Minireview Article

Genetics to Genomics in Clinical Medicine

ABSTRACT

Biomedical research and knowledge has grown exponentially since the completion of the Human Genome Project in the year 2000. There has been a gradual shift from 'genetics' (study of genes) to 'genomics' (study of the whole genome) in medicine. Advances such as sequencing of the human genome, genome enrichment, epigenetics and bioinformatics have transformed the face of translational research and are beginning to have a major impact on clinical practice. In order to take advantage of the full potential of genomic research in clinical practice, clinicians will need to understand and embrace a significant conceptual shift from 'Mendelian genetics' to 'Post Mendelian genomics'.

A relative lack of genetics to genomics knowledge has been reported amongst senior physicians in major health plans in the United States. This is also true of physicians practicing in the United Kingdom as reflected in the reports by the British Royal Society (BRS), Wellcome Trust and UK department of Health. While large sections of the academic medical community is driving this conceptual shift, a significant proportion of practicing clinicians are not actively involved in these developments. Here we describe the continuum from genetics to genomics in medicine by giving a brief overview of the shift from single gene disorders and chromosomal aberrations to functional genomics and our current understanding of the more dynamic relationship between genotype and phenotype.

Keywords: clinical medicine, genetics, genomics, genotype, phenotype, translational research

1. INTRODUCTION

It is essential that advances in basic science ultimately translate into benefit to our patients and one prerequisite for this to happen is an understanding of the fundamentals of innovation by clinicians.

Many practicing clinicians will have last learned about the genetic basis of disease during their university days. However, over the last 20 years our understanding of the relationship between genetic information and phenotype has evolved significantly. The study of the genetic basis of health and disease is one of the most active and promising areas of basic science research. It also holds great potential to bring new diagnostic and therapeutic modalities to the bedside. Therefore, clinicians need to have an understanding of the techniques involved, their potential and limitations. However, the investigations into the human genome and its role in disease are evolving at an astonishing pace and it is increasingly difficult for a practicing clinician to keep abreast with these developments. While large sections of the academic medical community is driving the conceptual shift in genetics, a significant proportion of the practicing clinicians have not yet familiarised themselves with

these developments. Whilst it is not necessary or expected for a non-academic physician to follow the cutting edge of genome related research, certain milestones have been reached which the forward-thinking clinician may wish to understand. For these advances to translate into patient benefit it is essential that there is active communication between researchers and clinicians, with a mutual understanding of each other's language, challenges and achievements.

We now understand that the fairly mechanistic and rigid model of strictly mendelian genetics, which has led to the discovery of some ground-breaking links between genotype and phenotype is too limited a concept for the vast majority of pathology, or indeed variations in health and performance.

The human genome is a complex macromolecule comprised of 3.2 billion repeating nucleotides of adenine (A), cytosine (C), guanine (G), and thymine (T). The primary base sequence is not the sole explanation for the complex way that genetics governs our biological function. For example the presence of a gene (a series of nucleotides that form a genetic code) does not automatically imply that it is being read and expressed to produce a complementary RNA copy of the DNA sequence (transcription) or that this is then used to generate proteins (translation) and there are a number of genetic and environmental factors that govern this process. Therefore genetics, the study of inherited traits or phenotypes with the basic unit of inheritance being the gene, contrasts with genomics, which refers to the study of functions and interactions of all genes in a genome. This includes the entirety of inherited DNA sequences and a recognition that information in one region (or locus) of the genome is modified by information at many other loci and by non-genetic factors. Functional genomics also includes the study of the dynamic changes in gene products (transcripts, proteins, metabolites) and how these changes mediate normal and abnormal biological function. The term 'Omics' encompasses comprehensive methodologies that attempt to capture the exhaustive output of an organism's genes (genomics), RNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics). The systems biology approach allows the study of networks of interactions, in addition to dissecting the role of individual molecular components [1].

2. MENDELIAN GENETICS

Traditionally, genetic disorders were considered to be caused by defects in the DNA sequence of single genes that are transmitted in Mendelian fashion to the offspring (Figure 1).

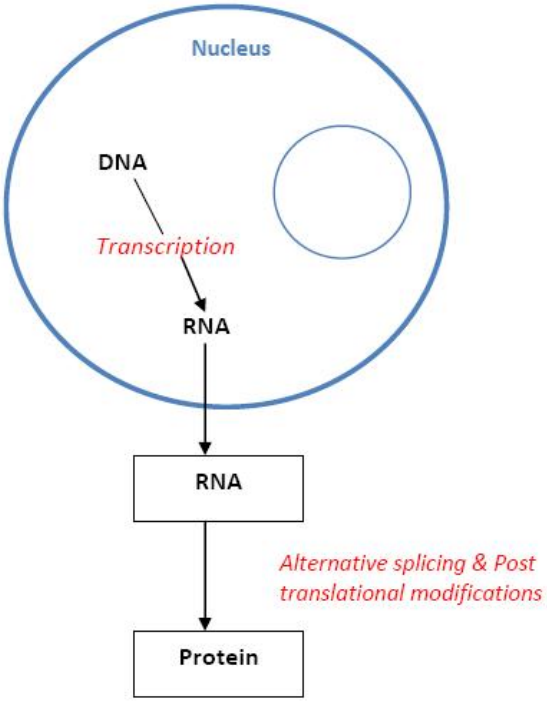

Figure 1: Central Dogma (DNA → RNA → Protein)

Such mutations are responsible for a large number of human diseases e.g. cystic fibrosis, sickle cell anaemia, marfan's syndrome, hypertrophic cardiomyopathy and other pathologies. With a prevalence of 1.4%, they account for considerable morbidity and mortality.

Understanding genetic transmission of Mendelian disorders plays a critical role in diagnosing and managing diseases. For example, an individual with a family history of an autosomal dominant disorder can have an increased likelihood of disease from [1 in 500–5000] in the general population to [1 in 2] in some cases and hence warrants a different approach to assessment in comparison to an individual with sporadic disease [2].

Genetic testing is available for many single gene disorders and timely preventive treatment can be offered if diagnosed at an early age (newborn screening using CFTR-mutation testing has improved management of cystic fibrosis).

These discoveries have had a major clinical impact with gene therapy emerging as a successful treatment option for some single gene disorders. While this genetic approach has been successful in various infectious diseases as well (tuberculosis, malaria), it is currently not replicable in other more complex disorders including asthma and sepsis. Nevertheless, the recent national research fund of £3.1 million awarded to the UK CF Gene therapy consortium reflects the great potential this field holds.[3]

3. CHROMOSOMAL ABNORMALITIES

Mendelian genetics gave rise to cytogenetics, studying heredity and variation. Various methods were developed to visualise chromosome structure and organisation in order to determine genotype-phenotype relationships. Early studies identified associations between syndromic phenotypes and chromosome number abnormalities e.g. Downs, Turners and Kleinefelter's syndromes.

A major turning point was the discovery of the first chromosomal structural abnormality associated with chronic myeloid leukaemia in 1960[4]. Using this information the drug Imatinib was developed (2002), which revolutionised cancer treatment. Cytogenetics has also improved our understanding of Acute Myelogenous Leukemia, Prader-Willi and Angelman syndromes and has led to the identification of *PIK3CA* oncogene associated with ovarian cancer [5].

Cytogenetics entered routine clinical practice in pre-implantation and diagnostics in congenital abnormalities as well as degenerative diseases. Not only did cytogenetic approaches discover associations between human disease and chromosomal abnormalities, it lead to mapping of genes to specific chromosomes. Mapping of Duffy blood group locus to chromosome 1 is a fine example [5].

Various techniques are employed in the study of cytogenetics, including routine analysis of geimsa stained chromosomes, banding techniques, molecular analysis such as fluorescent in situ hybridisation (FISH), spectral karyotyping and comparative genomic hybridisation.

These are described online.

{Further reading: Cytogenetics (WEB1)}

4. POLYGENIC DISORDERS

Genetic disorders can be complex and caused by an interplay of genetic variants with environmental factors. Their pattern of inheritance is not clear-cut. Non-oncological examples include asthma, diabetes, obesity and heart disease.

Genetic insights have brought major advances in the field of cancer. An example of impact on patient care is the concept of preventive mastectomies for women with high-risk mutations in BRCA1 and BRCA2 genes.

5. GENOME WIDE ASSOCIATION STUDIES (GWAS)

While changes in a single DNA sequence imparting a large determinative effect can explain single gene disorders, this does not always hold true in complex phenotypes. Complex diseases result from the cumulative and interactive effects of a large number of gene regions (loci), each imparting a modest marginal effect on phenotypic expression[6]. This principle, commonly known as common disease-common variant hypothesis, suggests that a profile or pattern of multiple common alleles (one of two or more forms of a gene) contributes to the risk of developing common diseases. This underpins genome wide association studies (GWAS).

GWAS aim to find genetic variants associated with a particular disease by scanning markers across DNA sets of a species. GWAS search the genome for small variations called Single Nucleotide Polymorphisms (SNP, often called 'Snips'), which occur more frequently in people with a particular disease than in people without it. A SNP is a DNA sequence variation occurring when a single nucleotide (A,T,C or G) in the genome differs between members of a biological species or paired chromosomes in an individual. So the unit of genetic information examined is far smaller–a single nucleotide, rather than a sequence, but the investigation is designed to look at whole profiles of small genetic variations across many sites. GWAS might lead to examining the complete genomic sequence of individuals to identify all genetic variations, but currently we rely on the principle of linkage disequilibrium (LD) to identify a set of common variants that are statistical proxies for genetic variation at a particular frequency. LD describes the non-random association between two alleles at different locations.

{Further reading: GWAS (WEB2)}

The shift of research focus from Mendelian disorders to the current emphasis on GWAS was enabled by the completion of the Human genome project in the year 2000, public availability of vast amounts of detailed sequence information and development of high throughput genetic technologies. Advances in information technology are fundamental in harnessing this wealth of data[6].

GWAS have lead to a better understanding of the genetic basis of complex diseases in which the patients' risk is determined by a combination of many genetic variations e.g. coronary artery disease, hypertension, stroke susceptibility[7,8]. One of the most significant clinical impacts of GWAS has been their contribution to pharmacogenomics (effect of genetic variations on response to medication). In cardiovascular medicine, recognizing that 25% of patients have a sub-therapeutic antiplatelet response to clopidogrel, researchers have identified several genetic variants affecting the metabolism of clopidogrel, a prodrug, to its active metabolite. Of these, the CYP2C19 variant allele has been best linked to impaired clopidogrel metabolism, reduced platelet inhibition, and a higher risk of adverse cardiovascular events after percutaneous coronary interventions. Because of the cumulative data, the Food and Drug Administration has now altered the prescribing information for clopidogrel based on CYP2C19 genotype, a move that foreshadows the development of companion diagnostic testing and alternative inhibitors of ADP-mediated platelet activation that do not require metabolism by CYP2C19. Genotype guided clopidogrel prescription is a major advancement in the field of genomics [9].

By their very nature, GWAS focus on a small percentage of the total genome and explain a small proportion of heritability given the low odd ratios. Hence there is an increased risk of missing rare variants, irrespective of whether these are in coding or non-coding regions. Capturing all possible variation within a sample requires a sequencing strategy.

6. GENE SEQUENCING

Gene sequencing is the determination of the precise sequence of nucleotides in a DNA sample. Sequencing of the human genome, however, has been a daunting task, at least until the very recent years. The Human Genome Project, which was launched in 1990 with the primary goal of deciphering the sequence of the human genome, took more than a decade to complete, even in a draft form, and cost nearly $3 billion[10]. DNA sequencing technology, however, has undergone a colossal evolution since the beginning of the Sanger method in 1980. New techniques that sequence millions of DNA strands in parallel have been developed. The new technologies, which are collectively referred to as *next generation*

sequencing (NGS) platforms have increased DNA sequencing output and reduced the cost of DNA sequencing by 500,000-fold [11]. Recent advances in technology (3rd generation sequencers), may well deliver on the promise to provide the '1000 Dollar Genome': the ability to sequence the whole human genome for $1000.

{Further reading: Sequencing (WEB3)}

These advances in GWAS and sequencing could have a substantial impact on medical care. The results of the Encode project demonstrated multiple regulatory functions of so-called 'junk' DNA and its potential role in understanding conditions like diabetes and heart disease. These effects of non-coding RNA, for example, might explain GWAS hits in gene deserts. (Glossary) The vision is for increasingly personalised medicine, whereby healthcare interventions (treatment and prevention programs) would be based on individuals' genomic make up. An example includes the use of genomic information in the risk prediction models of coronary disease [12-14]. Genotype based risk prediction is fixed from birth, allows early risk prediction, is less susceptible to biological variation over life, is easy to obtain with minimal measurement error.

7. EPIGENETICS AND THE 'OMICS' ERA

Although DNA sequence variation plays a major role in determining phenotype and 'DNA→RNA→Protein' remains the central dogma in the 'omics' era, discoveries in genome science have revealed more complex interactions that determine clinical phenotype.

Functional genomics investigates dynamic changes in genes and gene products (transcripts, proteins, metabolites). *Epigenetics* identifies mechanisms independent of nucleotide sequence-such as DNA methylation, histone deacetylation or RNA epigenetics. Figure 2 depicts the dynamic genotype-phenotype relationship.

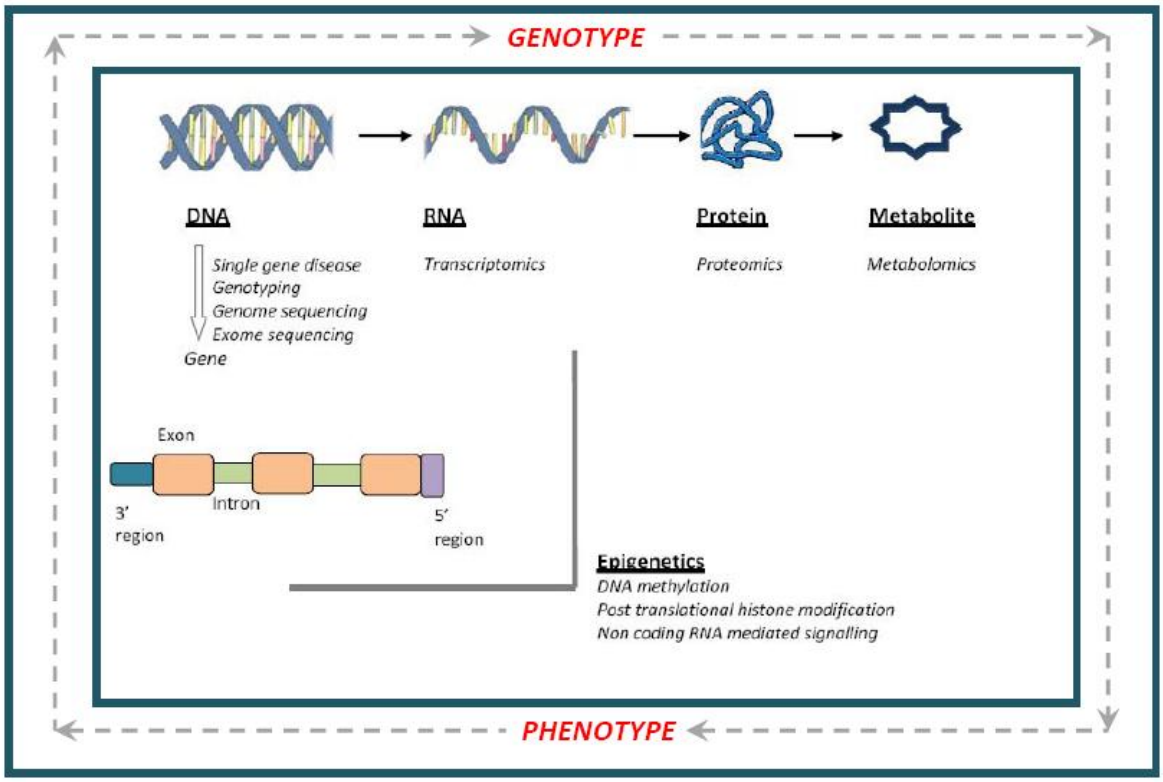

Figure 2: Genotype-Phenotype dynamics

Various '*omics*' terms are used to categorize concepts that interrogate these dynamic interactions. These are:

7.1 Transcriptomics

The mere presence of a gene does not mean that it is being read and expressed to produce a complementary RNA copy of the DNA sequence (transcription) or that this is then used to generate proteins (translation). Despite the identical genome, there is tremendous variability in gene expression in different tissues in response to environmental stimuli. This variation may play a significant role in governing health and disease.

Transcriptomics helps understand the link between the genetic code and molecules governing cell function by studying the RNA transcripts produced by the genome (transcriptome).

Over the last decade transcriptomics (microarray technology) has contributed enormously to our understanding of the molecular basis of cancer. It is now possible to develop potential biomarkers that could be useful in diagnosis and prognosis and would also help achieve the goal of individualized cancer treatment. This technology has also been successful in research into infectious diseases like tuberculosis. Microarray studies have lead to the identification of biomarkers differentiating active and latent TB and have also evaluated mechanisms underlying variability in efficacy of BCG vaccination globally along-with development of chemo/immune therapy [15].

Transcriptomic studies are accomplished by using gene expression microarrays, RNA sequencing or mRNA FISH to quantify the abundance of all transcripts expressed in a tissue of interest under a given biological state [1]. The resulting data contain a large amount of information regarding genes that are turned on or off in the setting of a disease. This information can be used to identify individual genes of interest or gene panels that change together.

Transcriptome databases have been created by the National Human Genome Research Institute (NHGRI) – Mammalian Gene Collection (mgc.nci.nih.gov) and Mouse Transcriptome Project (ncbi.nlm.nih.gov).

7.2 Epigenetics

Studies changes in gene function that occur without a change in DNA sequence. It refers to the functionally relevant modifications to the genome that do not involve a change in the nucleotide sequence. Epigenetic mechanisms explain the ability of certain chemicals to initiate biological perturbations that can lead to malignancy and have also established a causal link between certain infectious diseases and cancer [16]. The key feature that distinguishes epigenetic modifications from genetic changes is their reversible nature. Epigenetic modifications include DNA methylation, post-translational histone modifications and non-coding RNA mediated signaling pathways.

{Further reading: Epigenetics (WEB4)}

DNA methylation measurement techniques include measurement of global methylation content (amount of methylation in test DNA regardless of position), gene specific analysis and high coverage methods including genome wide scans and arrays. The global methylation content can be analysed by direct measurement of methyl group content by using high performance liquid chromatography, immuno-staining or digestion with methylation sensitive enzymes. Methylation can also be estimated in repeated elements based on pyrosequencing and methylight. Gene specific analysis can be done by pyrosequencing (highly quantitative with single site resolution), methylation specific PCR (qualitative with no single site resolution, more economical), real time PCR (low resolution but very economical). Genome wide scans include microarrays and next generation sequencing.

The histone modification analysis involves initial purification and isolation of histones. This is followed by histone detection by ELISA (Enzyme linked Immunosorbent Assay) or ChIP (Chromatin Immunoprecipitation). ELISA provides a global genomic content of a certain modification whereas ChIP qPCR is a gene specific measure of a certain modification next to a specific gene. The non-coding RNA analysis techniques include candidate miRNA analysis by real time PCR, microarrays, counter Nanostring analysis and deep sequencing.

7.3 Proteomics

Studies the entire protein complement of a cell. Most RNA transcripts are translated into proteins that exert physiological or pathological effects. The proteome consists of all proteins present in a cell at a given time and is far more complex than was originally proposed by the one-gene, one-transcript, one-protein hypothesis [17]. To date, it is estimated that the approximately 24,000 human genes encode for nearly one million proteins [17]. Alternative

splicing, by which a single gene can produce multiple versions of a protein, is a significant contributor to protein diversity, occurring in 35% to 60% of our genes [18].

{Further reading: Alternative Splicing (WEB5)}

Proteins have a functional role in phenotype determination, reflecting genetic constitution along-with environmental effects. This response to external stimuli is detected in the proteome. Measurable changes in protein profiles are being used to assess disease. In the differential diagnosis of benign versus malignant prostatic disease a difference in proteomic profiles is robust enough to be used as a predictive diagnostic tool [19].

{Further reading: Proteomics (WEB6)}

7.4 Metabolomics

Is the study of metabolites in a given biological state resulting from a complex interplay between gene expression, protein product and environmental factors. The functional state of an individual at a particular time-point and in response to specific drugs/environmental stimuli is represented by the metabolome.

Metabolomic studies can lead to a better understanding of disease mechanisms, new diagnostic markers and individual variation to drug response. Initial metabolomic signatures have already been reported for several conditions, including Alzheimer's, coronary disease and ovarian/breast cancer. These signatures are made up of metabolites that are de-regulated, with modified concentrations in the disease state or after drug exposure. As a result, analysis of these signatures and their components can show mechanisms of disease pathophysiology [20].

The various molecules studied in metabolomics can be analysed by using a combination of separation and detection techniques based on individual properties of the molecule being studied.

{Further reading: Metabolomic techniques (WEB7)}

The development of these analytical platforms that are capable of accurately measuring hundreds or thousands of small molecules in biological samples promise to substantially advance our understanding of disease pathophysiology and development of disease risk biomarkers [20].

8. BIONIFORMATICS AND SYSTEM GENETICS

Vast quantities of data are generated by genomic research. A biological database is a large organised body of data usually associated with computerised software designed to update, query and retrieve components of the data stored within the system. For researchers to benefit from stored data, easy access to information and a method to extract only the information required to answer a specific biological question are essential.

The need to utilize this new information in context of the existing genetic data has lead to the development of a new set of tools. Bioinformatics uses computer sciences to integrate large data sets and to answer biological questions. Interestingly, many of the concepts of bioinformatics were developed well before the human genome project, but functional genomics technologies, the internet, and a culture of data sharing have propelled the field,

which now touches nearly all domains of biomedical research. A major bioinformatics initiative to standardize the representation of gene and gene products across species and databases has lead to the development of the Gene Ontology Project (www.geneontology.org). It provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data as well tools to access and process data. The use of computational approaches in bioinformatics provides a global perspective in experimental design and helps to capitalize on the emerging technology of database mining [21].

Data mining (referencing data from different sources and summarising it into useful information) and common bioinformatics tools have for example been used for the selection of highly specific DNA probes, eliminating the need of traditional methods which are costly and time consuming. Bioinformatic tools allow performing many investigations into the genome 'in silico' as opposed to time and cost consuming wet lab work 'in-vitro' [22].

{Further reading: Bioinformatics and Systems Biology (WEB8)}

Systems genetics seeks to understand the complexity of phenotypic variation resulting from multiple complex interactions between genetic and environmental factors. The defining principle of systems genetics is understanding how genetic information is integrated and ultimately transmitted through molecular, cellular, and physiological networks to enable higher-order functions and emergent properties of biological systems [23]. Although the goal of understanding how genetic and phenotypic variants interact to create the functional diversity of organismal biology has not changed since Mendel, the experimental and computational methods of systems genetics will finally enable us to study previously intractable problems.

9. CONCLUSION

Biomedical research has grown exponentially in the last 20 years and remarkable advances have been achieved. However, we have been unable to translate the full potential of genomic research to clinical medicine due in part to a relative lack of education about genetics and genomics amongst the general non-academic physicians. Advances in genetic knowledge and an insight into genetic variation in human populations, manifested as disease risk through various genetic, epigenetic and environmental interactions, need to become commonplace in clinical practice. Genomics, in clinical practice, can lead to development of new targets for treatment and prevention of disease as well as realise the goal of personalised medicine. Efficient use and regulation of the vast amounts of information generated for the benefit of patients requires the physicians, geneticists and biomedical researchers to work closely together and to have a mutual understanding of the challenges and opportunities.

{Further reading: Glossary (WEB9)}

{Further reading: Online Figures (WEB10)}

COMPETING INTERESTS

Authors have declared that no competing interests exist.

CONSENT

Not applicable

ETHICAL APPROVAL

Not applicable

REFERENCES

1. Cappola TP, Margulies KB. Functional genomics applied to cardiovascular medicine. Circulation. 2011;124(1):87-94.

2. Kim L, Devereux RB, Basson CT. Impact of genetic insights into Mendelian disease on cardiovascular clinical practice. Circulation. 2011;123(5):544-50.

3. UK Cystic Fibrosis Gene Therapy Consortium. Accessed June 2014. Available: http://www.cfgenetherapy.org.uk/main/news.html

4. Nowell PC, Hungerford DA. A minute chromosome in human chronic granulocytic leukemia. Science. 1960;132:1497–01.

5. Chial H. Cytogenetic methods and disease: Flow cytometry, CGH, and FISH. Nature Education. 2008;1(1).

6. Marian AJ, Belmont J. Strategic Approaches to Unraveling Genetic Causes of Cardiovascular Diseases. Circ Res. 2011;108(10):1252-69.

7. Shunckert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011;43(4):333-38.

8. Arking DE, Chakravarti A. Understanding cardiovascular disease through the lens of genome-wide association studies. Trends Genet. 2009;25(9):387-94.

9. Ellis KJ, Stouffer GA, McLeod HL, Lee CR. Clopidogrel pharmacogenomics and risk of inadequate platelet inhibition: US FDA recommendations. Pharmacogenomics. 2009;10:1799 –1817.

10. National Human Genome Research Institute. Accessed June 2014. Available http://www.genome.gov/11006943

11. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J et al. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.

12. Ye S, Willeit J, Kronenberg F, Xu Q, Kiechl S. Association of genetic variation on chromosome 9p21 with susceptibility and progression of atherosclerosis: a population-based prospective study. J. Am. Coll. Cardiol. 2008;52:378e84.

13. Talmud PJ, Cooper JA, Palmen J, Lovering R, Drenos F, Hingorani AD et al. Chromosome 9p21.3 coronary heart disease locus genotype and prospective risk of CHD in healthy middle-aged men. Clin. Chem. 2008;54:467e74.

14. Gong Y, Beitelshees AL, Cooper-DeHoff RM, Lobmeyer MT, Langaee TY, Wu J et al. Chromosome 9p21 haplotypes and prognosis in Caucasian and African American patients with coronary artery disease. Circ Cardiovasc Genet. 2011;4:169e78.

15. Zárate-Bladés CR, Silva CL, Passos GA. The impact of transcriptomics on the fight against tuberculosis: focus on biomarkers, BCG vaccination, and immunotherapy. Clin Dev Immunol. 2011;192630.doi 10.1155/2011/192630.

16. Stein RA. Epigenetics and environmental exposures. J Epidemiol Community Health. 2012;66(1):8-13.

17. Humphery-Smith I. A human proteome project with a beginning and an end. Proteomics. 2004;4(9):2519-21.

18. Modrek B, Lee C. A genomic view of alternative splicing. Nat Genet. 2002;30(1):13-19.

19. Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Research. 2002;62:3609–14.

20. Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: a global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol. 2008;48:653-83.

21. Hagen JB. The origins of bioinformatics. Nat Rev Genet. 2000;1(3):231–36.

22. Zeng H, Weier HUG, Kwan J, Wang M, O'Brien B. Data mining empowers the generation of a novel class of chromosome-specific DNA probes. J Data Mining in Genom Proteomics. 2011;2:108.

23. Nadeau JH, Dudley AM. Systems Genetics. Science. 2011;331(6020):1015-16.

APPENDIX
(ONLINE MATERIAL)

WEB1: Cytogenetics

WEB2: Genome Wide Association Studies

WEB3: Gene Sequencing

WEB4: Epigenetics

WEB5: Alternative Splicing

WEB6: Proteomics

WEB7: Metabolomic Techniques

WEB8: Bioinformatics

WEB9: Glossary

WEB10: Online Figures

Online References