.
[23] Redon, R. et al. (2006) Global variation in copy number in the human genome. Nature 444, 444–454.
[24] Stranger, B.E. et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315, 848–853.
[25] Korbel, J.O. et al. (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318, 420–426.
[26] Reinhardt, J.A., Baltrus, D.A., Nishimura, M.T., Jeck, W.R., Jones, C.D. and Dangl, J.L. (2009) De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. Genome Res. 19 (2), 294–305.
[27] Hernandez, D., Francois, P., Farinelli, L., Osteras, M. and Schrenzel, J. (2008) De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res. 18, 802–809.
[28] Weber, A.P., Weber, K.L., Carr, K., Wilkerson, C. and Ohlrogge, J.B. (2007) Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol. 144, 32–42.
[29] Cloonan, N. and Grimmond, S.M. (2008) Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol. 9, 234.
[30] Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Meth. 5, 621–628.
[31] Yassour, M. et al. (2009) Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. Proc. Natl. Acad. Sci. USA 106, 3264–3269.
[32] Campbell, C.D., Kirby, A., Nemesh, J., Daly, M.J. and Hirschhorn, J.N. (2008) A survey of allelic imbalance in F1 mice. Genome Res. 18, 555–563.
[33] Serre, D. et al. (2008) Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. PLoS Genet. 4, e1000006.
[34] Petretto, E. et al. (2006) Heritability and tissue specificity of expression quantitative trait loci. PLoS Genet. 2, e172.
[35] Heinzen, E.L. et al. (2008) Tissue-specific genetic control of splicing: implications for the study of complex traits. PLoS Biol. 6, 2869–2879.
[36] Wold, B. and Myers, R.M. (2008) Sequence census methods for functional genomics. Nat. Meth. 5, 19–21.
[37] Valouev, A., Johnson, D.S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R.M. and Sidow, A. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat. Meth. 5, 829–834.
[38] Barski, A. et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129, 823–837.
[39] Mikkelsen, T.S. et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560.
[40] Simonis, M., Kooren, J. and de Laat, W. (2007) An evaluation of 3C-based methods to capture DNA interactions. Nat. Meth. 4, 895–901.
[41] Dostie, J. and Dekker, J. (2007) Mapping networks of physical interactions between genomic elements using 5C technology. Nat. Protoc. 2, 988–1002.