

Review

## Next-Generation Sequencing: From Understanding Biology to Personalized Medicine

Karen S. Frese<sup>1</sup>, Hugo A. Katus<sup>1,2</sup> and Benjamin Meder<sup>1,2,\*</sup>

<sup>1</sup> Department of Internal Medicine III, University of Heidelberg, Heidelberg 69120, Germany; E-Mail: karen.frese@med.uni-heidelberg.de

<sup>2</sup> DZHK (German Centre for Cardiovascular Research)—partner site, Heidelberg/Mannheim, Department of Internal Medicine III, University of Heidelberg, Im Neuenheimer Feld 350, Heidelberg 69120, Germany; E-Mail: hugo.katus@med.uni-heidelberg.de

\* Author to whom correspondence should be addressed; E-Mail: benjamin.meder@med.uni-heidelberg.de; Tel.: +49-0-6221-564835; Fax: +49-0-6221-564645.

Received: 21 January 2013; in revised form: 21 January 2013 / Accepted: 4 February 2013 /

Published: 1 March 2013

---

**Abstract:** Within just a few years, the new methods for high-throughput next-generation sequencing have generated completely novel insights into the heritability and pathophysiology of human disease. In this review, we wish to highlight the benefits of the current state-of-the-art sequencing technologies for genetic and epigenetic research. We illustrate how these technologies help to constantly improve our understanding of genetic mechanisms in biological systems and summarize the progress made so far. This can be exemplified by the case of heritable heart muscle diseases, so-called cardiomyopathies. Here, next-generation sequencing is able to identify novel disease genes, and first clinical applications demonstrate the successful translation of this technology into personalized patient care.

**Keywords:** next-generation sequencing; genomics; epigenomics; transcriptomics; cardiomyopathy; heart failure

---

## 1. From Genes, Biology and Disease

Every cell in our organism is a highly dynamic biological system that must continuously respond and adapt to multiple intrinsic and extrinsic factors. In this ever-changing system, the genome of the cell is its only constant and master plan for its behavior. Hence, if we want to understand the very complex molecular networks of cells in health and disease, we need to enlighten the secrets of the genome itself.

Genetic research has begun with fascinating investigations on simple phenotypic characteristics, with Gregor Mendel discovering the basics of inheritance in 1856 without even knowing the molecular composition of the genome. While he studied the color of pea seeds, his rules haven't changed and are applicable to many diseases today. By now, over 4,000 Mendelian diseases have been recognized, with new ones discovered every year. Additionally, the era of genome-wide association studies has led to new paradigms, discovering the role of non-coding and intergenic variants and their contribution to highly prevalent, complex diseases [1–3]. It seems desirable to understand the single genetic contribution to each of these diseases to allow personalized diagnosis and therapies, a promise that was made after the first human genome was successfully sequenced and which has not been fulfilled yet.

### *Why Four Letters of Genetic Code Are So Complicated*

It is evident that the genome is far more complex than previously thought. While the understanding of its coding regions has considerably advanced, 99% of the non-coding sequence is still challenging researchers from different disciplines to finally unravel all of the functions of the genome. Additionally, genetic variation across different individuals and populations is higher than estimated, and the transition from common to rare variants is fluid, making interpretation of their functional relevance difficult; structural changes, such as insertions, deletions or copy number variations, are far more frequent than previously thought. For instance, the Database of Genomic Variants (DGV) lists about 60,000 CNVs, 850 inversions and 30,000 insertion/deletions identified in healthy individuals [4].

The sequencing of the genome has laid the groundwork for many investigations that improved our knowledge on the biology and molecular principles, not only of Mendelian disorders, but of human disease in general. However, at the same time, it has raised a vast amount of new questions. For example, the completion of the human genome project revealed that there are just around 20,000 protein-coding genes, a surprisingly low number in comparison to the complexity of the human organism [4,5]. It has become apparent that a biological trait is not necessarily caused by a single gene/protein or its mutation. While rare variants are a prototype for Mendelian disorders, common variants are too frequent to be disease causing. In recent years, their disease contribution and associated biological mechanisms were successfully uncovered by genome-wide association studies. Any of the common variants alone may not affect a trait, but put together, they can add up to or result in a significant phenotypic difference. Furthermore, complex diseases are heterogeneous, often as a result of the cumulative effects of genetic and environmental influences, exerting disease susceptibility over time. Hence, the genotypic components of complex diseases are not causative, but rather mediate disease risk and further result from the cumulative effect of low penetrance variants that are frequently found in the general population, usually displaying an allele frequency >1–5%. Large-scale projects, such as

ENCODE, considerably increased our knowledge on these coding and non-coding regions and the functional implications of variations of the human genome [3,5]. They take the big challenge to elucidate the functional link between associated variants and phenotypic traits and the development of methods to dissect the role of common and rare variants in aggregate. Methods, such as NGS, and projects, such as ENCODE, help shed light into these unsolved mysteries, since they investigate the whole genome, transcriptome and epigenome and, therefore, are able to provide an unbiased and comprehensive view on biological systems not available before.

## 2. Next-Generation Sequencing—Towards Understanding Biology

The advances of sequencing technologies have successfully contributed in elucidating the function of the human genome. NGS technologies have gained the capacity to sequence gigabases of DNA in a high-throughput and highly efficient manner that has not been possible using traditional Sanger sequencing. While Sanger is based on gel separation of chain-terminated fragments from enzymatic synthesis [6,7], most NGS techniques are based on locally bound nano-clusters of template DNA and incorporation of fluorescent-labeled nucleotides by DNA polymerases or ligation processes. The read lengths of current NGS approaches are relatively short, due to the small sequencing colonies and progressive signal deterioration (35–500 bp), compared to traditional sequencing (1,000–1,200 bp), which in turn is compensated by its highly-paralleled fashion. Technical and chemical refinements are used to steadily increase the read lengths [8,9], but only novel technologies, as nanopore sequencing, will be able to provide substantially longer reads.

Novel platforms of the third generation are under development, which are based on real-time sequencing of the DNA templates without prior amplification [10]. Braslavsky *et al.* introduced one of the first techniques for single-molecule sequencing [11], and fluorescence-based single-molecule sequencing methods are now available from Pacific Bioscience or Helicos. Another innovative sequencing technique is the Oxford Nanopore DNA sequencer that is free of nucleotide labeling. The technology is based on an electrical current fingerprint of each nucleotide, which is produced by the nucleotides passing through a  $\alpha$ -hemolysin nanopore. Therefore, the nanopore is immersed in a conducting fluid, and after application of a potential voltage, an electric current, due to conduction of ions through the nanopore, can be observed [12–14]. These improvements and maturation of third generation sequencers will make the analysis of genetic variations in genomes more feasible in the near future.

The massive data produced by current NGS systems presents a significant challenge for data storage and analysis. A number of computational tools and databases have been newly developed to handle base calling, alignment of sequence reads to a reference, *de novo* assembly, variant detection/filtering and annotation [15–18]. This basic analysis already is demanding, but the interpretation of the large number of genetic variants is far more complex. An excellent overview of suitable software tools and databases is provided by Bao *et al.* [16,19–21].

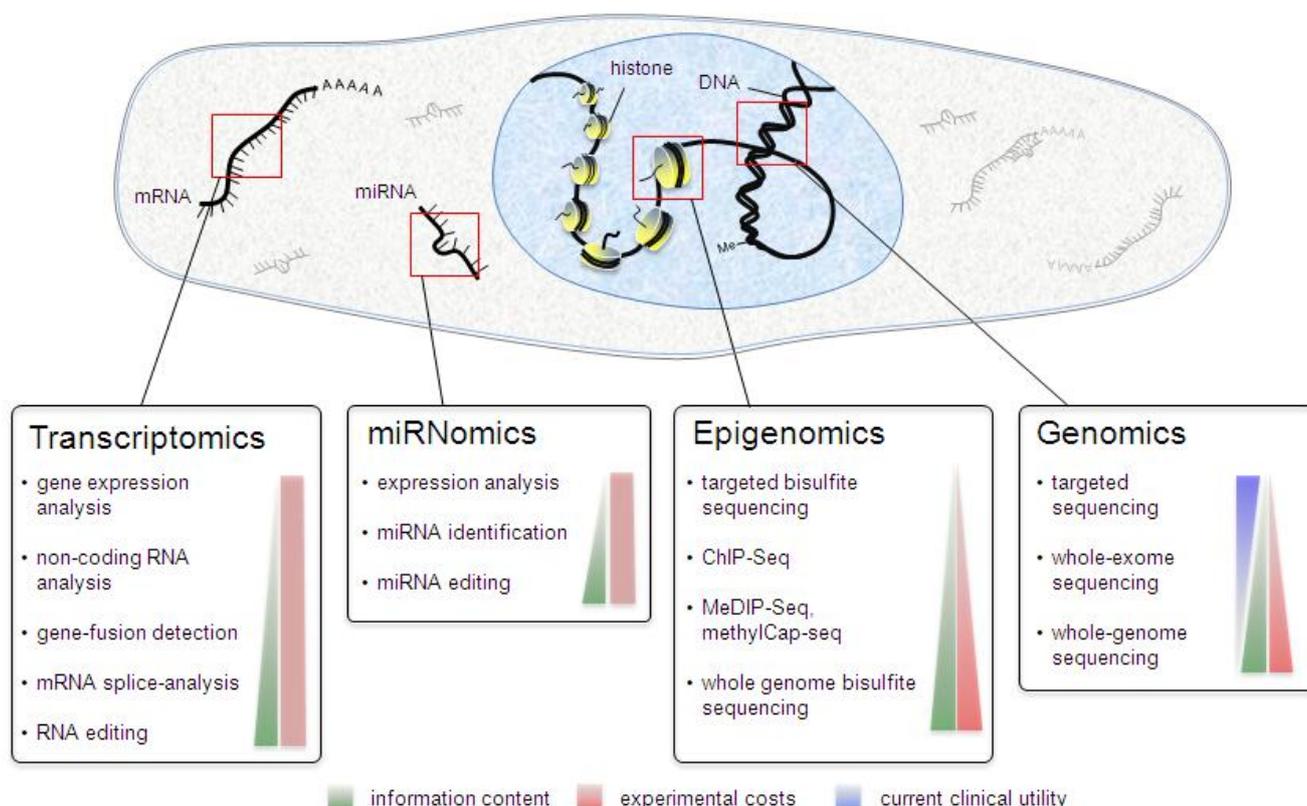
So far, several diseases and syndromes have been dissected by NGS approaches. Now, the systematic detection and annotation of the complete genome, as well as correct interpretation of its variations, transcription start and polyadenylation sites, exon-intron structures, splice variants and regulatory sequences is required to advance our understanding of biology. The recently published

ENCODE project helped to systematically map the regions of gene transcription, transcription factor binding and chromatin modifications, assigning functional properties to 80% of the whole genome.

### 3. Genomics

In the following paragraphs, we want to provide an overview on different NGS applications, starting with genomic sequencing (Figure 1).

**Figure 1.** Next-generation sequencing applications. Schematogram depicting the different methods for transcriptomic, miRNomic, epigenomic and genomic studies.



DNA sequencing by NGS can be applied within different applications, such as partial-exome (PES), whole-exome (WES) or whole-genome sequencing (WGS). The broad range of applications opens new and more affordable possibilities to study numerous cellular processes at the single-base resolution. However, both WES and WGS produce massive amounts of data, which presents significant challenges for data storage, distribution, analysis and interpretation. In the nearer future, this will remain one of the main bottlenecks of all described approaches.

#### 3.1. Exome Sequencing

Partial and whole-exome sequencing is a relatively cost-efficient method to detect genome-wide variations in exons and adjacent splice-sites. Considering that Mendelian diseases typically affect the protein-coding regions of the genome, exome sequencing is particularly relevant to discover such rare-variants. The advances made so far in NGS technologies makes WES, with an average cost of approximately 700–1,000 US\$ per sample, a widely applicable tool for discovering rare alleles

underlying Mendelian phenotypes and complex traits [22,23]. To date, WES has been successfully used to identify several disease-causing genes [24], including Miller syndrome, Freeman-Sheldon syndrome [25], Floating-Harbor syndrome [26], Kabuki syndrome [27] and spino-cerebellar ataxia [28]. In the cardiovascular field, exome sequencing has successfully identified novel causal genes, including *SHROOM3* for heterotaxy [29,30] and *ANGPTL3* in cases of familial combined hypolipidemia [30,31] or novel mutations in known disease genes for dilated and hypertrophic cardiomyopathy [32–34]. In some cases, WES alone also fails to identify the causal variant. Galmiche *et al.* used exome sequencing in conjunction with genetic mapping to identify a mutation in the mitochondrial ribosomal protein, MRPL3, in a family with mitochondrial cardiomyopathy [35]. A combined approach of WES and copy-number variation helped Norton *et al.* to bypass missing sequencing depth to reveal a deletion in *BAG3* to be causative for familial dilated cardiomyopathy [36]. Although studies showed that WES can detect variants missed by WGS due to coverage reasons [37], it is likely that this approach will soon be superseded by high-coverage, high-quality whole-genome sequencing.

### 3.2. Whole-Genome Sequencing

Alterations in regulatory sequences and non-coding regions account for a significant proportion of genetic susceptibility to common and complex diseases. WGS holds the great advantage that it enables the grasping of variations, not only in the protein coding genes, but it also assesses the large non-coding parts of the genome. Meanwhile, WGS is only approximately five- to ten-fold more expensive than exome sequencing, and costs are expected to further decrease dramatically in the next few years [38].

Especially in cancer research, it was recognized early on that it is important to target all types of somatic/germ-line genetic alterations, including nucleotide substitution, small insertions and deletions, CNVs and chromosomal rearrangements, also of non-coding regions [39]. However, also in neurological and cardiovascular diseases, WGS is now successfully applied to dissect causative variants. Lupski *et al.*, for instance, identified by WGS a family with a recessive form of Charcot-Marie-Tooth disease, a clinically relevant heterozygous mutation in the *SH3TC2* gene [40]. Despite these achievements, the functional understanding of the millions of identified variants per genome is still challenging. Hence, integrative systems biology approaches in combination with genetic model systems, such as iPS-cells, zebrafish or mice, provide powerful tools to analyze genetic alterations and their biological effects [41,42]. Hence, only an integrative approach of *in silico*, *in vitro* and *in vivo* model systems with WGS may hold the key to facilitate the interpretation of genomic variation and allow a more accurate prediction of the clinical impact of coding, as well as non-protein-coding variations.

## 4. Transcriptomics

The DNA in multicellular organisms contains the same genetic information in every cell (with the exception of gametes or neoplastic cells); the transcriptome of different cells, however, largely varies depending on the cell type, its function and temporal state. The transcriptome describes the complete set of all RNA molecules in a cell, in the sum determined by the genes that are actively expressed and the RNAs that underlay active or passive degradation processes. The exploration of complex cellular

processes, like gene expression, alternative splicing, allele-specific expression and RNA editing, are still challenging. Next-generation sequencing techniques allow analyzing RNA-levels, RNA-editing and isoform-analysis in a single, unbiased experiment. Furthermore, rRNA, tRNA and other non-coding RNAs (miRNA, ncRNA, siRNA) can be investigated.

#### 4.1. Defective RNA Processing and Disease

During mRNA maturation, pre-mRNAs undergo a sequence of chemical and structural changes, such as capping, editing, splicing and polyadenylation. Alternative splicing is one main step in RNA maturation, and transcriptomic diversification is a tissue-specific and developmentally strictly regulated process [43,44]. Regulation of alternative splicing is dependent on sequence motifs in the genes to be spliced and by various splicing factors and associated proteins. Recent genome-wide analyses of alternative splicing show that at least 60% of human genes have alternatively spliced variants [45], suggesting that alternative splicing is one of the most important mechanisms to create the functional complexity of eukaryotic cells [46,47].

All of the described steps need to be well controlled [48–50], and hence, defects in this fine-tuned processes are increasingly recognized as probable causes of inherited human diseases [51–56]. Dysregulation of cell type-specific alternative splicing and mutations in several splicing factors have been characterized in cancer, cardiomyopathies and neurological disorders [44,57–59]. Currently, it is estimated that 50–60% of inherited diseases involve defective splicing, making the understanding of splicing mechanisms and regulation an important area of research [60].

#### 4.2. NGS Methods to Study the Transcriptome

Until now, microarrays are still the most commonly used technique for measuring gene expression, allowing high throughput analysis of thousands of target genes in parallel. Nevertheless, microarrays have some considerable drawbacks, such as problems with unspecific hybridization and limited dynamic range. Also, they cannot be easily used to detect splice-events or previously unknown transcripts [61].

Next-generation sequencing protocols for RNAs, often referred to as RNA-seq, provide direct access to the transcript levels and sequences without the prior knowledge about the targets to be analyzed. RNA-seq is an unbiased, rapid, precise and, meanwhile, not too expensive method to quantify the expression of genes and to detect tissue-specific transcript isoforms, even without a reference genome or predesigned probes [62]. Depending on the read length, RNA-seq is able to pinpoint the location of transcription boundaries and reveals important information about how exons are connected [63,64]. Further, RNA-seq shows a high level of technical reproducibility [65], as well as a high accuracy in expression quantification [66]. In comparison to conventional microarray-based methods, it also allows the identification of sequence polymorphisms and posttranscriptional mRNA editing [62]. Further, software for *de novo* reconstruction of transcriptomes from RNA-seq data, such as Trinity, are promising approaches that allow assembly of full-length transcripts, even without a complete reference sequence [67], particularly useful for model organisms with limited knowledge about their genomes [62].

#### 4.3. Gene-Expression Analyses and mRNA Splicing

In recent years, RNA-seq was successfully applied to dissect gene dysregulation that contributes to the pathogenesis of cardiovascular diseases, cancer, Chronic Obstructive Pulmonary Disease or Type 2 Diabetes, in some cases directly or indirectly caused by genetic variants in proteins involved in different steps during RNA processing [68]. Hence, variation in gene expression levels also can be heritable [69], and polymorphisms that affect the expression levels of genes are most often found near the gene itself (cis-regulation) and especially near the transcriptional start sites [70–72]. Pickrell *et al.* used RNA-seq to generate a map of the transcriptional landscape of 69 lymphoblastoid cell lines derived from unrelated Nigerian individuals who had been completely genotyped by the HapMap consortium. Thus, by using this genotype data, they identified over 1,000 genes at which genetic variation influences expression levels or splicing. They also demonstrate that eQTLs near genes mostly act by a mechanism involving allele-specific expression and that variation that influences the inclusion of an exon is enriched within or near the consensus splice sites [72].

Various studies using RNA-seq have shown an improvement of assessing alternative splicing and detection of novel transcripts in comparison to splicing-arrays. RNA-seq was used on several human tissues and cell lines uncovering a larger number of alternative splicing events in humans than previously thought [43,73]. Isoforms differ most drastically between tissues, whereas differences between individuals are almost three-fold less common [74]. The dysregulation of cell-type-specific alternative splicing and mutations in several splicing factors could already be associated with various diseases [44,57,58,75,76]. It is, for instance, known that aberrant alternative splicing is tightly associated with the development of heart failure, with aberrant splicing of cardiac troponins linked to the progression of cardiomyopathies [77–79]. Since it is known that genetic mutations in the alternative-splicing regulators, such as RBM20 or RBM24, are associated with cardiomyopathy [80–83], RNA-seq approaches are now of major interest for target identification of these splicing factors.

#### 4.4. RNA Editing and Non-Coding RNAs

Transcriptome diversity is further increased by RNA editing, which results in a different product than that encoded by the DNA template. RNA editing is a process of site-specific modification of the mRNA sequence. This process involves deamination of adenosines into inosines, which are read as guanines. The substitution of adenosine to inosine is catalyzed by members of the double-stranded RNA-specific Adenosine Deaminase enzymes (ADAR). Since RNA editing can lead to the formation of an altered protein if editing results in a codon exchange, this process may be an essential post-transcriptional mechanism for expanding the proteomic diversity [84]. Altered RNA-editing patterns were found to be associated with a number of human pathologies, including inflammation, epilepsy, depression, amyotrophic lateral sclerosis (ALS) and cancerogenesis [85–88].

For a long time, only a handful of editing sites within coding sequences have been well characterized [89]. However, this poorly understood process is becoming clearer now, due to advances in RNA-seq technologies. Bioinformatics analyses have predicted that RNA-editing is apparently more abundant than previously thought, affecting thousands of human genes [90,91]. A pioneering RNA-seq

study of human brain and other tissues has revealed hundreds of new RNA editing sites, many of them located in non-coding RNAs [12].

In recent years, sequencing technologies have revealed that at least 90% of the genome is actively transcribed, giving rise to thousands of non-coding transcripts [92–94]. Interestingly, non-protein-coding sequences have been found to be rapidly evolving in vertebrate genomes [95] and increases proportionally with organism complexity [96], whereas the part of protein-coding genes remains relatively unchanged. ENCODE further elucidated that 80% of the genome contains functional elements defined as discrete genome segments that encode a product, for instance, protein or non-coding RNA, or display a reproducible biochemical role, e.g., transcription factor binding [1]. The ENCODE consortium mapped functional sites at high resolution across the genome integrating results from 147 different cell types and other resources, such as candidate regions from GWAS and evolutionarily constrained regions. The most prevalent functional elements identified were regions being transcribed into RNAs, including transfer RNA, microRNA, small nuclear RNA and small nucleolar RNA (tRNA, miRNA, snRNA and snoRNA). Of note, these regions covered 62% of the genome, mainly inside introns or near genes [97].

Different classes of small and large non-coding RNAs (ncRNAs) have been shown to regulate gene expression at the transcriptional level through a direct interaction with the transcriptional machinery. This results in either transcriptional activation or transcriptional repression determined by physiological and developmental processes. Some ncRNA are strongly linked to epigenetic regulation influencing chromatin-remodeling complexes, chromatin architecture, post-transcriptional processing and translation [92,98–102]. However, the precise functional significance of most of these non-coding transcripts remains unclear. Some of them could be considered biological noise [103], but there are already many ncRNAs that are known to have diverse functions in developmental and disease pathways [104–108], shown for cancer, central nervous system disorders, neurodegenerative disease and cardiovascular disease [104,109–111]. A recent RNA-seq analysis by Lee *et al.* revealed that more than 100 lncRNAs were differentially expressed in hypertrophic mouse hearts [112], suggesting a relevant role in proper heart function. This assumption is supported by the analysis of transcriptional levels of *ANRIL* and *MIAT* ncRNAs, which was linked to the pathogenesis of CAD and myocardial infarction [113–115]. Short ncRNAs (sncRNA) include miRNAs, which are the best-studied ncRNAs. They are known to be involved in the specific regulation of protein-coding genes, by post-transcriptional silencing or infrequently by activation [116,117]. Their pivotal role in several diseases has been dissected extensively in many studies about cardiovascular diseases [118–120], and their application as therapeutics and biomarkers is just in sight [121–123]. Interestingly, data from recent reports reveal that miRNAs can also regulate the expression of other types of ncRNAs (such as long ncRNAs), suggesting that miRNAs can even impact on independent regulatory networks [124–129].

## 5. Epigenomics

Transcriptional regulation is frequently controlled epigenetically by mechanisms that do not directly depend on the underlying DNA sequence [130]. Hence, to better understand the complex interactions of different regulatory factors with the genome, it is essential to perform multilayered

approaches, including the analyses of gene transcription, alternative splicing, histone modification and DNA methylation.

### 5.1. Role of Epigenetic Alterations

Epigenetic modifications of the genome are now known to be involved in many cellular processes, such as embryonic development, transcription, chromatin structuring, X chromosome inactivation, genomic imprinting and chromosome stability [131]. Hence, differences in epigenetic modifications might explain changes in disease susceptibility and progression without underlying variations in the DNA sequence. Since epigenetic regulation is sensitive to environmental changes, it is thought to be a major mechanism by which external stimuli induce (an inheritable) response, together called genome-environment interaction [132].

To date, most of the epigenetic studies have focused on aberrant DNA methylation patterns, particularly in embryonic development and cancer biology [131,133,134]. Hence, methylation abnormalities were often found to occur in signaling pathways that regulate proliferation, migration, growth, differentiation, transcription and death signals. The studying of epigenetic mechanisms in different tumor types revealed that cytosine methylation is one of the earliest events in tumorigenesis. Consequently, several epigenetic markers have been identified for cancer detection, diagnosis and treatment [135], and the exploration of epigenetic alterations in other human diseases has become a special focus of current research. However, surprisingly few studies have addressed the role of epigenetic regulation in the pathogenesis of cardiovascular disease, although it is commonly recognized that not only the genetic background, but lifestyle and yet unknown factors influence cardiovascular morbidity [136,137]. This may be in part due to limited access to myocardial tissue from patients. A recent report profiled for epigenetic modifications in explanted hearts of end-stage heart failure patients [138]. This study from Movassagh *et al.* found different methylation patterns between end-stage heart failure and control human hearts in CGIs within gene promoters and gene bodies. Moreover, the observed decreased gene promoter methylation that correlated with upregulated transcripts, but not *vice versa*. By a genome-wide approach, we could recently identify alterations in cardiac DNA methylation that are associated with human dilated cardiomyopathy. By consecutive fine-mapping and biological validation in independent cohorts and *in vivo* characterization in zebrafish, it was underlined in this study that epigenetic modifications of distinct pathways and modifiers are functionally involved in the pathogenesis of DCM [139].

A recently published review by Leung *et al.* comprehensively summarizes studies that demonstrated that single nucleotide polymorphisms are assumed to be associated with altered DNA methylation and chromatin accessibility, implicating that genomic variants can modify epigenetic patterns [140,141]. Hence, the integration of DNA methylation information in current population-based studies might help to explain the impact of variants or disease-causing alleles to the onset and progression of common and complex diseases.

Unlike genome-wide variation data, which is included and steadily updated in a wealth of databases, such as the Human Genome Project, 1,000 Genomes Project or HapMap, whole-epigenome data only started to be systematically identified and catalogue. The now increasing amount of methylation data for many tissues, pathological conditions and species are deposited, for instance, in

the NGSmethDB [142] or ENCODE [143] databases that will be useful for multilayered analytical approaches to improve our knowledge about epigenetic mechanisms.

### 5.2. NGS Methods to Assess the Epigenome

Bisulfite sequencing or MeDIP-Seq (methylated DNA immunoprecipitation coupled with sequencing) are used to capture global DNA methylation changes, while ChIP-Seq is applied to identify histone modifications and to analyze transcription factor binding sites [144,145]. So far, the “gold standard” for detection of cytosine methylation comprises a sodium bisulfite conversion of the DNA, followed by a sequencing step. Cytosine methylation, which is the addition of a methyl group at the carbon 5 position of cytosine through DNA methyltransferase enzymes (DNMT), plays an important role in transcriptional regulation and is the most extensively studied epigenetic modification. Recently, various high-throughput NGS approaches have been combined with bisulfite DNA conversion for genome-wide analysis of DNA methylation by discriminating methylated and unmethylated cytosines [146–149]. Alternative methods include immunoprecipitation of methylated DNA (MeDIP) [150] or Methyl-Capture sequencing (MethylCap-seq). Yu *et al.* demonstrated, for instance, the applicability of MethylCap-seq, which combines precipitation of methylated DNA by the recombinant methyl-CpG binding domain of MBD2 protein with NGS, to dissect genome-wide DNA methylation profiles of the Cisplatin-sensitive ovarian cancer cell line [151]. MeDIP has been widely used to explore the methylomes of plants, mice and human cells [152–157], and a recently improved protocol enables MeDIP-seq analysis with very low DNA concentrations [158].

Histone modifications, chromatin-remodeling factors and binding sites for transcription factors are today mostly analyzed using chromatin immunoprecipitation sequencing in combination with NGS (ChIP-seq). As compared to previously used ChIP-chip approaches, which are also based on chromatin immunoprecipitation, but on a microarray platform [159,160], ChIP-seq, on the one hand, does not have the typical microarray-specific limitations and, on the other hand, offers the opportunity for *de novo* motif discovery. Robertson *et al.* and Euskirchen *et al.*, for instance, catalogued binding sites of the transcription factors STAT1 and NRSF in human cells by ChIP-seq and highlighted the excellent resolution and low necessity of extensive replicates for the method [161,162].

## 6. NGS—Towards Personalized Medicine

The introduction of NGS technologies has tremendously changed the landscape of genomic research. As described above, NGS has led to important discoveries in biomedical research and has already been implemented in clinical diagnostics, too. Early studies reported such a successful translation of NGS into clinical workflows using whole-genome, whole-exome and enrichment-based sequencing approaches for different diseases. Especially in oncology, NGS-based diagnostic testing has already achieved a broader clinical impact. Walsh *et al.*, for instance, demonstrated the clinical applicability of NGS in cancer diagnostics using target region capture and NGS to detect germ-line mutations in 21 tumor suppressor genes in genomic DNA from women with primary ovarian, peritoneal or fallopian tube carcinoma. Other publications analyzed the versatility of NGS for clinical applications for diseases, such as retinitis pigmentosa, inflammatory bowel disease, neurofibromatosis, Charcot-Marie-tooth neuropathy, Kabuki syndrome and others [40,163–166]. As recently published,

we could demonstrate the diagnostic capabilities of NGS as a clinical diagnostic tool for dilated and hypertrophic cardiomyopathy [32]. Importantly, the enrichment-driven NGS approach in this study yielded consistently high sequence and target coverage, as well as good specificity and sensitivity compared to Sanger sequencing. Considering that costs of high-coverage WGS are still substantial, targeted sequencing approaches are currently delivering the best data quality for clinical applications in which a known panel of genes needs to be tested.

In biomarker discovery, NGS is already applied in many screening studies, e.g., for miRNA or epigenomic signatures. It is foreseeable that NGS technologies can also be applied in clinical biomarker assays, especially when complex, maybe temporal or multivariate biomarker signatures are used to increase diagnostic performance [167]. At the same time, it must be noted that the benefits of NGS technologies still brings with it a number of other challenges that must be meticulously addressed before they can be transferred from the research field into routine clinical application. In particular, a comprehensive and transparent analysis strategy of the large amount of sequence information and their interpretation is indispensable. This poses a challenge to both laboratory and clinical geneticists and requires appropriate training of different disciplines, which may directly facilitate the application of genome-based medicine. Some clinical pilot studies relying on NGS in daily practice underscore the value of such a multidisciplinary team dedicated to the collection and interpretation of NGS data and establishment of “best practices” for clinical NGS [168]. Finally, we are facing numerous political, ethical and social challenges by the advancements and progression of translational genomics, bearing fears of discrimination, breach of confidentiality and data security [169–171]. It is in our hands to carefully address these issues and establish NGS in the clinics for the sake of our patients.

### Acknowledgements

Our research is supported by grants from the German Ministry of Education and Research (BMBF): NGFN-plus 01GS0836, NGFN-transfer 01GR0823, INSIGHT DCM, DZHK (“Deutsches Zentrum für Herz-Kreislauf-Forschung”—German Centre for Cardiovascular Research); the University of Heidelberg (Innovationsfond FRONTIER) and the European Union (FP7 BestAgeing and INHERITANCE).

### References and Notes

1. Dunham, I.; Kundaje, A.; Aldred, S.F.; Collins, P.J.; Davis, C.A.; Doyle, F.; Epstein, C.B.; Frietze, S.; Harrow, J.; Kaul, R.; *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74.
2. Sastre, L. Clinical implications of the encode project. *Clin. Transl. Oncol.* **2012**, *14*, 801–802.
3. Frazer, K.A. Decoding the human genome. *Genome Res.* **2012**, *22*, 1599–1601.
4. Database of Genomic Variants. Available online: <http://projects.tcag.ca/> (accessed on 5 July 2012).
5. Ecker, J.R.; Bickmore, W.A.; Barroso, I.; Pritchard, J.K.; Gilad, Y.; Segal, E. Genomics: Encode explained. *Nature* **2012**, *489*, 52–55.
6. Sanger, F.; Nicklen, S.; Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463–5467.
7. Sanger, F.; Coulson, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **1975**, *94*, 441–448.

8. Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L.; Law, M. Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, *2012*, 251364.
9. Genomeweb. Available online: <http://genomeweb.com/> (accessed on 12 December 2012).
10. Schadt, E.E.; Turner, S.; Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet.* **2010**, *19*, R227–R240.
11. Braslavsky, I.; Hebert, B.; Kartalov, E.; Quake, S.R. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3960–3964.
12. Pareek, C.S.; Smoczynski, R.; Tretyn, A. Sequencing technologies and genome sequencing. *J. Appl. Genet.* **2011**, *52*, 413–435.
13. Astier, Y.; Braha, O.; Bayley, H. Toward single molecule DNA sequencing: Direct identification of ribonucleoside and deoxyribonucleoside 5'-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J. Am. Chem. Soc.* **2006**, *128*, 1705–1710.
14. Rusk, N. Focus on next-generation sequencing data analysis. *Nat. Methods* **2009**, *6*, S1.
15. Lee, H.C.; Lai, K.; Lorenc, M.T.; Imelfort, M.; Duran, C.; Edwards, D. Bioinformatics tools and databases for analysis of next-generation sequence data. *Brief Funct. Genomics* **2012**, *11*, 12–24.
16. Torri, F.; Dinov, I.D.; Zamanyan, A.; Hobel, S.; Genco, A.; Petrosyan, P.; Clark, A.P.; Liu, Z.; Eggert, P.; Pierce, J.; *et al.* Next generation sequence analysis and computational genomics using graphical pipeline workflows. *Genes* **2012**, *3*, 545–575.
17. Afgan, E.; Chapman, B.; Taylor, J. Cloudman as a platform for tool, data, and analysis distribution. *BMC Bioinformatics* **2012**, *13*, 315.
18. Schadt, E.E.; Linderman, M.D.; Sorenson, J.; Lee, L.; Nolan, G.P. Computational solutions to large-scale data management and analysis. *Nat. Rev. Genet.* **2010**, *11*, 647–657.
19. Abeel, T.; van Parys, T.; Saeys, Y.; Galagan, J.; van de Peer, Y. Genomeview: A next-generation genome browser. *Nucleic Acids Res.* **2012**, *40*, e12.
20. Bao, S.; Jiang, R.; Kwan, W.; Wang, B.; Ma, X.; Song, Y.Q. Evaluation of next-generation sequencing software in mapping and assembly. *J. Hum. Genet.* **2011**, *56*, 406–414.
21. Cordero, F.; Beccuti, M.; Donatelli, S.; Calogero, R.A. Large disclosing the nature of computational tools for the analysis of next generation sequencing data. *Curr. Top. Med. Chem.* **2012**, *12*, 1320–1330.
22. Hershberger, R.E.; Siegfried, J.D. Update 2011: Clinical and genetic issues in familial dilated cardiomyopathy. *J. Am. Coll Cardiol.* **2011**, *57*, 1641–1649.
23. Bamshad, M.J.; Ng, S.B.; Bigham, A.W.; Tabor, H.K.; Emond, M.J.; Nickerson, D.A.; Shendure, J. Exome sequencing as a tool for mendelian disease gene discovery. *Nat. Rev. Genet.* **2011**, *12*, 745–755.
24. Ng, S.B.; Buckingham, K.J.; Lee, C.; Bigham, A.W.; Tabor, H.K.; Dent, K.M.; Huff, C.D.; Shannon, P.T.; Jabs, E.W.; Nickerson, D.A.; *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **2010**, *42*, 30–35.
25. Ng, S.B.; Turner, E.H.; Robertson, P.D.; Flygare, S.D.; Bigham, A.W.; Lee, C.; Shaffer, T.; Wong, M.; Bhattacharjee, A.; Eichler, E.E.; *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **2009**, *461*, 272–276.
26. Hood, R.L.; Lines, M.A.; Nikkel, S.M.; Schwartzentruber, J.; Beaulieu, C.; Nowaczyk, M.J.; Allanson, J.; Kim, C.A.; Wiczorek, D.; Moilanen, J.S.; *et al.* Mutations in srcap, encoding

- snf2-related crebbp activator protein, cause floating-harbor syndrome. *Am. J. Hum. Genet.* **2012**, *90*, 308–313.
27. Ng, S.B.; Bigham, A.W.; Buckingham, K.J.; Hannibal, M.C.; McMillin, M.J.; Gildersleeve, H.I.; Beck, A.E.; Tabor, H.K.; Cooper, G.M.; Mefford, H.C.; *et al.* Exome sequencing identifies mll2 mutations as a cause of kabuki syndrome. *Nat. Genet.* **2010**, *42*, 790–793.
  28. Wang, J.L.; Yang, X.; Xia, K.; Hu, Z.M.; Weng, L.; Jin, X.; Jiang, H.; Zhang, P.; Shen, L.; Guo, J.F.; *et al.* Tgm6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* **2010**, *133*, 3510–3518.
  29. Tariq, M.; Belmont, J.W.; Lalani, S.; Smolarek, T.; Ware, S.M. Shroom3 is a novel candidate for heterotaxy identified by whole exome sequencing. *Genome Biol.* **2011**, *12*, R91.
  30. Lara-Pezzi, E.; Dopazo, A.; Manzanares, M. Understanding cardiovascular disease: A journey through the genome (and what we found there). *Dis. Model Mech.* **2012**, *5*, 434–443.
  31. Musunuru, K.; Pirruccello, J.P.; Do, R.; Peloso, G.M.; Guiducci, C.; Sougnez, C.; Garimella, K.V.; Fisher, S.; Abreu, J.; Barry, A.J.; *et al.* Exome sequencing, angptl3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **2010**, *363*, 2220–2227.
  32. Meder, B.; Haas, J.; Keller, A.; Heid, C.; Just, S.; Borries, A.; Boisguerin, V.; Scharfenberger-Schmeer, M.; Stahler, P.; Beier, M.; *et al.* Targeted next-generation sequencing for the molecular genetic diagnostics of cardiomyopathies. *Circ. Cardiovasc. Genet.* **2011**, *4*, 110–122.
  33. Herman, D.S.; Lam, L.; Taylor, M.R.; Wang, L.; Teekakirikul, P.; Christodoulou, D.; Conner, L.; DePalma, S.R.; McDonough, B.; Sparks, E.; *et al.* Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* **2012**, *366*, 619–628.
  34. Gerull, B.; Gramlich, M.; Atherton, J.; McNabb, M.; Trombitas, K.; Sasse-Klaassen, S.; Seidman, J.G.; Seidman, C.; Granzier, H.; Labeit, S.; *et al.* Mutations of ttn, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat. Genet.* **2002**, *30*, 201–204.
  35. Galmiche, L.; Serre, V.; Beinat, M.; Assouline, Z.; Lebre, A.S.; Chretien, D.; Nietschke, P.; Benes, V.; Boddaert, N.; Sidi, D.; *et al.* Exome sequencing identifies mrpl3 mutation in mitochondrial cardiomyopathy. *Hum. Mutat.* **2011**, *32*, 1225–1231.
  36. Norton, N.; Li, D.; Rieder, M.J.; Siegfried, J.D.; Rampersaud, E.; Zuchner, S.; Mangos, S.; Gonzalez-Quintana, J.; Wang, L.; McGee, S.; *et al.* Genome-wide studies of copy number variation and exome sequencing identify rare variants in bag3 as a cause of dilated cardiomyopathy. *Am. J. Hum. Genet.* **2011**, *88*, 273–282.
  37. Clark, M.J.; Chen, R.; Lam, H.Y.; Karczewski, K.J.; Euskirchen, G.; Butte, A.J.; Snyder, M. Performance comparison of exome DNA sequencing technologies. *Nat. Biotechnol.* **2011**, *29*, 908–914.
  38. Norton, N.; Li, D.; Hershberger, R.E. Next-generation sequencing to identify genetic causes of cardiomyopathies. *Curr. Opin. Cardiol.* **2012**, *27*, 214–220.
  39. Meyerson, M.; Gabriel, S.; Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* **2010**, *11*, 685–696.
  40. Lupski, J.R.; Reid, J.G.; Gonzaga-Jauregui, C.; Rio Deiros, D.; Chen, D.C.; Nazareth, L.; Bainbridge, M.; Dinh, H.; Jing, C.; Wheeler, D.A.; *et al.* Whole-genome sequencing in a patient with charcot-marie-tooth neuropathy. *N. Engl. J. Med.* **2010**, *362*, 1181–1191.

41. Smith, K.A.; Joziase, I.C.; Chocron, S.; van Dinther, M.; Guryev, V.; Verhoeven, M.C.; Rehmann, H.; van der Smagt, J.J.; Doevendans, P.A.; Cuppen, E.; *et al.* Dominant-negative alk2 allele associates with congenital heart defects. *Circulation* **2009**, *119*, 3062–3069.
42. Meder, B.; Laufer, C.; Hassel, D.; Just, S.; Marquart, S.; Vogel, B.; Hess, A.; Fishman, M.C.; Katus, H.A.; Rottbauer, W. A single serine in the carboxyl terminus of cardiac essential myosin light chain-1 controls cardiomyocyte contractility *in vivo*. *Circ. Res.* **2009**, *104*, 650–659.
43. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413–1415.
44. Wang, G.S.; Cooper, T.A. Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* **2007**, *8*, 749–761.
45. Chacko, E.; Ranganathan, S. Comprehensive splicing graph analysis of alternative splicing patterns in chicken, compared to human and mouse. *BMC Genomics* **2009**, *10*, S5.
46. Modrek, B.; Lee, C. A genomic view of alternative splicing. *Nat. Genet.* **2002**, *30*, 13–19.
47. Matlin, A.J.; Clark, F.; Smith, C.W. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 386–398.
48. Bentley, D. Coupling rna polymerase ii transcription with pre-mRNA processing. *Curr. Opin. Cell Biol.* **1999**, *11*, 347–351.
49. Minvielle-Sebastia, L.; Keller, W. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.* **1999**, *11*, 352–357.
50. Maniatis, T.; Reed, R. An extensive network of coupling among gene expression machines. *Nature* **2002**, *416*, 499–506.
51. Philips, A.V.; Cooper, T.A. RNA processing and human disease. *Cell. Mol. Life Sci.* **2000**, *57*, 235–249.
52. Stoss, O.; Olbrich, M.; Hartmann, A.M.; Konig, H.; Memmott, J.; Andreadis, A.; Stamm, S. The star/gsg family protein rslm-2 regulates the selection of alternative splice sites. *J. Biol. Chem.* **2001**, *276*, 8665–8673.
53. Jensen, K.B.; Dredge, B.K.; Stefani, G.; Zhong, R.; Buckanovich, R.J.; Okano, H.J.; Yang, Y.Y.; Darnell, R.B. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **2000**, *25*, 359–371.
54. Mendell, J.T.; Dietz, H.C. When the message goes awry: Disease-producing mutations that influence mrna content and performance. *Cell* **2001**, *107*, 411–414.
55. Caceres, J.F.; Kornblihtt, A.R. Alternative splicing: Multiple control mechanisms and involvement in human disease. *Trends Genet.* **2002**, *18*, 186–193.
56. Nissim-Rafinia, M.; Kerem, B. Splicing regulation as a potential genetic modifier. *Trends Genet.* **2002**, *18*, 123–127.
57. Faustino, N.A.; Cooper, T.A. Pre-mRNA splicing and human disease. *Genes Dev.* **2003**, *17*, 419–437.
58. Cooper, T.A.; Wan, L.; Dreyfuss, G. RNA and disease. *Cell* **2009**, *136*, 777–793.
59. Venables, J.P. Aberrant and alternative splicing in cancer. *Cancer Res.* **2004**, *64*, 7647–7654.
60. Hammond, S.M.; Wood, M.J. Genetic therapies for rna mis-splicing diseases. *Trends Genet.* **2011**, *27*, 196–205.

61. Tang, F.; Lao, K.; Surani, M.A. Development and applications of single-cell transcriptome analysis. *Nat. Methods* **2011**, *8*, S6–S11.
62. Wang, Z.; Gerstein, M.; Snyder, M. RNA-seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63.
63. Raghavachari, N.; Barb, J.; Yang, Y.; Liu, P.; Woodhouse, K.; Levy, D.; O'Donnell, C.J.; Munson, P.J.; Kato, G.J. A systematic comparison and evaluation of high density exon arrays and rna-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. *BMC Med. Genomics* **2012**, *5*, 28.
64. Feng, J.; Li, W.; Jiang, T. Inference of isoforms from short sequence reads. *J. Comput. Biol.* **2011**, *18*, 305–321.
65. Marioni, J.C.; Mason, C.E.; Mane, S.M.; Stephens, M.; Gilad, Y. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **2008**, *18*, 1509–1517.
66. Nagalakshmi, U.; Wang, Z.; Waern, K.; Shou, C.; Raha, D.; Gerstein, M.; Snyder, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **2008**, *320*, 1344–1349.
67. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; *et al.* Full-length transcriptome assembly from rna-seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652.
68. Herbert, A.; Rich, A. RNA processing in evolution. The logic of soft-wired genomes. *Ann. N. Y. Acad. Sci.* **1999**, *870*, 119–132.
69. Kwan, T.; Benovoy, D.; Dias, C.; Gurd, S.; Serre, D.; Zuzan, H.; Clark, T.A.; Schweitzer, A.; Staples, M.K.; Wang, H.; *et al.* Heritability of alternative splicing in the human genome. *Genome Res.* **2007**, *17*, 1210–1218.
70. Cheung, V.G.; Spielman, R.S.; Ewens, K.G.; Weber, T.M.; Morley, M.; Burdick, J.T. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **2005**, *437*, 1365–1369.
71. Veyrieras, J.B.; Kudaravalli, S.; Kim, S.Y.; Dermitzakis, E.T.; Gilad, Y.; Stephens, M.; Pritchard, J.K. High-resolution mapping of expression-qtls yields insight into human gene regulation. *PLoS Genet.* **2008**, *4*, e1000214.
72. Pickrell, J.K.; Marioni, J.C.; Pai, A.A.; Degner, J.F.; Engelhardt, B.E.; Nkadori, E.; Veyrieras, J.B.; Stephens, M.; Gilad, Y.; Pritchard, J.K. Understanding mechanisms underlying human gene expression variation with rna sequencing. *Nature* **2010**, *464*, 768–772.
73. Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B.B.; Siddiqui, A.; *et al.* mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **2009**, *6*, 377–382.
74. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtkova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative isoform regulation in human tissue transcriptomes. *Nature* **2008**, *456*, 470–476.
75. Backes, C.; Meese, E.; Lenhof, H.P.; Keller, A. A dictionary on micrnas and their putative target pathways. *Nucleic Acids Res.* **2010**, *38*, 4476–4486.

76. David, C.J.; Manley, J.L. Alternative pre-mRNA splicing regulation in cancer: Pathways and programs unhinged. *Genes Dev.* **2010**, *24*, 2343–2364.
77. Biesiadecki, B.J.; Elder, B.D.; Yu, Z.B.; Jin, J.P. Cardiac troponin t variants produced by aberrant splicing of multiple exons in animals with high instances of dilated cardiomyopathy. *J. Biol. Chem.* **2002**, *277*, 50275–50285.
78. Neagoe, C.; Kulke, M.; del Monte, F.; Gwathmey, J.K.; de Tombe, P.P.; Hajjar, R.J.; Linke, W.A. Titin isoform switch in ischemic human heart disease. *Circulation* **2002**, *106*, 1333–1341.
79. Philips, A.V.; Timchenko, L.T.; Cooper, T.A. Disruption of splicing regulated by a cug-binding protein in myotonic dystrophy. *Science* **1998**, *280*, 737–741.
80. Poon, K.L.; Tan, K.T.; Wei, Y.Y.; Ng, C.P.; Colman, A.; Korzh, V.; Xu, X.Q. RNA-binding protein rbm24 is required for sarcomere assembly and heart contractility. *Cardiovasc. Res.* **2012**, *94*, 418–427.
81. Refaat, M.M.; Lubitz, S.A.; Makino, S.; Islam, Z.; Frangiskakis, J.M.; Mehdi, H.; Gutmann, R.; Zhang, M.L.; Bloom, H.L.; MacRae, C.A.; *et al.* Genetic variation in the alternative splicing regulator rbm20 is associated with dilated cardiomyopathy. *Heart Rhythm.* **2012**, *9*, 390–396.
82. Brauch, K.M.; Karst, M.L.; Herron, K.J.; de Andrade, M.; Pellikka, P.A.; Rodeheffer, R.J.; Michels, V.V.; Olson, T.M. Mutations in ribonucleic acid binding protein gene cause familial dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **2009**, *54*, 930–941.
83. Guo, W.; Schafer, S.; Greaser, M.L.; Radke, M.H.; Liss, M.; Govindarajan, T.; Maatz, H.; Schulz, H.; Li, S.; Parrish, A.M.; *et al.* Rbm20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat. Med.* **2012**, *18*, 766–773.
84. Bass, B.L. RNA editing and hypermutation by adenosine deamination. *Trends Biochem. Sci.* **1997**, *22*, 157–162.
85. Maas, S.; Patt, S.; Schrey, M.; Rich, A. Underediting of glutamate receptor glur-b mrna in malignant gliomas. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14687–14692.
86. Patterson, J.B.; Samuel, C.E. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: Evidence for two forms of the deaminase. *Mol. Cell Biol.* **1995**, *15*, 5376–5388.
87. Kawahara, Y.; Ito, K.; Sun, H.; Aizawa, H.; Kanazawa, I.; Kwak, S. Glutamate receptors: RNA editing and death of motor neurons. *Nature* **2004**, *427*, 801.
88. Dominissini, D.; Moshitch-Moshkovitz, S.; Amariglio, N.; Rechavi, G. Adenosine-to-inosine rna editing meets cancer. *Carcinogenesis* **2011**, *32*, 1569–1577.
89. Seeburg, P.H.; Higuchi, M.; Sprengel, R. RNA editing of brain glutamate receptor channels: Mechanism and physiology. *Brain Res. Brain Res. Rev.* **1998**, *26*, 217–229.
90. Athanasiadis, A.; Rich, A.; Maas, S. Widespread a-to-i RNA editing of alu-containing mRNAs in the human transcriptome. *PLoS Biol.* **2004**, *2*, e391.
91. Levanon, E.Y.; Eisenberg, E.; Yelin, R.; Nemzer, S.; Hallegger, M.; Shemesh, R.; Fligelman, Z.Y.; Shoshan, A.; Pollock, S.R.; Sztybel, D.; *et al.* Systematic identification of abundant a-to-i editing sites in the human transcriptome. *Nat. Biotechnol.* **2004**, *22*, 1001–1005.
92. Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigo, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T.; Thurman, R.E.; *et al.* Identification and analysis of

- functional elements in 1% of the human genome by the encode pilot project. *Nature* **2007**, *447*, 799–816.
93. Costa, F.F. Non-coding rnas: Meet thy masters. *Bioessays* **2010**, *32*, 599–608.
  94. Mattick, J.S. Non-coding RNAs: The architects of eukaryotic complexity. *EMBO Rep.* **2001**, *2*, 986–991.
  95. Pang, K.C.; Frith, M.C.; Mattick, J.S. Rapid evolution of noncoding RNAs: Lack of conservation does not mean lack of function. *Trends Genet.* **2006**, *22*, 1–5.
  96. Costa, V.; Angelini, C.; de Feis, I.; Ciccodicola, A. Uncovering the complexity of transcriptomes with RNA-seq. *J. Biomed. Biotechnol.* **2010**, *2010*, 853916.
  97. Fratkin, E.; Bercovici, S.; Stephan, D.A. The implications of encode for diagnostics. *Nat. Biotechnol.* **2012**, *30*, 1064–1065.
  98. Tilgner, H.; Knowles, D.G.; Johnson, R.; Davis, C.A.; Chakraborty, S.; Djebali, S.; Curado, J.; Snyder, M.; Gingeras, T.R.; Guigo, R. Deep sequencing of subcellular rna fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncnas. *Genome Res.* **2012**, *22*, 1616–1625.
  99. Amaral, P.P.; Mattick, J.S. Noncoding rna in development. *Mamm. Genome* **2008**, *19*, 454–492.
  100. Mattick, J.S.; Amaral, P.P.; Dinger, M.E.; Mercer, T.R.; Mehler, M.F. RNA regulation of epigenetic processes. *Bioessays* **2009**, *31*, 51–59.
  101. Mattick, J.S. The genetic signatures of noncoding rnas. *PLoS Genet.* **2009**, *5*, e1000459.
  102. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; *et al.* The gencode v7 catalog of human long noncoding rnas: Analysis of their gene structure, evolution, and expression. *Genome Res.* **2012**, *22*, 1775–1789.
  103. Struhl, K. Transcriptional noise and the fidelity of initiation by rna polymerase II. *Nat. Struct. Mol. Biol.* **2007**, *14*, 103–105.
  104. Taft, R.J.; Pang, K.C.; Mercer, T.R.; Dinger, M.; Mattick, J.S. Non-coding RNAs: Regulators of disease. *J. Pathol.* **2010**, *220*, 126–139.
  105. Orom, U.A.; Derrien, T.; Guigo, R.; Shiekhatar, R. Long noncoding rnas as enhancers of gene expression. *Cold Spring Harb. Symp. Quant. Biol.* **2010**, *75*, 325–331.
  106. Orom, U.A.; Derrien, T.; Beringer, M.; Gumireddy, K.; Gardini, A.; Bussotti, G.; Lai, F.; Zytnicki, M.; Notredame, C.; Huang, Q.; *et al.* Long noncoding rnas with enhancer-like function in human cells. *Cell* **2010**, *143*, 46–58.
  107. Hung, T.; Wang, Y.; Lin, M.F.; Koegel, A.K.; Kotake, Y.; Grant, G.D.; Horlings, H.M.; Shah, N.; Umbrecht, C.; Wang, P.; *et al.* Extensive and coordinated transcription of noncoding rnas within cell-cycle promoters. *Nat. Genet.* **2011**, *43*, 621–629.
  108. Huarte, M.; Guttman, M.; Feldser, D.; Garber, M.; Koziol, M.J.; Kenzelmann-Broz, D.; Khalil, A.M.; Zuk, O.; Amit, I.; Rabani, M.; *et al.* A large intergenic noncoding rna induced by p53 mediates global gene repression in the p53 response. *Cell* **2010**, *142*, 409–419.
  109. Christov, C.P.; Trivier, E.; Krude, T. Noncoding human y RNAs are overexpressed in tumours and required for cell proliferation. *Br. J. Cancer* **2008**, *98*, 981–988.
  110. Angeloni, D.; ter Elst, A.; Wei, M.H.; van der Veen, A.Y.; Braga, E.A.; Klimov, E.A.; Timmer, T.; Korobeinikova, L.; Lerman, M.I.; Buys, C.H. Analysis of a new homozygous deletion in the

- tumor suppressor region at 3p12.3 reveals two novel intronic noncoding rna genes. *Genes Chromosomes Cancer* **2006**, *45*, 676–691.
111. Koob, M.D.; Moseley, M.L.; Schut, L.J.; Benzow, K.A.; Bird, T.D.; Day, J.W.; Ranum, L.P. An untranslated ctg expansion causes a novel form of spinocerebellar ataxia (sca8). *Nat. Genet.* **1999**, *21*, 379–384.
  112. Lee, J.H.; Gao, C.; Peng, G.; Greer, C.; Ren, S.; Wang, Y.; Xiao, X. Analysis of transcriptome complexity through rna sequencing in normal and failing murine hearts. *Circ. Res.* **2011**, *109*, 1332–1341.
  113. Harismendy, O.; Notani, D.; Song, X.; Rahim, N.G.; Tanasa, B.; Heintzman, N.; Ren, B.; Fu, X.D.; Topol, E.J.; Rosenfeld, M.G.; *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* **2011**, *470*, 264–268.
  114. Ishii, N.; Ozaki, K.; Sato, H.; Mizuno, H.; Saito, S.; Takahashi, A.; Miyamoto, Y.; Ikegawa, S.; Kamatani, N.; Hori, M.; *et al.* Identification of a novel non-coding RNA, miat, that confers risk of myocardial infarction. *J. Hum. Genet.* **2006**, *51*, 1087–1099.
  115. Jarinova, O.; Stewart, A.F.; Roberts, R.; Wells, G.; Lau, P.; Naing, T.; Buerki, C.; McLean, B.W.; Cook, R.C.; Parker, J.S.; *et al.* Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler. Thromb. Vasc. Biol.* **2009**, *29*, 1671–1677.
  116. Bartel, D.P. Micrnas: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233.
  117. Krol, J.; Loedige, I.; Filipowicz, W. The widespread regulation of microrna biogenesis, function and decay. *Nat. Rev. Genet.* **2010**, *11*, 597–610.
  118. Ramsingh, G.; Koboldt, D.C.; Trissal, M.; Chiappinelli, K.B.; Wylie, T.; Koul, S.; Chang, L.W.; Nagarajan, R.; Fehniger, T.A.; Goodfellow, P.; *et al.* Complete characterization of the micrornaome in a patient with acute myeloid leukemia. *Blood* **2010**, *116*, 5316–5326.
  119. Latronico, M.V.; Condorelli, G. Micrnas and cardiac pathology. *Nat. Rev. Cardiol.* **2009**, *6*, 419–429.
  120. Calin, G.A.; Croce, C.M. Investigation of microrna alterations in leukemias and lymphomas. *Methods Enzymol.* **2007**, *427*, 193–213.
  121. Keller, A.; Backes, C.; Leidinger, P.; Kefer, N.; Boisguerin, V.; Barbacioru, C.; Vogel, B.; Matzas, M.; Huwer, H.; Katus, H.A.; *et al.* Next-generation sequencing identifies novel micrnas in peripheral blood of lung cancer patients. *Mol. Biosyst.* **2011**, *7*, 3187–3199.
  122. Keller, A.; Leidinger, P.; Bauer, A.; Elsharawy, A.; Haas, J.; Backes, C.; Wendschlag, A.; Giese, N.; Tjaden, C.; Ott, K.; *et al.* Toward the blood-borne mirnome of human diseases. *Nat. Methods* **2011**, *8*, 841–843.
  123. Meder, B.; Keller, A.; Vogel, B.; Haas, J.; Sedaghat-Hamedani, F.; Kayvanpour, E.; Just, S.; Borries, A.; Rudloff, J.; Leidinger, P.; *et al.* Microrna signatures in total peripheral blood as novel biomarkers for acute myocardial infarction. *Basic Res. Cardiol.* **2011**, *106*, 13–23.
  124. Calin, G.A.; Pekarsky, Y.; Croce, C.M. The role of microrna and other non-coding RNA in the pathogenesis of chronic lymphocytic leukemia. *Best Pract. Res. Clin. Haematol.* **2007**, *20*, 425–437.
  125. Calin, G.A.; Liu, C.G.; Ferracin, M.; Hyslop, T.; Spizzo, R.; Sevignani, C.; Fabbri, M.; Cimmino, A.; Lee, E.J.; Wojcik, S.E.; *et al.* Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **2007**, *12*, 215–229.

126. Rossi, S.; Sevignani, C.; Nnadi, S.C.; Siracusa, L.D.; Calin, G.A. Cancer-associated genomic regions (cagrs) and noncoding RNAs: Bioinformatics and therapeutic implications. *Mamm. Genome* **2008**, *19*, 526–540.
127. Braconi, C.; Kogure, T.; Valeri, N.; Huang, N.; Nuovo, G.; Costinean, S.; Negrini, M.; Miotto, E.; Croce, C.M.; Patel, T. MicroRNA-29 can regulate expression of the long non-coding RNA gene *meg3* in hepatocellular cancer. *Oncogene* **2011**, *30*, 4750–4756.
128. Jeggari, A.; Marks, D.S.; Larsson, E. Mircode: A map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* **2012**, *28*, 2062–2063.
129. Chen, X.; Liang, H.; Zhang, C.Y.; Zen, K. Mirna regulates noncoding RNA: A noncanonical function model. *Trends Biochem. Sci.* **2012**, *37*, 457–459.
130. Egger, G.; Liang, G.; Aparicio, A.; Jones, P.A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **2004**, *429*, 457–463.
131. Robertson, K.D. DNA methylation and human disease. *Nat. Rev. Genet.* **2005**, *6*, 597–610.
132. Ordovas, J.M.; Smith, C.E. Epigenetics and cardiovascular disease. *Nat. Rev. Cardiol.* **2010**, *7*, 510–519.
133. Szyf, M. The role of DNA hypermethylation and demethylation in cancer and cancer therapy. *Curr. Oncol.* **2008**, *15*, 72–75.
134. Szyf, M.; Pakneshan, P.; Rabbani, S.A. DNA methylation and breast cancer. *Biochem. Pharmacol.* **2004**, *68*, 1187–1197.
135. Banerjee, H.N.; Verma, M. Epigenetic mechanisms in cancer. *Biomark Med.* **2009**, *3*, 397–410.
136. Vinci, M.C.; Polvani, G.; Pesce, M. Epigenetic programming and risk: The birthplace of cardiovascular disease? *Stem Cell Rev.* **2012**, doi:10.1007/s12015-012-9398-z.
137. Barker, D.J. Fetal programming of coronary heart disease. *Trends Endocrinol. Metab.* **2002**, *13*, 364–368.
138. Movassagh, M.; Choy, M.K.; Knowles, D.A.; Cordeddu, L.; Haider, S.; Down, T.; Siggins, L.; Vujic, A.; Simeoni, I.; Penkett, C.; *et al.* Distinct epigenomic features in end-stage failing human hearts. *Circulation* **2011**, *124*, 2411–2422.
139. Haas, J.; Frese, K.S.; Park, Y.J.; Keller, A.; Vogel, B.; Lindroth, A.M.; Weichenhan, D.; Franke, J.; Fischer, S.; Bauer, A.; *et al.* Alterations in cardiac DNA methylation in human dilated cardiomyopathy. *EMBO Mol. Med.* **2013**, *5*, 1–17.
140. Bjornsson, H.T.; Fallin, M.D.; Feinberg, A.P. An integrated epigenetic and genetic approach to common human disease. *Trends Genet.* **2004**, *20*, 350–358.
141. Leung, A.; Schones, D.E.; Natarajan, R. Using epigenetic mechanisms to understand the impact of common disease causing alleles. *Curr. Opin. Immunol.* **2012**, *24*, 558–563.
142. Hackenberg, M.; Barturen, G.; Oliver, J.L. Ngsmethdb: A database for next-generation sequencing single-cytosine-resolution DNA methylation data. *Nucleic Acids Res.* **2011**, *39*, D75–D79.
143. Ryu, H.J.; Kim do, Y.; Park, J.Y.; Chang, H.Y.; Lee, M.H.; Han, K.H.; Chon, C.Y.; Ahn, S.H. Clinical features and prognosis of hepatocellular carcinoma with respect to pre-s deletion and basal core promoter mutations of hepatitis b virus genotype c2. *J. Med. Virol.* **2011**, *83*, 2088–2095.
144. Mardis, E.R. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* **2008**, *9*, 387–402.

145. Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends Genet.* **2008**, *24*, 133–141.
146. Suzuki, M.M.; Bird, A. DNA methylation landscapes: Provocative insights from epigenomics. *Nat. Rev. Genet.* **2008**, *9*, 465–476.
147. Laird, P.W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **2010**, *11*, 191–203.
148. Beck, S.; Rakan, V.K. The methylome: Approaches for global DNA methylation profiling. *Trends Genet.* **2008**, *24*, 231–237.
149. Fouse, S.D.; Nagarajan, R.O.; Costello, J.F. Genome-scale DNA methylation analysis. *Epigenomics* **2010**, *2*, 105–117.
150. Jacinto, F.V.; Ballestar, E.; Esteller, M. Methyl-DNA immunoprecipitation (medip): Hunting down the DNA methylome. *Biotechniques* **2008**, *44*, 35, 37, 39 passim.
151. Yu, W.; Jin, C.; Lou, X.; Han, X.; Li, L.; He, Y.; Zhang, H.; Ma, K.; Zhu, J.; Cheng, L.; *et al.* Global analysis of DNA methylation by methyl-capture sequencing reveals epigenetic control of cisplatin resistance in ovarian cancer cell. *PLoS One* **2011**, *6*, e29450.
152. Farthing, C.R.; Ficiz, G.; Ng, R.K.; Chan, C.F.; Andrews, S.; Dean, W.; Hemberger, M.; Reik, W. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **2008**, *4*, e1000116.
153. Dindot, S.V.; Person, R.; Strivens, M.; Garcia, R.; Beaudet, A.L. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res.* **2009**, *19*, 1374–1383.
154. Down, T.A.; Rakan, V.K.; Turner, D.J.; Flicek, P.; Li, H.; Kulesha, E.; Graf, S.; Johnson, N.; Herrero, J.; Tomazou, E.M.; *et al.* A bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* **2008**, *26*, 779–785.
155. Koga, Y.; Pelizzola, M.; Cheng, E.; Krauthammer, M.; Sznol, M.; Ariyan, S.; Narayan, D.; Molinaro, A.M.; Halaban, R.; Weissman, S.M. Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome Res.* **2009**, *19*, 1462–1470.
156. Straussman, R.; Nejman, D.; Roberts, D.; Steinfeld, I.; Blum, B.; Benvenisty, N.; Simon, I.; Yakhini, Z.; Cedar, H. Developmental programming of cpg island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.* **2009**, *16*, 564–571.
157. Weber, M.; Hellmann, I.; Stadler, M.B.; Ramos, L.; Paabo, S.; Rebhan, M.; Schubeler, D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **2007**, *39*, 457–466.
158. Taiwo, O.; Wilson, G.A.; Morris, T.; Seisenberger, S.; Reik, W.; Pearce, D.; Beck, S.; Butcher, L.M. Methylome analysis using medip-seq with low DNA concentrations. *Nat. Protoc.* **2012**, *7*, 617–636.
159. Weinmann, A.S.; Yan, P.S.; Oberley, M.J.; Huang, T.H.; Farnham, P.J. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and cpg island microarray analysis. *Genes Dev.* **2002**, *16*, 235–244.
160. Ballestar, E.; Paz, M.F.; Valle, L.; Wei, S.; Fraga, M.F.; Espada, J.; Cigudosa, J.C.; Huang, T.H.; Esteller, M. Methyl-cpg binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J.* **2003**, *22*, 6335–6345.

161. Robertson, G.; Hirst, M.; Bainbridge, M.; Bilenky, M.; Zhao, Y.; Zeng, T.; Euskirchen, G.; Bernier, B.; Varhol, R.; Delaney, A.; *et al.* Genome-wide profiles of stat1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **2007**, *4*, 651–657.
162. Euskirchen, G.M.; Rozowsky, J.S.; Wei, C.L.; Lee, W.H.; Zhang, Z.D.; Hartman, S.; Emanuelsson, O.; Stolc, V.; Weissman, S.; Gerstein, M.B.; *et al.* Mapping of transcription factor binding regions in mammalian cells by chip: Comparison of array- and sequencing-based technologies. *Genome Res.* **2007**, *17*, 898–909.
163. Neveling, K.; Collin, R.W.; Gilissen, C.; van Huet, R.A.; Visser, L.; Kwint, M.P.; Gijzen, S.J.; Zonneveld, M.N.; Wieskamp, N.; de Ligt, J.; *et al.* Next-generation genetic testing for retinitis pigmentosa. *Hum. Mutat.* **2012**, *33*, 963–972.
164. Fokstuen, S.; Munoz, A.; Melacini, P.; Iliceto, S.; Perrot, A.; Ozcelik, C.; Jeanrenaud, X.; Rieubland, C.; Farr, M.; Faber, L.; *et al.* Rapid detection of genetic variants in hypertrophic cardiomyopathy by custom DNA resequencing array in clinical practice. *J. Med. Genet.* **2011**, *48*, 572–576.
165. Worthey, E.A.; Mayer, A.N.; Syverson, G.D.; Helbling, D.; Bonacci, B.B.; Decker, B.; Serpe, J.M.; Dasu, T.; Tschannen, M.R.; Veith, R.L.; *et al.* Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med.* **2011**, *13*, 255–262.
166. Chou, L.S.; Liu, C.S.; Boese, B.; Zhang, X.; Mao, R. DNA sequence capture and enrichment by microarray followed by next-generation sequencing for targeted resequencing: Neurofibromatosis type 1 gene as a model. *Clin. Chem.* **2010**, *56*, 62–72.
167. Vogel, B.; Keller, A.; Frese, K.; Kloos, W.; Kayvanpour, E.; Sedaghat-Hamedani, F.; Hassel, S.; Marquart, S.; Beier, M.; Giannitis, E.; *et al.* Refining diagnostic microRNA signatures by whole-mirnome kinetic analysis in acute myocardial infarction. *Clin. Chem.* **2012**, *59*, 410–418.
168. Davies Thirty Groups Enter Clarity Clinical Genome Interpretation Challenge. Available online: <http://genes.childrenshospital.org/> (accessed on 12 July 2012).
169. Soden, S.E.; Farrow, E.G.; Saunders, C.J.; Lantos, J.D. Genomic medicine: Evolving science, evolving ethics. *Per. Med.* **2012**, *9*, 523–528.
170. De Lecea, M.G.; Rossbach, M. Translational genomics in personalized medicine—Scientific challenges en route to clinical practice. *HUGO J.* **2012**, *6*, 2.
171. Tester, D.J.; Ackerman, M.J. Genetic testing for potentially lethal, highly treatable inherited cardiomyopathies/channelopathies in clinical practice. *Circulation* **2011**, *123*, 1021–1037.