DRUG DISCOVERY TARGETING GENOMIC APPROACH

A Dissertation submitted to the Department of Pharmacy, East West University, in partial fulfillment of the requirements for the degree of Master of Pharmacy.

Submitted By

Mst. Tania Ashraf

ID: 2014-3-79-006

Department Of Pharmacy

East West University

This thesis paper

is dedicated

to my beloved Parents...

DECLARATION BY THE CANDIDATE

I, Mst.Tania Ashraf, hereby declare that this dissertation, entitled “Drug discovery targeting genomic approach” submitted to the Department of Pharmacy, East West University, in the partial fulfillment of the requirement for the degree of Master of Pharmacy, is a genuine & authentic research work carried out by me under the guidance of Dr. Repon Kumer Saha, Assistant Professor, Department of Pharmacy, East West University, Dhaka. The contents of this dissertation, in full or in parts, have not been submitted to any other Institute or University for the award of any Degree or Diploma of Fellowship.

Mst. Tania Ashraf

ID: 2014-3-79-006

Department of Pharmacy

East West University

Jaharul Islam city, Aftabnagar, Dhaka

CERTIFICATION BY THE SUPERVISOR

This is to certify that the desertion, entitled “Drug discovery targeting genomic approach” is a benefice research work done, under our guidance and supervision by Mst. Tania Ashraf (ID: 2014-3-79-006), in partial fulfillment of the requirement for the degree of Master of Pharmacy.

Dr. Repon Kumer Saha

Assistant Professor

Department of Pharmacy

East West University

Jaharul Islam city, Aftabnagar, Dhaka

ENDORSEMENT BY THE CHAIRPERSON

This is to certify that the dissertation, entitled **"Drug discovery targeting proteomic approach"** is a bonafide research work done by Mst.Tania Ashraf (ID: 2014-3-79-006), in partial fulfillment of the requirements for the degree of Master of Pharmacy.

Dr. Shamsun Nahar Khan

Chairperson and Associate Professor

Department of Pharmacy

East West University

Jaharul Islam city, Aftabnagar, Dhaka

ACKNOWLEDGEMENTS

All praise is for Almighty for all the bounties granted to me and only with His guidance and help this achievement has become possible.

It is my pleasure and proud privilege to express my heartiest regards and gratitude to my respected teacher and supervisor Dr. ReponKumerSaha, Assistant Professor, Department of Pharmacy, East West University, for his expert supervision, constructive criticism, valuable advice, optimistic counseling, constant support and continuous backup and encouragement throughout every phase of the project as well as to prepare this dissertation.

I would also like to put forward my most sincere regards and profound gratitude to Dr. ShamsunNahar Khan, Chairperson& Associate Professor, Department of Pharmacy, East West University,for giving me the opportunity to conduct such an interesting project and for facilitating a smooth conduction of my study

I would also like to extend my thanks to all the research students in the lab,lab officers and other staffs of the Department of Pharmacy for their help and assistance, friendly behavior and earnest co-operation which enabled me to work in a very congenial and comfortable ambience.

I owe special thanks to my fellow research group members for their immense support and contribution in my research work.

Last but not the least, I would like to thank my family, and friends for their care and encouragement during my research work.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract

The drug discovery process is not a predefined series of steps. Modern approaches include target-based drug discovery in which researchers need to survey genetic profile like never before. Genomics, particularly high-throughput sequencing and characterization of expressed human genes, has created new opportunities for drug discovery. Knowledge of all the human genes and their functions may allow effective preventive measures, and change drug research strategy and drug discovery development processes. Pharmacogenomics is the application of genomic technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. It applies the large-scale systematic approaches of genomics to speed the discovery of drug response markers, whether they act at the level of the drug target, drug metabolism, or disease pathways. The two most important needs for this type of technology are to find more effective biomarkers for disease detection and discover gene to which therapeutic drugs can be targeted. It is well-known that the risks are high in drug discovery process and there are long timelines to be passed before it is known whether a candidate drug will succeed or fail. Making accurate decisions within an accelerated process is the key to success to the pharmaceutical companies.

Genomics revolution had a very positive impact upon these issues and now proteomics is in the field as a powerful new partner of genomics. Each step of the process from target discovery to clinical trials is accessible to genomics. Scientists are able to see every dimension of their biological focus, from genes, mRNA, proteins and their sub cellular localization. This will greatly assist our understanding of the fundamental mechanistic basis of human disease and will allow discovery of improved, speedier, less toxic and hopefully, inexpensive drugs.

Key words: Genomics, genes, mRNA, biomarkers, pharmacogenomics.

Chapter One
Introduction

1.1 Drug discovery using genomic approach

Genome research centers worldwide are engaged in the Human Genome Project (HGP) with the ultimate goal of elucidating and characterizing the complete sequence of the 3×10^9 base pairs (bp) arranged in about 85 000 genes of the human genome. An even greater task is to determine their function and interplay. The genomic approach to mapping and sequencing the genome project has accelerated the rate of gene discovery. In 1990, 1772 human genes were identified and mapped to a specific chromosome or region of the genome. In September of 1996, this number was 3868 genes—a more than two-fold increase. As of June 1996, 62 human genes linked to human diseases had been isolated by genomic technologies and, of these, 51 (82%) were available in the public domain as clones or as DNA sequences. Moreover, biomedical research is rapidly defining the molecular mechanisms of pharmacological effects, genetic determinants of disease pathogenesis, and functionally important polymorphisms in genes that govern drug metabolism and disposition. A radical new, but complementary, approach to drug development is now emerging which promises dramatic improvements in the efficiency and speed of drug development. This approach uses the emerging technological expertise from pharmacogenetics; pharmacogenomics and functional genomics to dissect predict and monitor the nature of the individual response to medications. Ultimately, this may lead to smaller and faster clinical studies and to individually tailored pharmacological treatments, in which patients are screened to identify which therapeutic option most suits their genetic and physiological makeup and accurately monitored for their response. This approach is likely to have radical consequences in the planning, conduct of clinical trials and medical treatment of diseases.

One important outgrowth of molecular medicine is the development of technologies for the transfer of therapeutic genes to cells in culture and tissues *in vivo*, with potential applications both to medical research and the practice of clinical medicine. The use of genomic databases to find new targets for drug discovery and the rapid accumulation of human gene sequences is

promising for clinical medicine, if the molecular level can be translated into improved interventions. If it can, therapeutic agents with specific molecular functions can be produced, be they gene products which are deficient or abnormal in the patients, or drugs with direct transcriptional or molecular effects. Individual genetic testing, with knowledge of disease genes, will help early diagnosis and early treatment. For example, recent advances in the genetics of complex traits (for example, diabetes, coronary heart disease and Alzheimer's disease) have to some extent reshaped disease phenotypic descriptions. The techniques developed for automated sequencing and analysis of DNA may eventually allow inexpensive screening of multiple loci for polymorphisms.

Molecular genetics techniques may also translate into gene therapy. The ability to clone and manipulate genes responsible for human disease and to re-introduce functional copies of normal genes into living cells and tissues is one such therapeutic objective. The potential clinical applications of gene therapy are numerous, and a number of specific human genetic or environmentally-induced diseases that result from a lesion in a single gene have been proposed as candidates for gene therapy (Yao SN, et al., 1991). For some of these diseases, the introduction of a functional homologue of the defective gene and the production of even small amounts of the missing gene product would have a beneficial effect; for example 10–20% production of the normal levels of factor IX can alleviate severe hemophilia B (Yao SN, et al., 1991). At the same time, over expression of the gene product would not be expected to have deleterious effects. Thus, these genetic disorders are candidates for gene therapy because the expression of the transduced gene need not be strictly regulated. In contrast, it is not always necessary to correct the genetic lesion in the cell type that shows the defect. In such cases, a therapeutic gene may be introduced into another cell type so that the genetically modified cells functionally replace the defective cell type.

Table 1: Examples of modifier genes and their phenotypic effects (Nadeau, J. H, 2001)

Target modified gene	Modifier gene	Modifier effect	Nature of the modified phenotype
Examples in mice (allele)			
disorganization (Ds)	Genetic background (C57BL/6L or C3H)	Penetrance	Whether or not an affected mouse shows

			Ds-like birth defects
Undulated ($pax1^{wt}$)	Genetic background (C57BL/6J×CBA)	Dominance modification	Presence or absence of foramina transversaria
short-ear ($Bmp5^{st}$)	Genetic background (C57BL/6J or CBA)	Dominance modification	Presence or absence of foramina transversaria
brachyury	Genetic background (selection experiment based on mixed background)	Expressivity	Tail length
Citr knockout	Cfm1	Suppressor	Suppresses meconium ileus
Apc^{mn}	Phospholpase A group IIA $(Mom1)^2$	Suppressor	Reduces polyp number
Apc^{d716}	Cox2 (knockout)	Suppressor	Reduces polyp number
Apc^{d716}	$cPLA_2$	Suppressor	Reduces polyp number
$Pax3^{sp}$	Fidgetin-fidget (Fign)	Suppressor	Suppresses spina bifida
ashen (Rab27a)	dsu	Suppressor	Coat colour suppressor
leaden (In)	dsu	Suppressor	Coat colour suppressor
ruby-eye (ru)	dsu	Suppressor	Coat colour suppressor
ruby-eye2 (ru2)	dsu	Suppressor	Coat colour suppressor
Patch (Ph)	Pax1	Novel phenotype	Spina bifida occulta in Ph/+un/un mice
Pax3sp	Curly tail (ct)	Novel phenotype	More extreme spina bifida in Sp/+ct/ct mice
Peripherin 1 (PRP1)	ROM1	Dominance Modifiation	Retinitis pigmentosa in PRP1/+heterozygotes
Familial hypercholesterolaemi (FH)	nnamed modifier on 13q	Suppressor	Reduces Cholesterol level in FH

Familial Mediterranearn fever (FMF)	Serum amyloid A (SAA)	Pleiotropy	Whether FMF Cases show renal amyloidosis
Cystic fibrosis transmembrane conductance regulator (CFTR)	Cystic fibrosis (CFM1)	Suppressor	Suppresses meconium ileus
Non-syndromic dearness (DFNB26)	At least one gene on chromosome 7	Penetrance	Suppresses deafness
Mitochondrial 12S ribosomal gene	At least one gene near D8S277	Penetrance	Suppresses deafness

Abbreviations

Apc-adenomatous polyposis coli: Cfm1- cystic fibrosis modifier 1: Cox-cyclooxygenase: cPLA

cystolic phospholipase A: dsu- dilute suppressor: pax-paired box transcription factor: ROM- rod outer segment protein.

1.2 Genomics

Genomics is a discipline in genetics that applies recombinant DNA, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyze the function and structure of genomes (the *complete* set of DNA within a single cell of an organism)(National Human Genome Research Institute, 2010; Concepts of genetics , 2012).Advances in genomics have triggered a revolution in discovery-based research to understand even the most complex biological systems such as the brain (SV, 2013). The field includes efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy and other interactions between loci and alleles within the genome (Pevsner J, 2009). In contrast, the investigation of the roles and functions of single genes is a primary focus of molecular biology or genetics and is a common topic of modern medical and biological research. Research of single genes does not fall into the definition of genomics unless the aim of this genetic, pathway, and functional information analysis is to elucidate its effect on, place in, and response to the entire genome's networks (National Human Genome Research Institute, 2010; Culver KW and Labow MA, 2002).

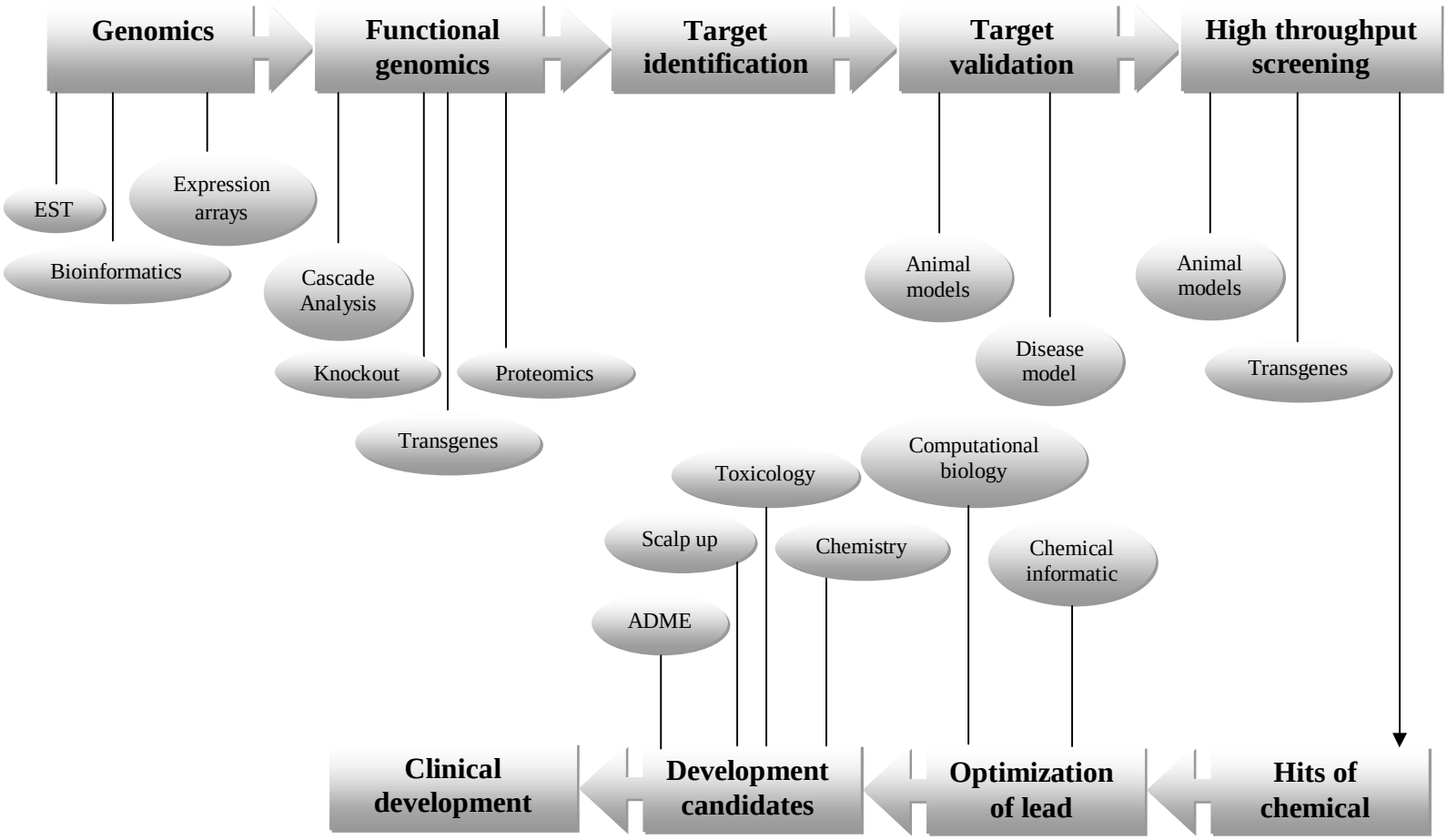


Figure 1: Illustrate the potential contribution of genomics on drug development process (Emilien G, et al., 2000)

1.2.1 Types of genomic

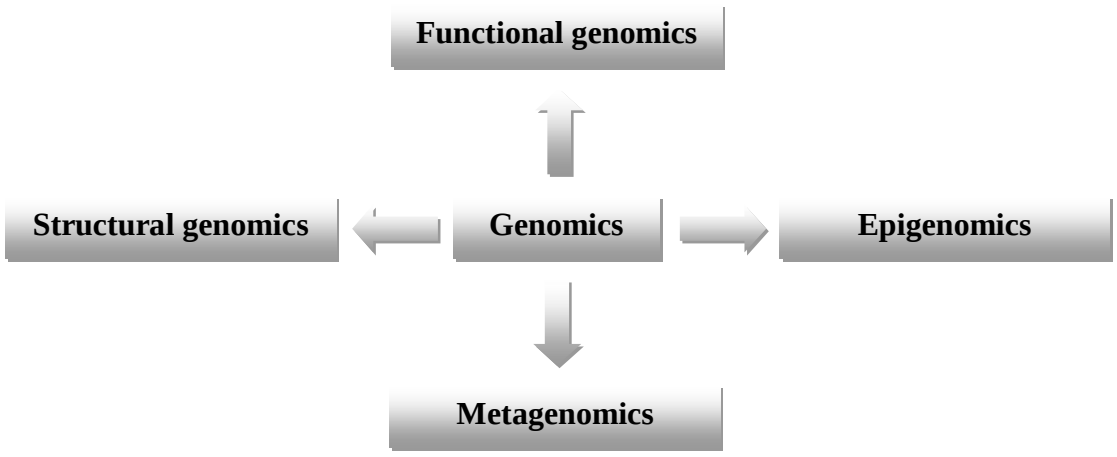


Figure 2: Illustrate the types of genomics

Functional genomic

Functional genomics is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (and protein) functions and interactions. Functional genomics focuses on the dynamic aspects such as gene transcription, translation, and protein–protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products. A key characteristic of functional genomics studies is their genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional "gene-by-gene" approach.

Structural genomic

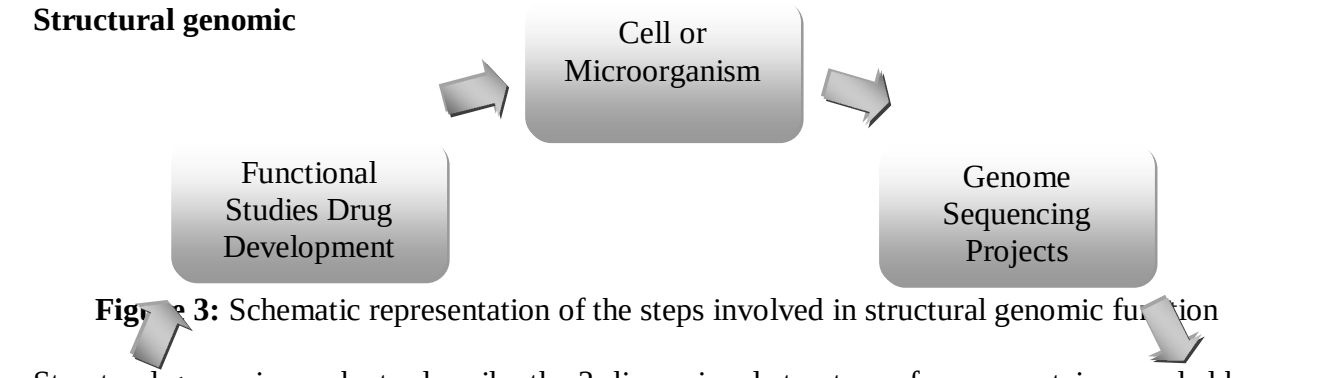


Figure 3: Schematic representation of the steps involved in structural genomic function

Structural genomics seeks to describe the 3-dimensional structure of every protein encoded by a rsden RL, et al., 2007; Brenner SE and Levitt M 2000).T r a high-throughput method of structure determination by experimental and modeling approaches. The principal difference between structural genomics and tra onal structural prediction is that structural genomics attempts to dete ne the structure of d by the genome, rather than focu r protein. With full-ge lable, structure prediction can be through a combinatio d modeling approaches, especia ability of large number of sequenced genomes ar structures allows scientists to model protein structure on the structu ved homologs. Structural genomics involves taking a large number of approaches to structure determination, including experimental methods using genomic sequences or modeling-based approaches based on sequence or structural homology to a protein of known structure or based on chemical and physical principles for a protein with no homology to any known structure. As opposed to traditional structural

biology, the determination of a protein structure through a structural genomics effort often (but not always) comes before anything is known regarding the protein function. This raises new challenges in structural bioinformatics, i.e. determining protein function from its 3D structure (Brenner SE, 2001).

Epigenomics

Epigenomics is the study of the complete set of epigenetic modifications on the genetic material of a cell, known as the epigenomic (Francis RC, 2011). Epigenetic modifications are reversible modifications on a cell's DNA or histones that affect gene expression without altering the DNA sequence (Russell 2010 p. 475). Two of the most characterized epigenetic modifications are DNA methylation and histone modification. Epigenetic modifications play an important role in gene expression and regulation, and are involved in numerous cellular processes such as in differentiation/development and tumorigenesis (Francis RC, 2011). The study of epigenetic on a global level has been made possible only recently through the adaptation of genomic high-throughput assays (Laird PW, 2010).

Metagenomics

Metagenomics is the study of metagenomes, genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics. While traditional microbiology and microbial genome sequencing rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes (often the 16S rRNA gene) to produce a profile of diversity in a natural sample. Such work revealed that the vast majority of microbial biodiversity had been missed by cultivation-based methods (Hugenholtz P, et al., 1998).Recent studies use "shotgun" Sanger sequencing or massively parallel pyrosequencing to get largely unbiased samples of all genes from all the members of the sampled communities (Eisen JA, 2007).Because of its power to reveal the previously hidden diversity of microscopic life, metagenomics offers a powerful lens for viewing the microbial world that has the potential to revolutionize understanding of the entire living world (Marco D, 2010; Marco D, et al., 2011).

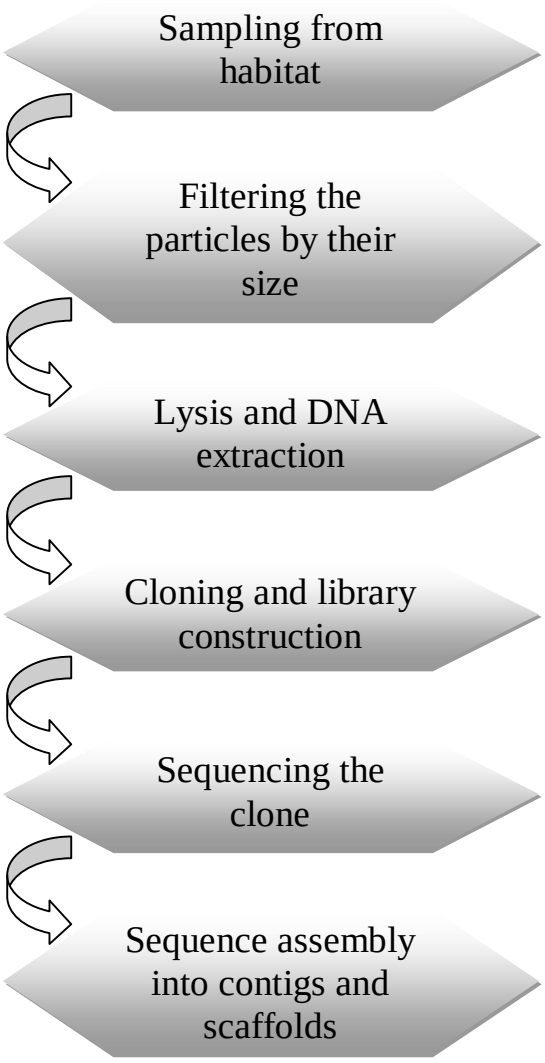


Figure 4: Schematic representation of the metagenomic process

Environmental Shotgun Sequencing (ESS) is a key technique in metagenomics

1.3 The development of genomics: a brief history

From the Greek ΓΕΝ (George H, et al., 2015) *gen*, "gene" (gamma, epsilon, nu, epsilon) meaning "become, create, creation, birth", and subsequent variants: genealogy, genesis, genetics, genic, genomere, genotype, genus etc. While the word *genome* (from the German *Genom*, attributed to Hans Winkler) was in use in English as early as 1926, (Genome n, 2008) the term *genomics* was coined by Tom Roderick, a geneticist at the Jackson Laboratory (Bar Harbor, Maine), over beer at a meeting held in Maryland on the mapping of the human genome in 1986 (Yadav SP, 2007.

Early sequencing efforts

Following Rosalind Franklin's confirmation of the helical structure of DNA, James D. Watson and Francis Crick's publication of the structure of DNA in 1953 and Fred Sanger's publication of the Amino acid sequence of insulin in 1955, nucleic acid sequencing became a major target of early molecular biologists(Ankeny RA,2003).In 1964, Robert W. Holley and colleagues published the first nucleic acid sequence ever determined, the ribonucleotide sequence of alanine transfer RNA(Holley RW, et al., 1965; Holley RW, et al., 1965). Extending this work, Marshall Nirenberg and Philip Leder revealed the triplet nature of the genetic code and were able to determine the sequences of 54 out of 64 codons in their experiments (Nirenberg M, et al., 1965).In 1972, Walter Fiers and his team at the Laboratory of Molecular Biology of the University of Ghent (Ghent, Belgium) were the first to determine the sequence of a gene: the gene for Bacteriophage MS2 coat protein (Min Jou W et al., 1972).Fiers' group expanded on their MS2 coat protein work, determining the complete nucleotide-sequence of bacteriophage MS2-RNA (whose genome encodes just four genes in 3569 base pairs [bp]) and Simian virus 40 in 1976 and 1978, respectively (Fiers W, et al., 1976; Tamarin RH, 2004).

DNA sequencing technology developed

Frederick Sanger Walter Gilbert

Frederick Sanger and Walter Gilbert shared half of the 1980 Nobel Prize in chemistry for independently developing methods for the sequencing of DNA.

In addition to his seminal work on the amino acid sequence of insulin, Frederick Sanger and his colleagues played a key role in the development of DNA sequencing techniques that enabled the establishment of comprehensive genome sequencing projects (Pevsner J, 2009).In 1975, he and Alan Coulson published a sequencing procedure using DNA polymerase with radiolabelled nucleotides that he called the *Plus and Minus technique* (Tamarin RH, 2004; Sanger F, 1980).

This involved two closely related methods that generated short oligonucleotides with defined 3' termini. These could be fractionated by electrophoresis on a polyacrylamide gel and visualised using autoradiography. The procedure could sequence up to 80 nucleotides in one go and was a big improvement, but was still very laborious. Nevertheless, in 1977 his group was able to sequence most of the 5,386 nucleotides of the single-stranded bacteriophage φX174, completing the first fully sequenced DNA-based genome (Sanger F, et al., 1977).The refinement of the *Plus and Minus* method resulted in the chain-termination, or Sanger method (see below), which formed the basis of the techniques of DNA sequencing, genome mapping, data storage, and bioinformatic analysis most widely used in the following quarter-century of research(Kaiser O, et al., 2003; Sanger F, et al., 1977) . In the same year Walter Gilbert and Allan Maxam of Harvard University independently developed the Maxam-Gilbert method (also known as the *chemical method*) of DNA sequencing, involving the preferential cleavage of DNA at known bases, a less efficient method (Maxam AM and Gilbert W, 1977; Darden L and James Tabery, 2010). For their groundbreaking work in the sequencing of nucleic acids, Gilbert and Sanger shared half the 1980 Nobel Prize in chemistry with Paul Berg (recombinant DNA).

Complete genomes

The advent of these technologies resulted in a rapid intensification in the scope and speed of completion of genome sequencing projects. The first complete genome sequence of an eukaryotic organelle, the human mitochondrion (16,568 bp, about 16.6 kb [kilobase]), was reported in 1981,(Anderson S, et al., 1981) and the first chloroplast genomes followed in 1986(Shinozaki K, et ai., 1986; Ohyama K, et ai., 1986). In 1992, the first eukaryotic chromosome, chromosome III of brewer's yeast Saccharomyces cerevisiae (315 kb) was sequenced (van der Aart QJ, et al., 1992). The first free-living organism to be sequenced was that of Haemophilus influenzae (1.8 Mb [megabase]) in 1995(Fleischmann RD, et al., 1995). The following year a consortium of researchers from laboratories across North America, Europe, and Japan announced the completion of the first complete genome sequence of a eukaryote, S. cerevisiae (12.1 Mb), and since then genomes have continued being sequenced at an exponentially growing pace(Goffeau A, et al., 1996). As of October 2011[update], the complete sequences are available for: 2,719 viruses, 1,115 archaea and bacteria, and 36 eukaryotes, of which about half are fungi (NCBI, 2011; Genome Project Statistics, 2011).

Most of the microorganisms whose genomes have been completely sequenced are problematic pathogens, such as *Haemophilus influenzae*, which has resulted in a pronounced bias in their

phylogenetic distribution compared to the breadth of microbial diversity Zimmer C, 2009; Wu D, 2009) Of the other sequenced species, most were chosen because they were well-studied model organisms or promised to become good models. Yeast (*Saccharomyces cerevisiae*) has long been an important model organism for the eukaryotic cell, while the fruit fly *Drosophila melanogaster* has been a very important tool (notably in early pre-molecular genetics). The worm *Caenorhabditis elegans* is an often used simple model for multicellular organisms. The zebrafish *Brachydanio rerio* is used for many developmental studies on the molecular level, and the flower *Arabidopsis thaliana* is a model organism for flowering plants. The Japanese pufferfish (*Takifugu rubripes*) and the spotted green pufferfish (*Tetraodon nigroviridis*) are interesting because of their small and compact genomes, which contain very little non-coding DNA compared to most species (Yue GH, et al., 2006;).The mammals dog (*Canis familiaris*), National Human Genome Research Institute, 2004),brown rat (*Rattus norvegicus*), mouse (*Mus musculus*), and chimpanzee (*Pan troglodytes*) are all important model animals in medical research (Darden L, James Tabery, 2010).

A rough draft of the human genome was completed by the Human Genome Project in early 2001, creating much fanfare.(McElheny V, 2010)This project, completed in 2003, sequenced the entire genome for one specific person, and by 2007 this sequence was declared "finished" (less than one error in 20,000 bases and all chromosomes assembled). (McElheny V, 2010) In the years since then, the genomes of many other individuals have been sequenced, partly under the auspices of the 1000 Genomes Project, which announced the sequencing of 1,092 genomes in October 2012.(Abecasis GR, et al., 2012) Completion of this project was made possible by the development of dramatically more efficient sequencing technologies and required the commitment of significant bioinformatics resources from a large international collaboration. (Nielsen R, 2010) The continued analysis of human genomic data has profound political and social repercussions for human societies. (Barnes B and Dupré J, 2008)

The "omics" revolution

The English-language neologism omics informally refers to a field of study in biology ending in -omics, such as genomics, proteomics or metabolomics. The related suffix -ome is used to address the objects of study of such fields, such as the genome, proteome or metabolome respectively. The suffix -ome as used in molecular biology refers to a totality of some sort; similarly omics has come to refer generally to the study of large, comprehensive biological data

sets. While the growth in the use of the term has led some scientists (Jonathan Eisen, among others (Eisen JA, 2012) to claim that it has been oversold, (Hotz RL, 2012) it reflects the change in orientation towards the quantitative analysis of complete or near-complete assortment of all the constituents of a system.(Scudellari and Megan, 2011) In the study of symbioses, for example, researchers which were once limited to the study of a single gene product can now simultaneously compare the total complement of several types of biological molecules. (Chaston J and Douglas AE, 2012; McCutcheon JP and von Dohlen CD, 2011)

Chapter Two

Technologies

2.1 Genomic technologies

High-throughput genomic technologies have the potential to have a major impact on preclinical and clinical drug development and the selection and stratification of patients in clinical trials. These technologies, which are at varying stages of commercialization, include array-based comparative genomic hybridization, single-nucleotide polymorphism arrays, and (the most mature example) expression-based arrays. One of the rate-limiting steps in the routine clinical application of expression array-based technology is the need for suitable clinical samples. One of the major challenges moving forward, therefore, relates to the ability to use formalin-fixed, paraffin-embedded–derived tissue in expression profiling-based approaches.

Although proteomics is fast evolving, only genomic platforms are currently ready for high throughput analysis. This paper therefore focuses on genomics and, more specifically, on the impact of gene expression microarrays and their use in preclinical and clinical drug development. This spans the entire process from the identification of novel targets to the selection of patients most likely to benefit from molecules designed to hit those targets. It is hoped that microarray technologies will both speed drug development and improve the rate of success.

Table 2: Common proteomic technologies, and their applications, limitations, strength.

Technology	Application	Strength	Limitation
Array-Based CGH	Measure alterations in copy number across the entire genome These can be large genomic clones such as bacterial artificial chromosomes (BACs) and phage artificial chromosomes (PACs) and now also oligonucleotides (Carvalho B, 2004).	Array-based CGH is far more quantitative than the metaphase chromosome spreads that preceded it and also has superior resolution and dynamic range.	Cost-effective and it is unable to detect abnormalities (such as translocations) that do not result in changes in copy number.
SNP Arrays	Used to map regions of the genome linked to defined phenotypes	The technology is based on the immobilization of	The human genome contains approximately 10

	(Khan AS, 2004*;* Janne *PA, 2004;* Kallioniemi *OP;* Anzick *SL and Trent JM, 2002)*	oligonucleotide probes corresponding to both alleles of a specific SNP.	million SNPs (Twyman RM, 2004). The largest array in prospect will contain only one million. Lack of reliable bioinformatic tools and SNP arrays are more expensive than their CGH counterparts.
Microarrays	Microarrays provide the necessary genome-wide view of expression changes, although they do not measure entire transcriptomes.	The arrays measure changes in RNA (rather than DNA) expression.	That standardization across platforms is not good.
Epigenetics	Application of these methods led to the first full chromosome methylation analysis of three human chromosomes and represented a milestone in the analysis of epigenetic change (Beck S and Rakyan V, 2008)	Change in controlling the expression of genes has recently moved to the forefront of genomic analysis with the development of high throughput sequencing methods directed to the recognition of epigenomic change, in particular Alteration of DNA methylation patterns (Rotger M, et Al., 2007).	Global patterns of DNA methylation change on genes will be reflected in their gene expression profiles
Transcriptional genomics	The most widely applied of the genomics methods in drug development has been the analysis of gene transcription (mRNA levels) (Lockhart D and	Hybridization is detected and quantities by measuring the degree of fluorescence associated with each target (Lockhart D and Winzeler E, 2000;	There is often debate amongst genomic practitioners about the right approach to data mining, a compound that requires metabolic activation

	Winzeler E, 2000)	Schena M, et al., 1998)	for an effect may not map correctly when the gene expression profile is generated in a non-metabolically
mRNA splicing	Screen for molecules that modulate alternative splicing, a recent example of which is digoxin (Fehlbaum P, et al., 20	Splice variant of a gene is associated with a specific or particular disease then there is the potential to develop a Molecule to target the variant protein.	Polymorphic profiles in splice variants can give rise to differential pharmacokinetics which can cause adverse drug reactions
Antisense and RNAi	Particular application in the development of direct mRNA targeting therapeutic molecules, not just for the target identification but also for the recognition of off-target, or downstream mRNA effects	Identification of a transcript closely associated with a disease represents the possibility of a druggable target, and connectivity analysis the possibility of identifying a small molecule That may display the requisite pharmacological activity for the target (Corey D, 2007).	This mechanism can be saturated *in vivo* preventing the normal processing of miRNA (the Physiological equivalent of RNAi) species with undesirable consequences (Grimm D, et al., 2006).
miRNA and mRNA translation analysis in drug development	Genomic methods similar to those used for transcriptomic analysis can be used to identify differentially expressed miRNA species in a cell or tissue	These small RNA species are transcribed from the genome in the same manner as protein coding genes and are under the control of many of the same transcriptional regulatory mechanisms (Lee Y, et al., 2004)	These miRNA species may be involved with the normal physiological suppression of genes involved in the inflammatory process, and that if replaced may act to reduce the inflammation in such conditions by suppressing translation of the

			inflammatory genes (Sonkoly E, et al., 2008).
Translationalomics	Using transcriptomics to verify any off-target effects of a miRNA therapy is not as simple as that for siRNA because most miRNA species suppress translation and so do not cause Degradation of the mRNA.	It is possible to use such methods to identify both on, and off, target mRNA transcripts undergoing translational regulation in response to the use of miRNAs or Antagomirs as drugs.	Is not as simple as that for siRNA because most miRNA species suppress translation and so do not cause Degradation of the mRNA.

For many years, three major genomic technologies have been available: comparative genomic hybridization (array-based CGH), single-nucleotide polymorphism analysis (SNP arrays), and gene expression analysis (microarrays), which emerged from the Northern blot. Over the past few years, all three have been adapted for high throughput analysis.

2.1.1 Array-Based CGH

Given the capacity for high throughput, array-based CGH can now measure alterations in copy number across the entire genome. The assay is based on the comparative hybridization of differently labeled samples (typically normal tissue and tumor) on an array of genomic clones. These can be large genomic clones such as bacterial artificial chromosomes (BACs) and phage artificial chromosomes (PACs) and now also oligonucleotides (Carvalho B, 2004).

The technique detects regions of the genome that are either amplified or deleted, in the hope that these represent over expression of oncogenes or loss of tumor suppressor genes and that these gains or losses can be correlated with different disease settings. Array-based CGH is far more quantitative than the metaphase chromosome spreads that preceded it and also has superior resolution and dynamic range. The trend is now toward using oligonucleotide arrays, and these are proving a more reproducible and adaptable platform than the previous BAC and PAC arrays (Carvalho B, 2004).

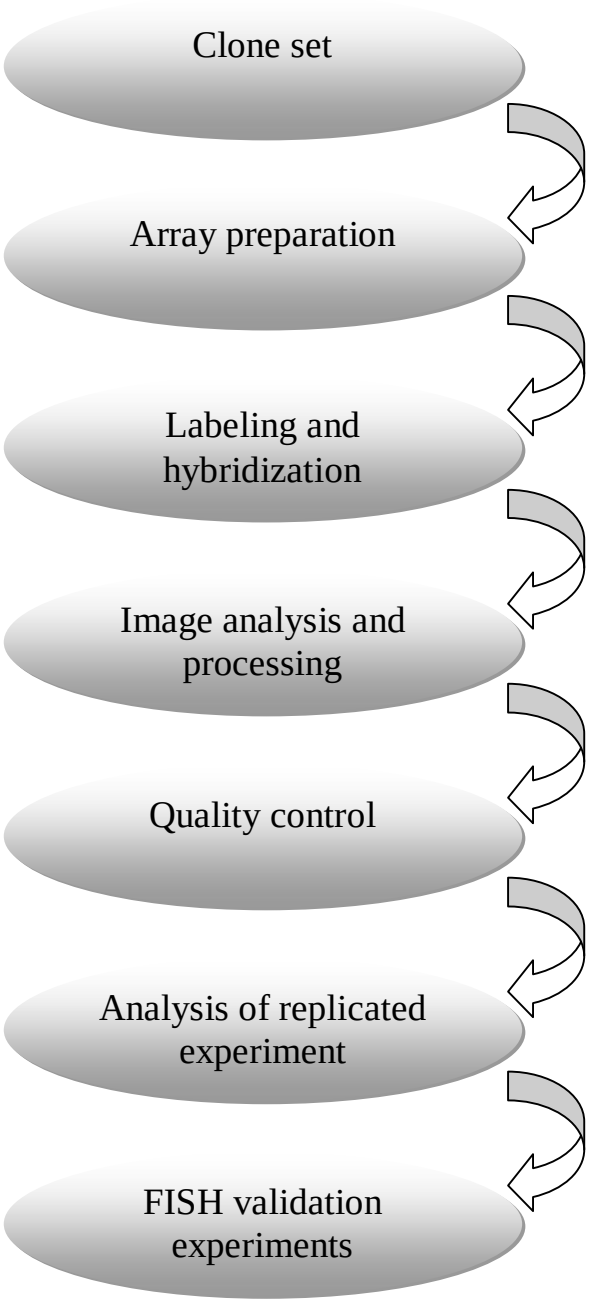


Figure 6: Schematic representation of steps involved in array-based CGH

However, while array-based CGH is reproducible, quick, based on genomic DNA (which has the advantage of stability), and relatively cost-effective, it is unable to detect abnormalities (such as translocations) that do not result in changes in copy number. Nor can it detect single copy changes, so it is not useful in investigating loss of heterozygosity in chromosome 15q, for example. Until recently, no commercially validated platforms have been available, and there is still a dearth of standardized software for data analysis. Array CGH is the process by which gene copy number variations (CNVs) in DNA are identified on a whole or partial genome basis (Barrett MT, et al., 2007; Pinkel D, et al., 1998). Hybridized to the array is genomic DNA isolated from either fresh tissue or frequently now from archived samples. The degree of Resolution for genomic change is dependent on the length of the probes on the microarray. Short probes spaced close together give the greatest resolution while longer probes such as those from bacterial artificial chromosome (BACS) give greater coverage (Pinkel D and Albertson D, 2005). The technique has been extensively employed in the identification of genomic change associated with tumor formation and progression, for example in breast carcinomas (Hyman E, et al., 2002; Pollack JR, et al., 2002). Furthermore, the method has been utilized to map the extent of human CNVs that may contribute to adverse drug reactions. What is the value of Array CGH in drug discovery? First is the applicability of the method to archived tissue. Many valuable human studies are stored as archived paraffin blocks and extracting usable mRNA from these blocks for an expression analysis is difficult. Even when mRNA extraction is possible it is never an absolute certainty that the composition of the complex mRNA has not been altered by selective degradation, which will give rise to variable and false positive results. DNA is however a stable molecule and relatively easily extracted from archived paraffin block samples (Joosse S, et al., 2007). This gives rise to the possibility of screening such samples for genomic changes that may indicate new drug targets. The expression of genes of interest from such an analysis can be verified later using fresh tissue using fewer samples than might otherwise have been required. A potential application of Array CGH in drug development is for the recognition of aneugens and clastogens. Both types of chemical can give rise to numerical changes in gene numbers on affected chromosome(s). One example of a chemical giving rise to a change in copy number is 4NQO, which increases the copy number of the SV40 gene in immortalized CO631 cells (Fahrig F and (Steinkamp-Zucht A, 1996). While this study was not conducted using an array analysis it indicates the potential applicability of Array CGH in this aspect of drug development. Recent studies employing the Array CGH method have shown that about 12% of the human genome carries CNVs(Redon R, et al., 2006), which could lead to differences in gene expression and

potentially phenotype. Several disease states (Beckmann J, et al., 2007) and acquired phenotypes such as drug resistance in cancer (Jones P and Baylin S, 2007; Ghoshal K and Bai S 2007) are linked to CNV. CNVs have also been found in inbred laboratory mice (Turton NJ, 2001). The effect of copy number and subsequent gene dosage in mediating adverse drug reactions has not been investigated to any great extent though one partial deletion in Cyp2b6 has been described (Egan C, et al., 2007). These data suggest that CNV alterations in genes responsible for metabolism of drugs could be responsible for altering pharmacokinetics, efficacy and may be partially or wholly responsible for some adverse drug reactions. This remains a substantially underexplored area but appears of critical importance both for drug development and usage in terms of personalized medicine.

2.1.2 SNP Arrays

These arrays are used to map regions of the genome linked to defined phenotypes (Khan AS, 2004; Janne PA, 2004; Kallioniemi OP; Anzick SL and Trent JM, 2002).The technology is based on the immobilization of oligonucleotide probes corresponding to both alleles of a specific SNP. Hybridizing a target sequence (from tumor or normal tissue) to the array can then determine if a specific SNP is AA, AB, or BB; and array-based profiling is now able to interrogate thousands of SNPs in a single experiment.

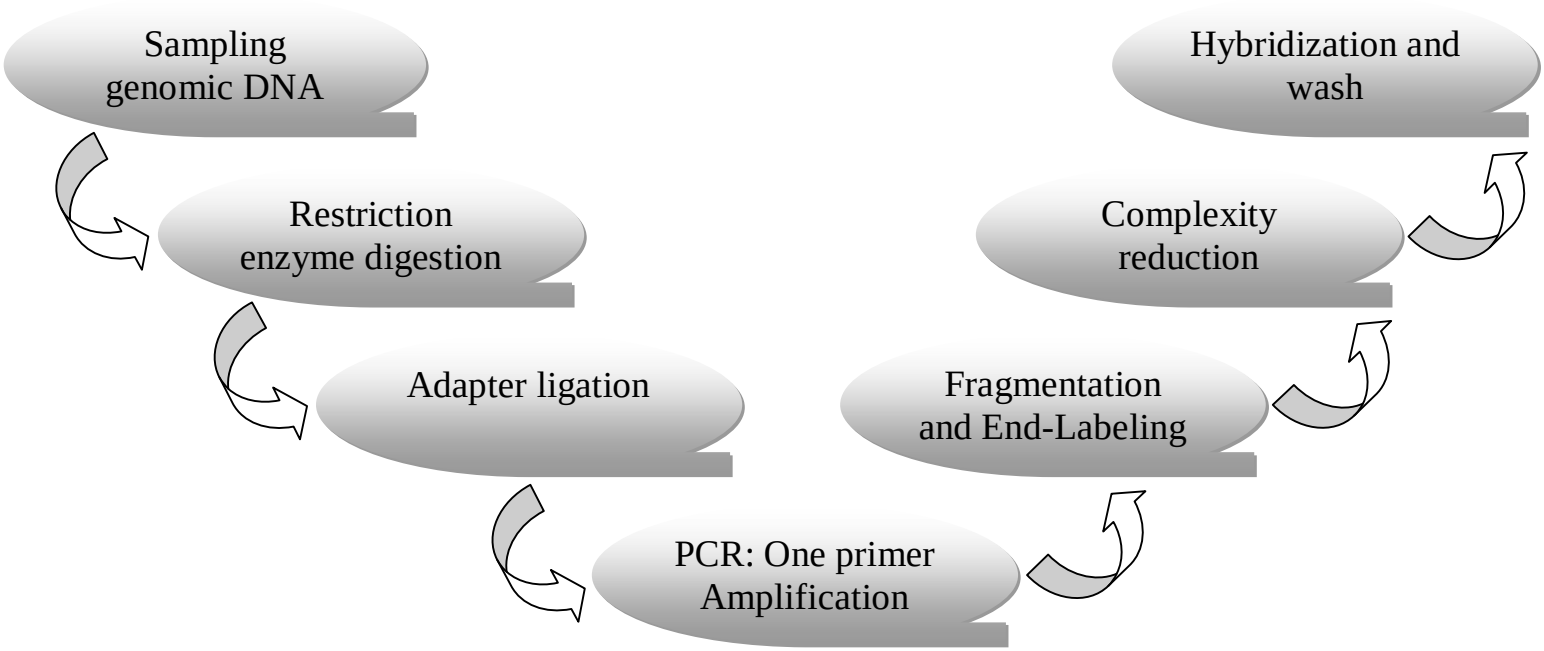


Figure 7: Schematic of the steps involved in preparing genomic DNA for Affymetrix array-based SNP analysis. Image courtesy of Affymetrix, Inc. Abbreviations: PCR, polymerase chain reaction; RE, restriction enzyme; SNP, single-nucleotide polymorphism.

As with array-based CGH, the technology is based on analysis of genomic DNA, and the protocol is relatively straightforward. However, the human genome contains approximately 10 million SNPs (Twyman RM, 2004). The largest array in prospect will contain only one million. The problem could be overcome by looking for SNPs that fall within regions of the genome which are conserved within haplotypes, assuming that one or two SNPs in that region will be representative of the haplotype block. There is again a lack of reliable bioinformatic tools, and SNP arrays are more expensive than their CGH counterparts.

2.1.3 Microarrays

Gene expression microarrays represent the most mature example of high throughput technology and have been validated in many basic and applied clinical studies across a range of tumor types (*Tai* IT, et al., 2005; Farmer P, et al., 2005; Workman P, 2004; Greiner TC, 2004; Adib TR, et. Al., 2004 Workman P, 2003). Within colorectal cancer, this is illustrated by the retrospective study of Wang et al., who identified a 23-gene signature that predicted recurrence of Dukes' B tumors (Wang Y, et al., 2004). The signature was validated in 36 patients and had an overall performance accuracy of 78%. The study demonstrated that it was possible to identify a poor-prognosis group of Dukes' B patients who could benefit from adjuvant chemotherapy.

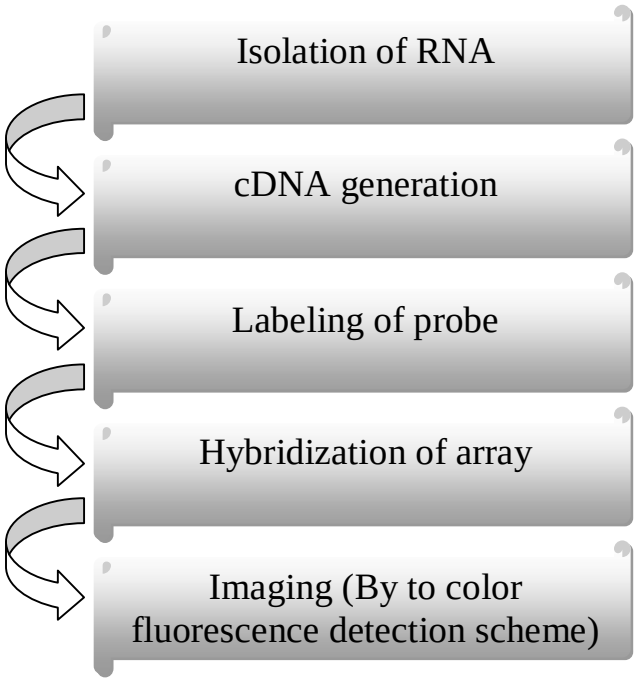


Figure 8: Steps in the microarray process

The microarray is scanned with a laser beam, first at one wavelength to collect fluorescence data representing one probe, and then is scanned at a second wavelength to collect data representing

the second probe. A computer compares the amount of fluorescence at each spot on the microarray for each probe. Through the use of computer software, the ratio of fluorescence is obtained and correlated with the clone address so the investigator knows which gene (spot on the slide) was expressed more in treated tissue as compared to control tissue.

The available platforms are all reliable (with Affymetrix [Santa Clara, CA] the market leader, other competing platforms include Agilent Technologies, Inc. [Palo Alto, CA] and Amersham [Little Chalfont, Buckinghamshire, U.K.]) Kreiner T and Buck KT, 2005; Hardiman G, 2004; Harkin DP, 2000), but their different content means that standardization across platforms is not good. Researchers are best advised to choose a platform and then stick with it.

We have seen the development of the international MIAMI (minimum information about a microarray experiment) guidelines, which should ensure that data generated in one laboratory can be duplicated in another, provided that the same platform is used. Within a platform, data analysis software is now reliable.

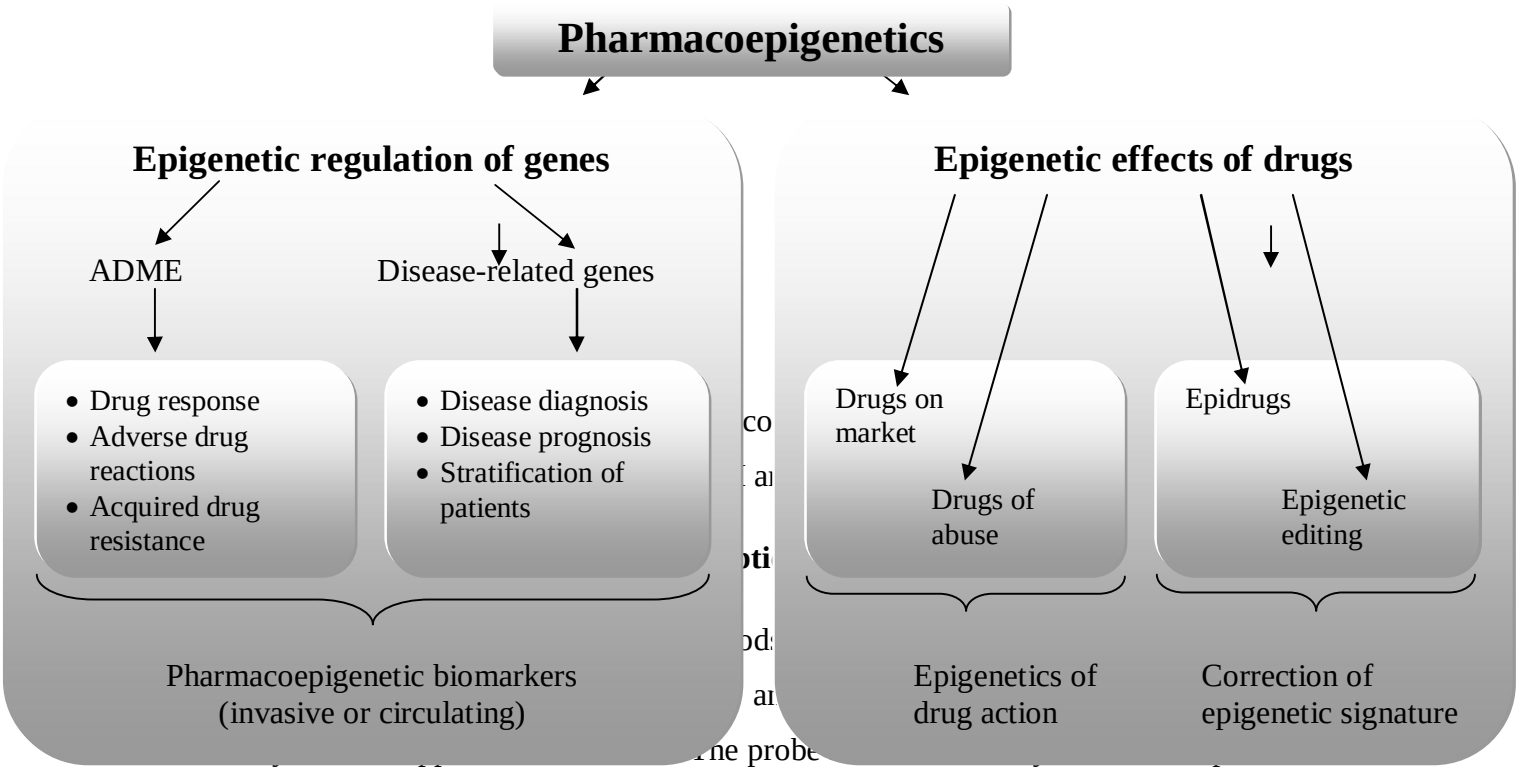
The arrays measure changes in RNA (rather than DNA) expression, and RNA is notoriously unstable. However, microarrays provide the necessary genome-wide view of expression changes, although they do not measure entire transcriptomes.

Concerns center on the availability and quality of clinical samples used for expression profiling. Traditionally, samples are fresh frozen in liquid nitrogen. However, a new preservative, “RNA Later,” is excellent at preserving RNA quality.

2.1.4 Epigenetics

The importance of epigenetic change in controlling the expression of genes has recently moved to the forefront of genomic analysis with the development of high throughput sequencing methods directed to the recognition of epigenomic change, in particular alteration of DNA methylation patterns(Rotger M, et Al., 2007). Application of these methods led to the first full chromosome methylation analysis of three human chromosomes and represented a milestone in the analysis of epigenetic change (Beck S and Rakyan V, 2008). On an array based format the technique of methyl cytosine (MeDIP) immunopreciptation coupled with analysis using a promoter region microarray allows the identification of those regions of DNA that have undergone methylation change in response to a altered state or drug(Eckhardt F, et l., 2006). However if changes in specific genes are of interest then direct sequencing combined with

bisulfite treatment is probably a better option(Jacinto F, et al., 2008). Global patterns of DNA methylation change on genes will be reflected in their gene expression profiles where transcriptomics has a major role to play in assessing the magnitude and type of these effects. Methylation changes are associated with several disease states and are frequently altered during cancer progression(Schilling E and Rehli M, 2007; Sawan C, et al., 2008).DNA methylases mediating DNA methylation therefore make promising drug targets and several small molecule inhibitors have been developed (Jones P and Baylin S, 2007; Ghoshal K and Bai S, 2007; Egger G, 2004)



of cRNA or cDNA derived from mRNA that is hybridized to an array containing the targets. Hybridization is detected and quantities by measuring the degree of fluorescence associated with

2000; Schena

data has typically taken one of three protocols. First global analysis for profiles indicative of an event using multivariate mathematical techniques such as clustering or principal component analysis, or other network or self learning type approaches. Second, for single differentially expressed genes whose altered expression in a disease state can act both as biomarker of the disease and as a prognostic or pathological indicator. The protein products of these genes may offer the prospect of a new 'druggable' target. Finally pathway analysis where genes are treated in groups according to their roles in established biochemical pathways. There is overlap from the

first analysis to the second, as multivariate approaches can successfully identify single genes which are discriminatory and potentially 'druggable'. This approach requires the application of rigorous statistical tests to avoid false positives (Zhang S-D and Gant T, 2014). All these methods have the potential to deliver new drug targets through the exploration of differences in gene expression, for example a normal cell and one that is transformed. This is the process of target identification.

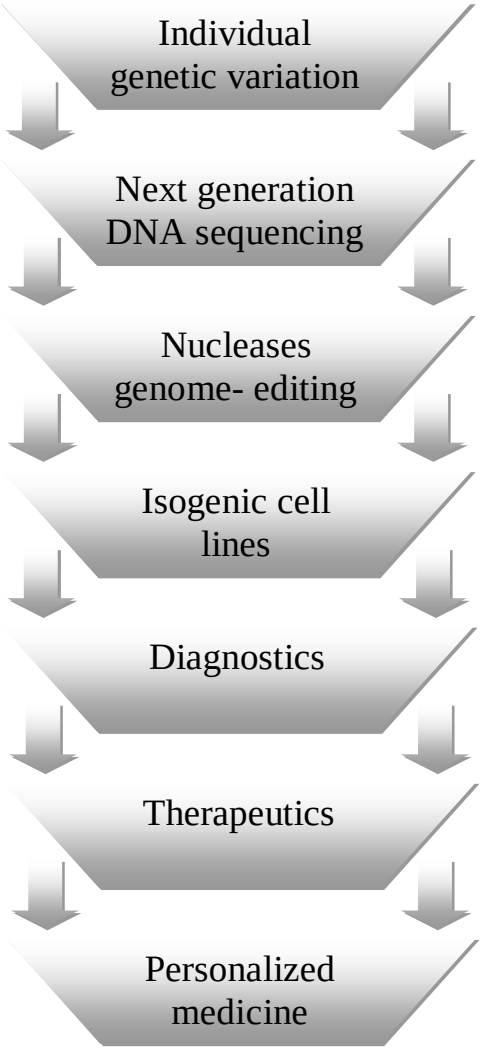


Figure 10: Translation genetic information into personalized medicine

One such example of target identification is OGN (osteoglycin). Using transcriptomic analysis OGN expression was recently shown to be correlated with left ventricular cardiac mass (LVM). Increased LVM leads to cardiac dysfunction. The close association of OGM expression with LVM identifies it as a potentially druggable target for LVM therapy. The association of OGN with LVM relied on an inherited trait in a model (the rat) which could be accurately measured and then developing an expression quantitative trait locus (eQTL) whose significance of association with the phenotype could be ascertained (Petretto E, et al., 2008). The above example

illustrates how identification of a potentially druggable target in a disease can be undertaken. However can transcriptomics also be used to actually identify molecules that may have potential as drugs? Traditional drug development against a target such as OGN takes the form of a cell line being developed with expression of high levels of the target of interest usually coupled to a reporter system or assay. Compound libraries are applied, and molecules with pharmacological activity identified by the read out from the reporter. This method however has a fundamental disadvantage in drug discovery; it is a single target screening approach. If it is subsequently decided that another target is of interest then screening has to take place again utilizing the new target. A more rational approach would be a generic screen utilizing a non-hypothesis driven approach with the potential of identifying different pharmacologies. In 2006 Lamb *et al* (Lamb J, et al., 2006) described such a method where gene expression profiles could be used to link pharmacological profiles to recognize novel pharmacology. This is known as the connectivity map and has been reviewed by Micknich (Micknich S, 2006). Screening takes place in cell lines that have not usually been engineered to over express any form of drug target. Gene expression profiles in response to the chemical are derived at multiple concentrations and then ordered by up and down regulated genes. These are then matched to a database and commonality is indicated by the number of matched genes. Lamb *et al* (Lamb J, et al., 2006; Zhang S and Gant T, 2008) showed the applicability of this method to a number of drug molecules. The method of analysis was reviewed and modified by Zhang and Gant (Wang G and Cooper T, 2007) to provide more statistical rigor. Recently the database of compounds has been increased substantially by Lamb et al. The net result of these efforts is the provision of are source that allows non-hypothesis driven drug discovery and therefore has the potential to recognize novel pharmacology, or off target pharmacology in existing molecules. Zhang and Gant (Wang G and Cooper T, 2007) tested the method on the estrogen receptor modulators and HDAC inhibitors using the data originally collected by Lamb and new data collected from public sources as the query signature. The new data was successfully used to identify not only estrogen receptor agonists from source data but also antagonists by virtue of negatively correlated gene expression maps. This can be argued to be a greater achievement than that with the estrogen receptor active compounds since the alteration of gene expression by HDAC inhibitors may potentially be more generic than that with estrogen receptor modulators, and therefore difficult to identify. This testing shows the potential of the connectivity mapping for drug discovery. Similar methods have the potential to recognize adverse toxicology in new chemical entities (NCE). For example, it is possible that genotoxins may produce a common gene expression signature that can be used to

recognize this particular activity in a NCE. However, the problem of generic signatures associated with cellular toxicity may render this analysis not possible. Thus connectivity analysis may only work acceptably when operated within an environment of a compound/nuclear receptor interaction where a specific set of genes under the control of the transcription factor recognition site will be activated. These hypotheses remain to be explored, but initial data derived by Zhang (not published) shows that the method has applicability. There is also a question of whether the biological system will allow the correct matching. For example, a compound that requires metabolic activation for an effect may not map correctly when the gene expression profile is generated in a non-metabolically competent cell line. This is likely to be a problem more relevant to the recognition of toxicology using this method than recognition of pharmacology where activity is more often a property of the parent molecule. The limits of the methods will only be discovered with extensive further testing.

2.1.6 mRNA splicing

Alternated RNA transcript splicing is a common feature of disease states and results in altered proteins (Fehlbaum P, et al., 2005). Differential splicing can, with care, be detected using microarrays utilizing the same methods as for transcriptomics (Fig-11). The difference is that the microarray will contain multiple probes for each genes directed against the exons known or hypothesized to be differentially spliced (Stoilov P, et al.,). There are two major applications of such technology in drug discovery and development. If a particular splice variant of a gene is associated with a specific or particular disease then there is the potential to develop a molecule to target the variant protein. Given that virtually every gene has the potential to be alternatively spliced into multiple forms this area of disease biology greatly expands the portfolio of potential drug targets. However a great deal of further analysis utilizing the methods above needs to be carried out to find and characterize those alternatively spliced mRNAs suitable as targets. This approach could also be used to screen for molecules that modulate alternative splicing, a recent example of which is digoxin (Fehlbaum P, et al., 2005). Furthermore polymorphic profiles in splice variants can give rise to differential pharmacokinetics which can cause adverse drug reactions.

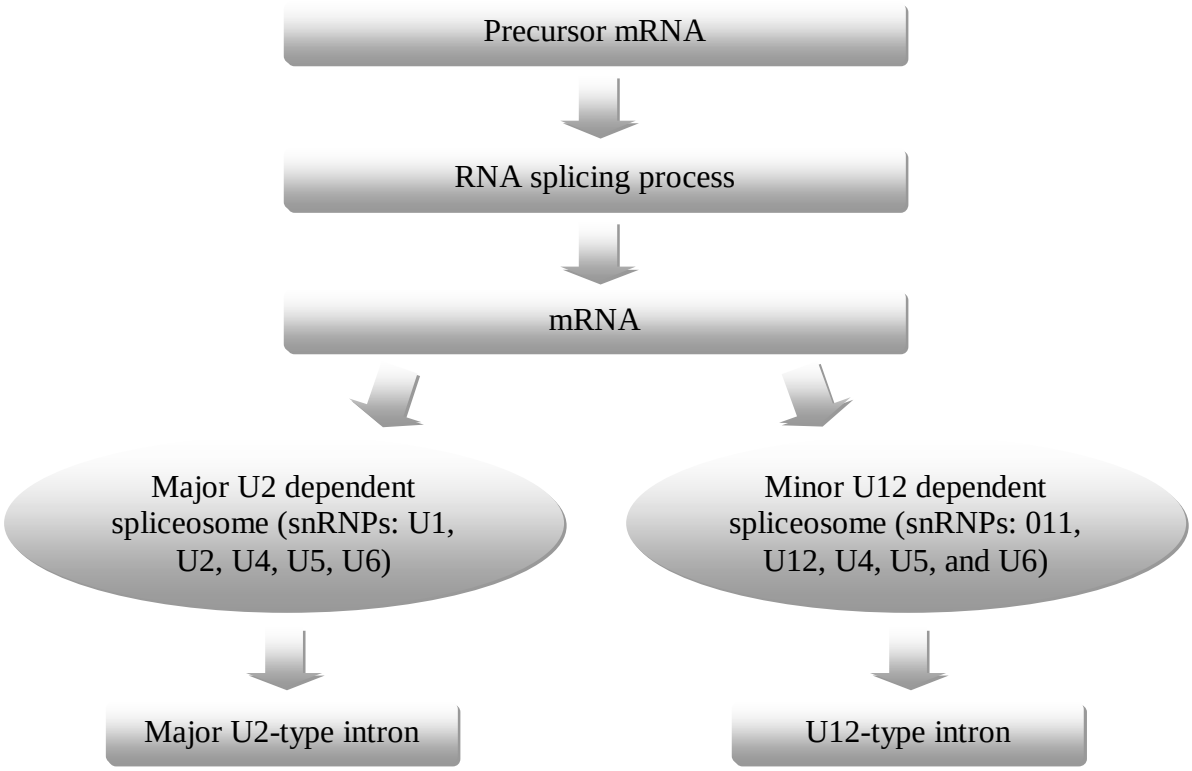


Figure 11: Schematic representation of precursor mRNA (pre-mRNA) splicing

2.1.7 Antisense and RNAi

Identification of a transcript closely associated with a disease represents the possibility of a druggable target, and connectivity analysis the possibility of identifying a small molecule that may display the requisite pharmacological activity for the target (Corey D, 2007). Another approach to modulation of the target is through the use of a biological therapeutic agent. One such biological drug approach is the direct antisense targeting of an mRNA strand using an oligonuleotide with the goal of causing mRNA degradation through activation of RNAses. Another approach is to use RNAi technology (RNAi = siRNA). Instead of using a single strand antisense molecule, RNAi utilizes a short (typically 23bp) double stranded RNA molecule which is activated by Dicer and incorporated into the RISC complex (Achenbach T, et al., 2003). Upon binding the mRNA is then targeted for degradation by RNAses. Despite the appeal of antisense RNAi therapeutic theory only one such molecule has made it to the clinic, Fomivirsen (5'-GCGTTTGCTCTTCTTCTTGCG-3') which is an antisense oligo used for the treatment of cytomegalovirus retinitis (Geary R, et al., 2002). Many more however are in clinical trials (Rayburn E and Zahng R, 2008).

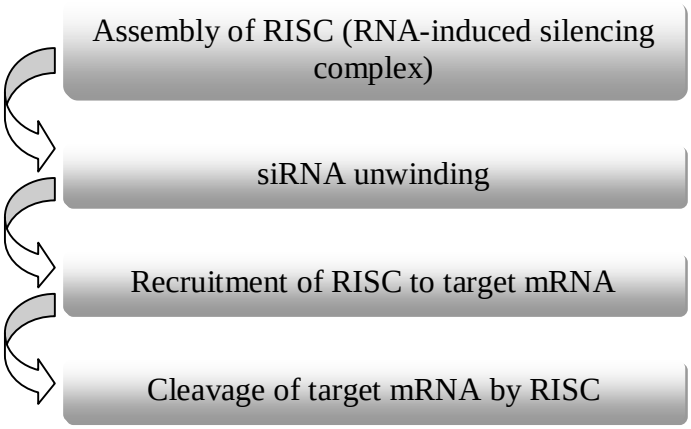


Figure 12: Schematic representation of the mechanism of RNA interference.

Transcriptomics has particular application in the development of direct mRNA targeting therapeutic molecules, not just for the target identification but also for the recognition of off target, or downstream mRNA effects. Similarly if downstream effects occur, for example on transcription rates, then these will be also easily recognized using whole genome profiling. RNAi transcriptomic analysis will be particularly important for RNAi technology where the mechanism of action relies on amplification through the RISC complex. This mechanism can be saturated *in vivo* preventing the normal processing of miRNA (the physiological equivalent of RNAi) species with undesirable consequences (Grimm D, et al., 2006). Both transcriptomic profiling of mRNA species and miRNA (see below) have potential in drug development for recognizing these undesirable off target RNAi effects.

2.1.8 miRNA and mRNA translation analysis in drug development

miRNA species are the physiological equivalent of RNAi though their mechanism of action is a little different. miRNAs do not require complete homology to the mRNA sequence, utilizing instead a 7-8bp target sequence in the 3'UTR of the mRNA to which they bind through the miRNA-RISC complex (Lewis B, et al., 2003). This leads to suppression of mRNA translation by repression of initiation or elongation (. Chu C-Y and Rana T, 2006) rather than causing the mRNA to be targeted for degradation.

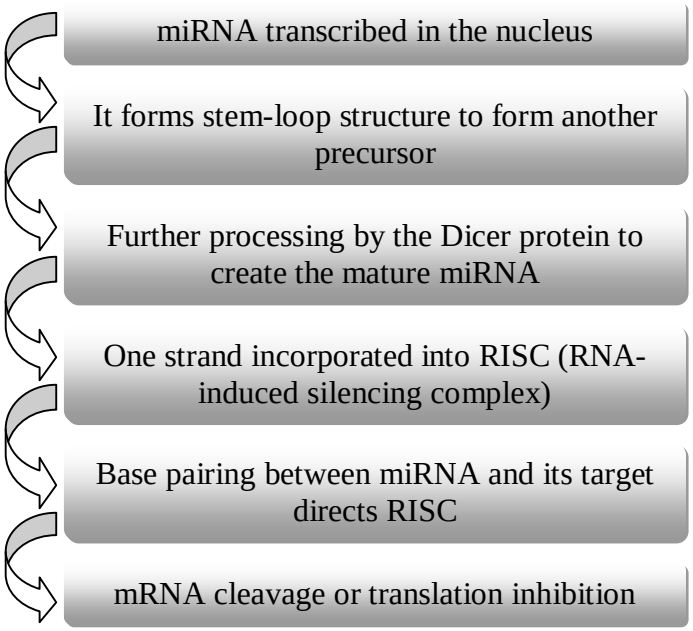


Figure 13: Steps in miRNA and mRNA translation

The precursor of a miRNA (pri-miRNA) is transcribed in the nucleus. It forms a stem−loop structure that is processed to form another precursor (pre-miRNA) before being exported to the cytoplasm. Further processing by the Dicer protein creates the mature miRNA, one strand of which is incorporated into the RNA-induced silencing complex (RISC). Base pairing between the miRNA and its target directs RISC to either destroy the mRNA or impede its translation into protein. The initial stem−loop configuration of the primary transcript provides structural clues that have been used to guide searches of genomic sequence for candidate miRNA genes

These small RNA species are transcribed from the genome in the same manner as protein coding genes and are under the control of many of the same transcriptional regulatory mechanisms (Lee Y, et al., 2004). Once processed and transported to the cytoplasm (for a review of miRNA biochemistry see (He L and Hannon G, 2004) they have the potential to alter the translation of mRNA species to which they have complementarities. Thus, these genes have network properties whereby a differential expression in a miRNA can affect the protein expression of many downstream genes. The effect of this translational suppression on specific protein levels has been

elegantly shown in two recent proteomic studies (Baek D, et al., 2008; Selbach M, et al., 2008). Many studies are now demonstrating the importance of miRNA expression in normal physiology and development and in the alteration of gene expression during xenobiotic exposure and disease progression (Pasquinelli A, et al., 2005). Genomic methods similar to those used for transcriptomic analysis can be used to identify differentially expressed miRNA species in a cell or tissue. The only essential differences in the technique are that the labeling is altered to a form of end labeling because the miRNA species are short and lack a poly A tail (Thompson J, et al., 2004; Thompson J, et al., 2007) and probes on the microarray can be altered to increase the stability of the hybridization kinetics, for example by including locked nucleic acids (Castoldi M, et al., 2006; Castoldi M, et al., 2008). Standard genomics methods can therefore be used to recognize druggable miRNA species. Furthermore, these species can be utilized as targets relatively simply by using the same technology as that for antisense methods. Antisense species to miRNA species are called 'Antagomirs' (Hammond S, 2006). Additionally the miRNA may itself be employed as a drug. For example, expression of several miRNA species are down regulated in inflammatory skin conditions such as Psoriasis (Sonkoly E, et al., 2007). This suggests that these miRNA species may be involved with the normal physiological suppression of genes involved in the inflammatory process, and that if replaced may act to reduce the inflammation in such conditions by suppressing translation of the inflammatory genes (Sonkoly E, et al., 2008). This type of approach has been used successfully in cardiac hypertrophy (Carè A, et l.,2007) and the whole miRNA field has huge potential for future drug development built on a miRNA genomic analysis foundation (Latronico M, et al., 2008).

2.1.9 Translationalomics

Using transcriptomics to verify any off-target effects of a miRNA therapy is not as simple as that for siRNA because most miRNA species suppress translation and so do not cause degradation of the mRNA. Therefore off-target mRNAs may not be detected by a microarray measure of mRNA levels. What is required in this instance is a measure of mRNA translation. This can be achieved by using sucrose density gradients to separate the mRNA species under active translation followed by analysis on microarrays (Melamed D and Arava Y, 2007)). Great care has to be taken with the experimental design and in particular data normalization to prevent the mistaken identification of transcriptional rather than translational events (Taylor E and Gant T, 2008).With this caveat however it is possible to use such methods to identify both on, and off, target mRNA transcripts undergoing translational regulation in response to the use of miRNAs or

antagomirs as drugs. It is also becoming clear, primarily driven by research into the physiological roles of miRNA, that translational control represents a fundamental process in normal development and can be disrupted in disease states. Using the above technique it has been shown that mir21 plays a role in the progression of colorectal cancer to a metastatic phenotype (Asangani I, et al., 2008). Furthermore Gabriely *et al* demonstrated that mir21 targets matrix metalloproteinases and by doing so promotes glioma invasion Gabriely G, et al., 2008). Therefore as proposed by these authors mir21 appears to be a very good target for antagomir drugs. There is highly likely to be a great deal more development in this field and novel genomic methods will have a significant role to play in both discovery and development

2.1.10 Proteins, cell and plasmid 'omics.

Finally, there is another whole world of arrays that are constructed with plasmids, proteins (antigens and antibodies), glycoproteins, cells and even chemicals. All of these have application in drug discovery and development, but are outside the scope of genomics.

Chapter Three

Genomics & Drug Discovery

3.1 New drug target

Molecular medicine and genomics technologies are inseparable for defining new molecular targets. cDNA databases and elementary informatic tools provide instantaneous glimpses of gene families or tissue-restricted expression patterns as a means of new target identification. In addition, cDNA microarrays and two-dimensional gel electrophoresis unmask the expression of genes with unassigned or unexpected functions. Depletion of mRNA with ribozymes or neutralization of proteins with intracellular antibodies enables investigators to reject or embrace new molecular hypotheses about the determinants of disease, pharmacology or toxicology.

- cDNA databases and elementary informatic tools
- cDNA microarrays and two-dimensional gel electrophoresis
- Depletion of mRNA with ribozymes or neutralization of proteins with intracellular antibodies

3.2 Chemical genomics

Chemical probes are cell permeable small molecules which potently inhibit protein function. Modern methods for identifying chemical probes rely on the screening of thousands of chemicals against a target(s) of interest *in vitro*. However, this approach is ineffective, and the number of approved new drugs for treating human diseases has been in steady decline over the past decade. This is because in target-based screening where the target is pre-defined and drug-target interactions are conducted *in vitro*, the potency of the chemical *in vitro* rarely translates to low-dosage and high specificity efficacy *in vivo*. It also limits our ability to discover other drug targets (Sams-Dodd F, 2005)

Chemical genetic approach to drug discovery

Target based approach to drug discovery

Figure 14: Target-based approach vs. chemical genetics approach to drug discovery

The target-based approach begins with the selection of a protein target based on its relevance to a disease state. Then, a high through put screen is set-up to look for chemicals that modulate target protein activities. Interesting hits are then validated and modified to improve its efficacy. On the other hand, with chemical genetics approach, a large chemical library is screened in a cell-based assay to find bioactive molecules for a disease state. After validation and modification, most interesting hits are subject to subsequent assays to identify cellular targets.

Phenotypic screening in model organismsis an alternative approach for discovering new small-molecule therapeutics which addresses the aforementioned shortfalls of current methodologies. Phenotypic screening is a method for the systematic, unbiased, and parallel testing of thousands of molecules for a desired cellular phenotype. Employing cell-based phenotypic screening addresses two drawbacks specifically associated with the conventional target-based screening methods. Firstly, screening using a cell-based assay directly measures drug potency in complex cellular environment, which allows for the early assessment of the biological activities and off-target potential of a drug candidate. Secondly, screening for chemicals that alter the entire cellular pathway bypasses the bias of target pre-selection, leading to the unbiased discovery of new druggable targets (Swinney DC and Anthony J, 2011).

Two criteria must be met in a drug discovery process (Cong F, et al., 2012). Firstly, large chemical libraries must be screened to identify biologically active compounds that elicit the desired phenotype. Secondly, once the compounds are found, the relevant protein targets need to be identified. Chemical genetics can specifically address these two areas in drug discovery.

Table 3: Examples of association between drug response and genetic variants (Roden DM, et al., 2006)

Drug	**Variable clinical effect**	**Gene with Associated variants**	**Possible mechanism**
Azathioprine and mercaptopurine	Bone marrow aplasia Reduced therapeutic effect at standard doses	TPMT	Hypofunctional alleles Wild type alleles

Some antidepressants and β-blockers	Increased side effect risk decreased efficacy	CYP2D6	Hypofunctional alleles Gene duplication
Omeprazole	*Helicobacter pylori* cure rate	CYP2C19	Hypofunctional alleles
Irinotecan	Neutropenia	UGT1A1	Decreased expression due to regulatory polymorphism
HIV protease inhibitor	Central Nervous System lebels	MDR1	Altered p-glycoprotein function
β-blockers	Blood pressure lowering and heart rate slowing	ADRB1	Altered receptor function or number
Inhaled β2 agonist	Bronchodilation	ADRB2	Altered receptor function or number
Diuretic	Blood pressure lowering	ADD1	Altered cytoskeletal function by adducin variants
warfarin	Anticoagulation	VKORC1 CYP2C9	Variant haplotypes in regulatory region leading to variable expression Coding region variants causing reduce 5-warfarin clearance
Abacavir	Immunologic reactions	HLA variants	Altered immunologic responses
QT-prolonging antiarrhythmics	Drug-induced arrhythmia	Ion channel genes	Exposure of subclinical reduction in repolarizing currents by drugs

General anesthetics	Malignant hyperthermia	RYR1	Anesthetic induced increased release of sarcoplasmic reticulum calcium by mutant channels
Inhaled steroids	Bronchodilation	CRHR1	Unknown
HMG-CoA reductase inhibitors (statins)	Low density lipoprotein Cholesterol lowering	HMGCR	Altered HMG CoA reductase activity

Abbreviations:

ADDI= the gene encoding α-adducin; ADRBI= the gene encoding the β1 adrenergic receptor; ADRB2= the gene encoding the β2 adrenergic receptor; CRHR1= the gene encoding the corticotrophin-releasing hormone receptor-1; CYP2C19: the gene encoding the 2C9 cytochrome P450 isoform; CYP2D6: the gene encoding the 2D6 cytochrome P450 isoform; HMG-CoA= 3-hydroxy-3-methylglutaryl coenzyme A; HMGCR: the gene encoding the HMG-CoA reductase; MDR1: the gene encoding p-glycoprotein; RYR1: the gene encoding the skeletal muscle calcium-release channels; TPMT: the gene encoding thiopurine methyltransferase; UGT1A1: the gene encoding uridine diphosphate glycosyltransferase 1 family, polypeptide A1: VKORC1= the gene encoding vitamin k epoxide reductase complex subunit 1.

Table 4: Examples of association between drug response and genetic variants (Roden DM, et al. 2006)

Mode of action	**Chemical family**	**Site of action**	**Resistant weed species (U.S)**
Synthesis of lipid	Arioxyphenoxy propionate cyclohexanedion	Acetyl CoA carboxylase (ACCase)	15
Amino acid synthesis	Sulfonylurea Glycine	Acetolactate synthease(ALS) 5-enolpyruvyl-shikimate-3-phosphate Synthase(EPSP)	38 7

Nitrogen metabolism Inhibitors of pigments	Phosphonic acid Isoxazolidinone Isoxazole,triketone	Glutamine synthase Diterpene synthase Hydroxyphenylpyruvate Dioxygenase	0 0 0
photosynthesis	Triazine, trazinone, Nitrile, Benzothiadiazole, Ureas Bipyridilium	Photosystem 2 electron transport Photosystem 1 electron transport	22 1 7
Disruptor of cell membrene	Dietylether N-phenylphthalimide Thiadiazole	PPO inhibitors	2 4
Seeding root growth Seeding shoot growth			
Growth regulators	Phenoxy-carboxylic acid Benzoic acid semicarbazone	Auxin receptor Auxin transport	7

Broad spectrum herbicides target a wide range of plant-specific processes, including photosynthesis and amino acid biosynthesis. The widespread use of herbicide has selected for resistance in some common weed species, necessitating the development of novel post control measures.

3.3 Use of genomics to identify biomarkers

Biomarkers are biological measurements that can be used to predict risk of disease, to enable early detection of disease, to improve treatment selection and to monitor the outcome of therapeutic interventions. One major motivation of the Human Genome Project was the identification and development of such biomarkers for ‘personalized, preventive and predictive medicine’.

Although the sequencing of the human genome has had profound impacts on biomedical research in many other fields, and while it is still too early to fully assess its impact on biomarker development (Lander, 2011), I will provide an interim analysis and identify some of the roadblocks to progress.

The development of genetic biomarkers for predicting risk of disease in individuals has had limited success to date. Numerous large whole-genome association studies (Ioannidis et al, 2010)involving thousands of patients have been conducted for many chronic diseases. These studies have genotyped cases and controls in order to identify germ-line polymorphisms that put individuals at higher risk for developing a specific disease. Many genetic loci have been identified as statistically significant and, in some cases, are providing valuable leads for understanding the biological basis of the diseases.

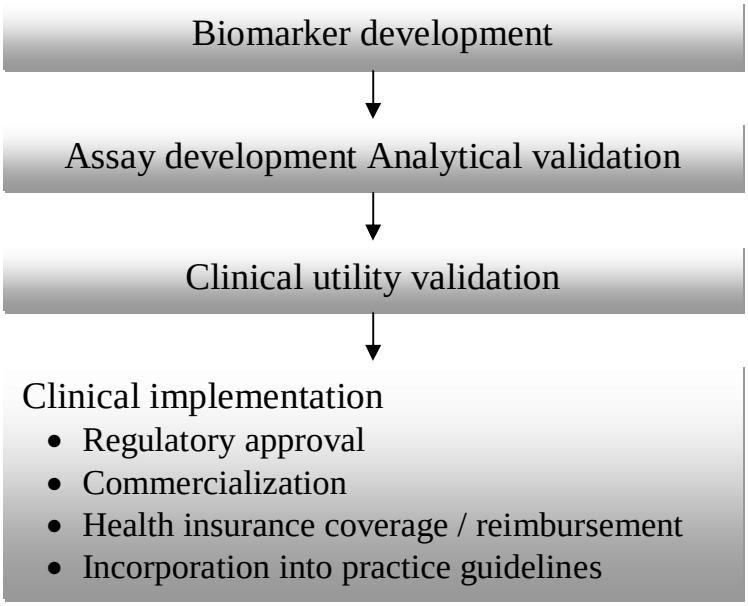


Figure 15: Progress in genomic biomarker development

Biomarkers are biological measurements that can be used for a variety of purposes, including identifying individuals who are at high risk of developing a disease, detecting disease early at a stage when it is treatable and diagnostic classification for personalized treatment based on a biological characterization of the disease of each individual patient. The sequencing of the human genome has provided an important body of information for the development of biomarkers for all of the purposes mentioned. In this paper, I provide a short and personal assessment of the progress achieved in these areas of genomic biomarker development. It is indicated that (i) progress has been slow in personalized risk prediction and in early detection; (ii) genome-wide association studies are more likely to provide leads for understanding the pathogenesis of diseases than useful information on personalized risk assessment; and (iii) development of biomarkers sufficiently sensitive and specific for early detection of diseases that will be life-threatening is very challenging and the validation of such biomarkers requires very large randomized screening trials. The development of biomarkers for personalizing treatment

selection, particularly in oncology, has seen greater progress. Key bottlenecks that limit progress in the translation of discoveries in genomics to biomarkers and treatments that reduce mortality and morbidity from chronic diseases are also discussed.

Table 5: Biomarkers in clinical use (Goossens N, et al., 2015)

Organ	**Cancer**	**Biomarker and Mechanism**	**Assay for measurement**	**Associated target and drug**	**Approximate proportion of positive tests**	**Stage of clinical validation**
Breast	Breast cancer	HER2: oncogene overexpression	ISH, IHC IHC, LBA	HER2: trastuzumab, Pertuzumab, ado-trastuzumab emtansine	18-20%	In clinical use
		ER/PR: suggests sensitivity to endocrine therapy		ER; endocrine therapy (tamoxifen, aromatase inhibitors)	75%	In clinical use
Gastro-intestinal	Colorectal cancer	KRAS: mutations activate RAS-RAF-MEK pathway and resistance to EGFR therapy KIT: mutation leads to consitutional activation	PCR	EGFR: cetuximab, panitumumab	40% mutated	In clinical use
	GIST	KIT: mutation leads to consitutional activation	IHC	BCR-ABL: imatinib	95%	In clinical use
	Esophago-gastric adenocarcinoma	HER:2 Oncogene overexpression	ISH, IHC	HER2: trastuzumab	7-22%	In clinical use

Hematol ogial	chronic myeloid leukemia	BCR-ABL: balanced t(9;22) leading to the formation of a constitutively active tyrosine kinase	Cytogenetics, FISH, RT-PCR	BCR-ABL: imatinib, dastinib, nilotinib	<90%	In clinical use
	Acute promyelocytic leukemia	PML-RARa: balanced t(15:17) leading to aberrant retinoid receptor	Cytogenetics, Fish, RT-PCR	PML-RARa: All trans retinoic acid	>90%	In clinical use
Lung	NSCLC	EGFR (HER1); mutations in tyrosine kinase domain	Sequencing, ISH	EGFR: Erlotinib, gefitinib, afatinib	15% adenocarci noms in USA (higher in Asians, women and nonsmoker s)	In clinical use
		ALK: Inversion in charomosome 2 leads to EML4-ALK fusion oncogene	FISH (HC)	ALK: crizotinib, certinib (alectinib under development)	adenocarci noma)	
	Lug adenocarcino ma	Multiple genes:				Continued
	Lug adenocarcino ma	Multiple genes:				Continued
		BRAF (V600E and non-V600E)	Multiplex sequencing	BRAF: AZD6244	2%	

		EGFR (HER1): mutations in tyrosine kinase domain		EGFR: eriotinib, gefitinib, afatinib, cetuximab	17%	
		HER2: oncogene overexpression		HER2: decinutubub neratinib, lapatinib, trastuzumab	3%	
		KRAS: mutations activate RAS-RAF- MEK pathway and resistance to EGFR therapy		KRAS: erlotinib, tivantinib, everolimus, ridaforalimus, AZD 6244	25%	
		ALK: inversion in chromosome 2 leads to EML4- ALK fusion oncogene MET		ALK: crizotinib, ceritinib	8%	
Skin	Melanoma	BRAF V600: 80-90% V600E mutation, a downstream mediator of RAS, leads to downstream activation of MEK and ERK	Sequencing	BRAF: vemurafenib, dabrafenib	4-60%	In clinical use

Abbreviations:

HER2-human epidermal growth factor 2: (F)ISH-(fluorescence)in situ hybridization: IHC-immunohistochemistry: ER-estrogen receptor: PR- progesterone receptor: LBA-ligand binding assay: MEK- mitogen-activated protein kinase: EGFR- epidermal growth factor receptor: (RT)-PCR- (reverse transcription) polymerase chain reaction: GIST- gastrointestinal stromal tumor: PLM- promelocytic leukemia gene: RARa- retinoic acid receptor-alpha: NSCLC-non small cell

lung cancer: ALK: anaplastic lymphoma kinase: EML4- echinoderm microtubule-associated protein-like 4: ERK-extra cellular-signal-regulated kinase.

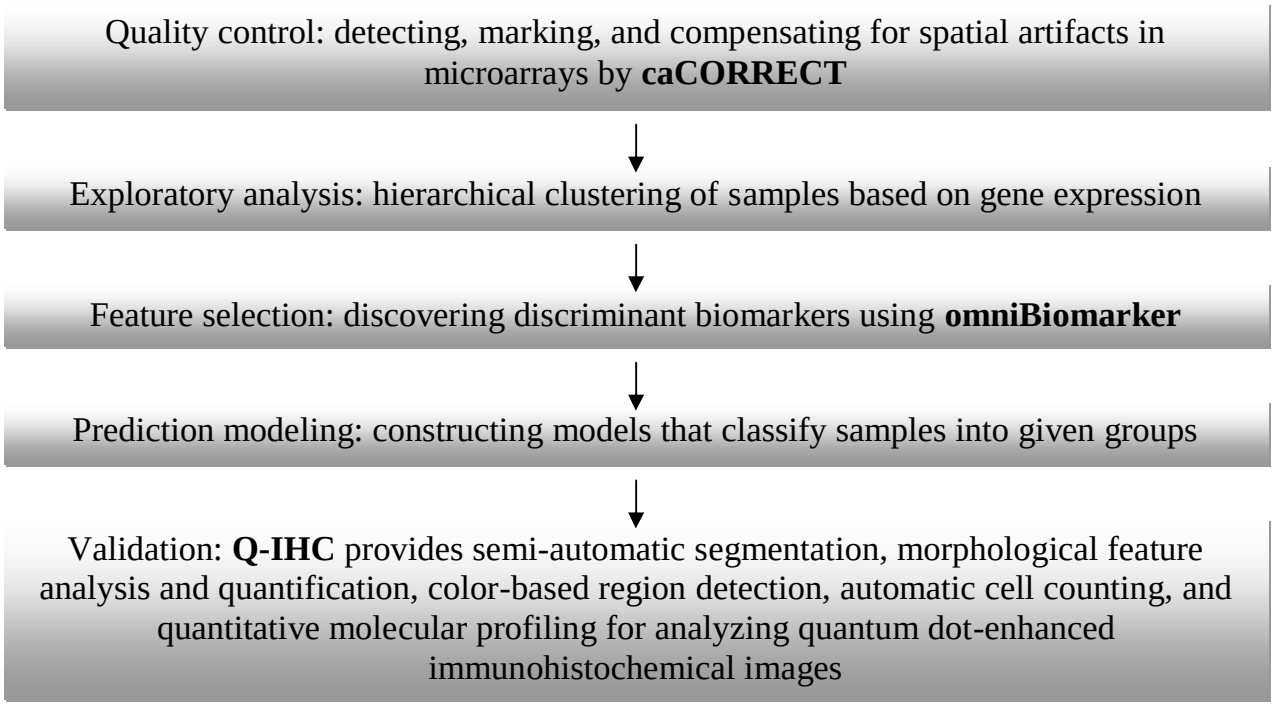


Figure 16: The translational bioinformatics for identifying biomarkers. (Yan Feng, et al., 2009)

caCORRECT improves the quality of collected microarray data by removing noise and artifacts (such as those produced by scratches, edge effects, and bubble effects) from the data, while retaining high-quality genes on the array. She notes that running QA on all microarray data is important before any biomarker selection. Once that has been done, omniBioMarker software can be used to identify and validate biomarkers from high-throughput gene expression datasets by searching healthy and cancer patient samples for genes most likely to be associated with cancer, as well as genes that appear to interact in cancer development. omniBioMarker also relates the selected biomarkers to clinical information. “The key focus for us is making it work for real medicine observes.

Q-IHC and omniVIsGrid—as well as others in development. Q-IHC analyzes and quantifies multi-spectral images, including quantum dot–stained histopathologic images, and onmiVisGrid is a grid-based tool that visualizes data and analysis processes of microarrays, biologic pathways, and clinical outcomes. “No matter what the data source is—genomics, proteomics, high-throughput, or next-generation sequencing—the goal is the same: QC, reproducibility, and reliability.

In the field of molecular biology, gene expression profiling is the measurement of the activity (the expression) of thousands of genes at once, to create a global picture of cellular function. These profiles can, for example, distinguish between cells that are actively dividing, or show how the cells react to a particular treatment. Many experiments of this sort measure an entire genome simultaneously, that is, every gene present in a particular cell.

DNA microarray technology (Microarrays Factsheet, 2007) measures the relative activity of previously identified target genes. Sequence based techniques, like serial analysis of gene expression (SAGE, Super SAGE) are also used for gene expression profiling. SuperSAGE is especially accurate and can measure any active gene, not just a predefined set. The advent of next-generation sequencing has made sequence based expression analysis an increasingly popular, "digital" alternative to microarrays called RNA-Seq. However, microarrays are far more common, accounting for 17,000 PubMed articles by 2006. (Kawasaki ES, 2006)

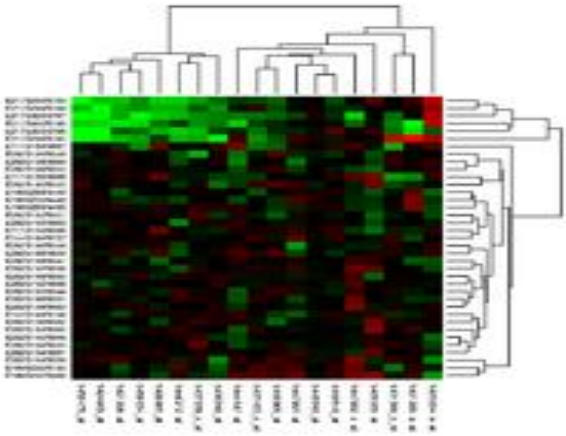


Figure 17: Heat maps of gene expression values show how experimental conditions influenced production (expression) of mRNA for a set of genes. Green indicates reduced expression. Cluster analysis has placed a group of down regulated genes in the upper left corner.

Expression profiling is a logical next step after sequencing a genome: the sequence tells us what the cell could possibly do, while the expression profile tells us what it is actually doing at a point in time. Genes contain the instructions for making messenger RNA (mRNA), but at any moment each cell makes mRNA from only a fraction of the genes it carries. If a gene is used to produce mRNA, it is considered "on", otherwise "off". Many factors determine whether a gene is on or off, such as the time of day, whether or not the cell is actively dividing, its local environment, and chemical signals from other cells. For instance, skin cells, liver cells and nerve cells turn on (express) somewhat different genes and that is in large part what makes them different.

Therefore, an expression profile allows one to deduce a cell's type, state, environment, and so forth.

Expression profiling experiments often involve measuring the relative amount of mRNA expressed in two or more experimental conditions. This is because altered levels of a specific sequence of mRNA suggest a changed need for the protein coded by the mRNA, perhaps indicating a homeostatic response or a pathological condition. For example, higher levels of mRNA coding for alcohol dehydrogenase suggest that the cells or tissues under study are responding to increased levels of ethanol in their environment. Similarly, if breast cancer cells express higher levels of mRNA associated with a particular transmembrane receptor than normal cells do, it might be that this receptor plays a role in breast cancer. A drug that interferes with this receptor may prevent or treat breast cancer. In developing a drug, one may perform gene expression profiling experiments to help assess the drug's toxicity, perhaps by looking for changing levels in the expression of cytochrome P450 genes, which may be a biomarker of drug metabolism. (Suter L, et al., 2004)Gene expression profiling may become an important diagnostic test. (Magic Z, et al., 2007; Cheung AN, 2007)

More commonly, expression profiling takes place before enough is known about how genes interact with experimental conditions for a testable hypothesis to exist. With no hypothesis, there is nothing to disprove, but expression profiling can help to identify a candidate hypothesis for future experiments. Most early expression profiling experiments, and many current ones, have this form (Chen JJ, 2007).which is known as class discovery. A popular approach to class discovery involves grouping similar genes or samples together using k-means or hierarchical clustering. Apart from selecting a clustering algorithm, user usually has to choose an appropriate proximity measure (distance or similarity) between data objects.(Jaskowia K,et al., 2014)The figure above represents the output of a two dimensional cluster, in which similar samples (rows, above) and similar gene probes (columns) were organized so that they would lie close together. The simplest form of class discovery would be to list all the genes that changed by more than a certain amount between two experimental conditions.

Class prediction is more difficult than class discovery, but it allows one to answer questions of direct clinical significance such as, given this profile, what is the probability that this patient will respond to this drug? This requires many examples of profiles that responded and did not respond, as well as cross-validation techniques to discriminate between them.

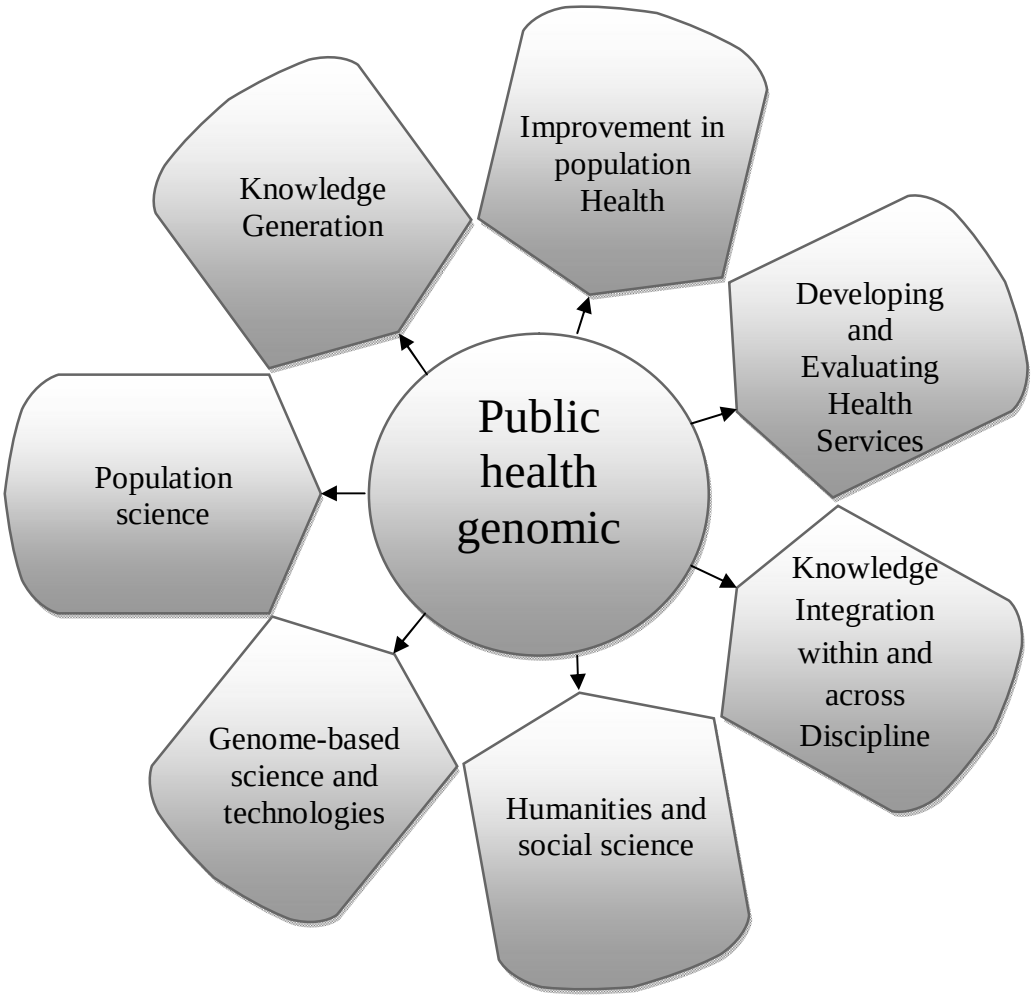


Figure 18: Impact of Public Health Genomic

Table 6: Some examples of genetic diseases

Genetic disease	**Deficient gene product**	**Cell type**
Thomsen's disease (Autosomal dominant myotonia congenita)	Allelic disorders associated with mutations in a gene	Muscle
Becker's disease (Autosomal recessive myotonia congenita)	coding for skeletal muscle chloride channel (CLCN1)	Muscle
Familial hyperkalaemic periodic paralysis	α1 subunit of the sodium channel (SCN4A)	Muscle
Paramyotonia congenita	α1 subunit of the sodium channel (SCN4A)	Muscle

Parkinson's disease Neurological channelopathies Voltage-gated	Dopamine	Substantia nigra
Duchenne muscular dystrophy	Dystrophin	Muscle
Lesch-Nyhan syndrome	Hypoxanthine transferase phosphoribosyl	Basal ganglia
Familial hypercholesterolemia	Low-density lipoprotein receptor	Liver cells
Familial paroxysmal choreoathetosis	Chromosome 1p where a cluster of potassium channel	
Nocturnal frontal lobe epilepsy	α4 subunit of the nicotinic acetylcholine receptor	
Familial startle disease	α1 subunit of the glycine receptor	
Episodic ataxia type 1	Potassium channel genes	Cerebellum and peripheral nerve
Episodic ataxia type 2	α1 subunit of a brain specific calcium channel	Brain

3.4 Identification and assignment of candidate target

High-throughput gene sequencing has revolutionized the process used to identify novel molecular targets for drug discovery. Thousands of new genes

Sequences have been generated but only a limited number of these can be converted into validated targets likely to be involved in disease. We describe here some of the approaches used at SmithKline Beecham to select and validate novel targets. These include the identification of selective tissue gene product expression, such as for cathepsin K, a novel osteoclast-specific cysteine protease. We also describe the discovery and functional characterization of novel members of the G protein coupled receptor super family and their pairing with natural ligands. Lastly, we discuss the promises of gene microarrays and proteomics, developing technologies

that allow the parallel analyses of tissue expression patterns of thousands of genes or proteins, respectively.

Genetic screening is an effective strategy for target identification, especially in organisms where whole genome library collections are available. There are three gene-dosage based assays in yeast that are commonly used to identify drug targets. All three assays are unbiased growth-based assays that do not require prior knowledge of a compound's mechanism of action (Cong F, et al., 2012; Ericson E, et al., 2010). HaploInsufficiency Profiling (HIP) assay is based on the principle that a decreased dosage of a drug target gene can result in an increase in drug sensitivity (Figure 3A). In a HIP assay, heterozygous deletion mutants are treated with a drug of interest. Strains that exhibit increased sensitivity in the form of growth inhibition will be identified. The HIP assay reveals the direct target as well as other components in the same pathway.

A: The HIP assay uses heterozygous deletion collection to look for strains that are sensitive to drug treatment, which reveals direct target and other genes in the same pathway

B: The HOP assay works in a similar way to the HIP assay. It uses homozygous deletion collection and identifies genes that buffer the drug target pathway

C: The MSP works in the opposite way of the HIP assay. It uses the over-expression DNA library to look for strains that are resistant to drug treatment, thereby identifying direct

Figure 19: Chemical genetics-based dosage-dependent assays for target identification.

Homozygous Profiling (HOP) assay is similar to the HIP assay except that both copies of a nonessential gene are deleted.Instead of identifying a direct target, the HOP assay often identifies genes that buffer the drug target pathway. This is because the HOP assay mimics a double deletion mutant in that the second gene disruption is achieved via compound inhibition. Direct target identification, however, can be inferred by integrating HOP chemical genetic profiles with genetic interaction profile. The chemical genetic profile of a drug will be similar to the genetic interaction profile of its target gene.

Multicopy suppression profiling (MSP) assay is based on the principle that increased dosage of a drug target gene can confer resistance to drug-mediated growth inhibition, essentially a mirror to the HIP assay. In a MSP assay, strains that contain high-copy over expression plasmids are treated with a drug of interest, and individuals that exhibit growth advantage in the presence of

the drug will be identified. Like the HIP assay, MSP assay often identifies direct target of a drug.

All three gene-dosage based screening assays are often performed in a single liquid culture using bar-coded strains to assess the competitive growth of the entire collection (Giaever G, et al., 2002). However, these assays can also be performed on solid media using colony size to measure individual colony fitness (Ho CH, et al., 2011).

S.cerevisiae is ideally suited for high-throughput phenotypic screening for several reasons. Firstly, it has a short doubling time and simple growth requirements. Secondly, its genome is the best characterized among eukaryotic models. Thirdly, many core cellular processes in *S. cerevisiae*, such as cell cycle control, DNA repair and various metabolic pathways, are conserved in humans (D B, et al., 1997; Botstein D and Fink GR, 2011). Finally, many laboratories have experience using yeast strains to model disease state in human cells and screen for promising compounds (St Onge R, et al., 2012).

However, it usually requires a much higher dosage of chemicals to perform genetic screens in yeast than in mammalian cells. This is because yeast has a cell wall, making it hard for some chemicals to enter the cells (Dielbandhoesing SK, et al., 1998). In addition because yeast have a dynamic chemical defense, which uses efflux pumps that actively transport drugs out of the cells (Ernst R, et al., 2010; Kolaczkowski M, et al., 1998). Special yeast strains that are mutated with these genes will increase the drug sensitivity of a yeast cell (Rogers B, et al., 2001).

3.5 Toxicity determinations

Microarray profiling has become a major omics tool for the characterization of drug toxicities by allowing for large-scale determination of gene expression changes associated with a defined pathology. A key step in toxic genomics is the ability to link a chemical-elicited phenotype with gene expression changes, termed “phenotypic anchoring”(Paules R , 2003). Experimentally determined gene expression signatures can serve as a guide to determining biomarkers that are indicative of toxicological responses that may be as-yet sub-clinical, with no observable morphological changes(Powell CL, et al., 2006).To confirm the utility of the approach, Heinloth et al. (Wang EJ, et al., 2008)demonstrated that patterns of gene expression perturbations observed at sub-toxic doses of acetaminophen in rats may indicate subtle cellular injury that was not detectable by histopathology or clinical chemistry methods within the liver. At toxic doses, expression changes in the same subset of genes associated with mitochondrial dysfunction and

oxidative stress were more exaggerated and changes were detected in additional genes associated with these processes. These data indicate that gene expression profiling has the potential to identify subtle markers of cellular injury that precipitate overt organ toxicity.

Identification of sensitive biomarkers that will assist in monitoring drug therapy for evidence of toxicity or therapeutic outcome, and (in acute poisoning cases) to predict exposure levels, is a critical gap where omics technologies hold promise. Genomic biomarkers of toxicity have recently been identified for a wide variety of toxicants including nephrotoxic agents. (Wang EJ, et al., 2008), testicular toxicants (Khor TO, et al., 2006), and for keratinocyte proliferation in papilloma murine skin model (Ridd K, et al., 2006), to name only a few. The potential of using this technology to identify safety biomarkers is great and may help to create better diagnostic tools for traditionally difficult toxic dynamic monitoring, such as in patients receiving immunosuppressive therapy(Christians U, et al., 2008).

The majority of recent investigations have used microarrays to study toxicity in target organ tissue or in cultured cells. While these experiments often yield important insight into the mechanism of toxicity, they provide limited information for monitoring drug safety in patient populations through non-invasive means. To address this limitation, Bushel et al.(Christians U, et al., 2008) investigated the utility of using gene expression signatures in peripheral blood as an early indicator of pathological changes in the liver following administration of varying doses of acetaminophen. In this study, a prediction algorithm using classifiers and a pattern-based method that was weighted toward non-injurious exposure levels was used to discriminate sub-toxic and toxic exposure doses. Characterization of acetaminophen-induced liver injury using gene expression profiles derived from blood was shown to better predict acute chemical exposure levels than clinical chemistry, hematology, or histopathology analysis, indicating that transcript profiles derived from blood may be a good marker for specific organ toxicity.

Table 7: Genetic biomarkers mentioned in labels of drug products approved by FDA. (Hong H, et al., 2010)

Biomarker	Drug (Brand name)	Drug(Generic name)	Section in Label
CYP2C19	PLAVIX	Clopidogrel bisulfate	Clinical pharmacology: Precaution: Dosage and Administration
	VFEND	Voriconazole	Clinical pharmacology: Use in specific populations.
	EFFIENT	Prasugrel	Clinical pharmacology: Clinical studies
	CELEBREX	Celecoxib	Clinical pharmacology:
CYP2C9	EFFIENT	Prasugrel	Use in specific populations. Clinical pharmacology: Clinical studies
	COUMADIN	Warfarin	Clinical pharmacology: precautions
	CELEBREX	Celecoxib	Clinical pharmacology:
CYP3A4	Codeine sulfate	Codeine sulfate	Drug interaction: Clinical pharmacology:
CYP3A5	EFFIENT	Prasugrel	Use in specific populations.

			Clinical pharmacology: Clinical studies
CYP2B6	EFFIENT STRATTERA	Prasugrel Atomoxetine	Use in specific populations. Clinical pharmacology: Clinical studies Dosage and Administration Warning and precautions: drug interaction clinical pharmacology
CYP2D6	PROZAC Codeine sulfate tablets	Fluoxetine Codeine sulfate	Clinical pharmacology: precautions Warning and precautions: drug interaction Use in specific populations. Clinical pharmacology
VKORC1	COUMADIN CAMPTOSER	Warfarin Levofloxacin	Clinical pharmacology: precautions Clinical pharmacology: Warning Dosage and Administration
UGT1A1	TASIGNA	Nilotinib	Drug interaction: Clinical pharmacology:

HLA-B* 1502	TAGRETOL	Carbamazepine	Warning and precautions
HLA-B* 5701	ZIAGEN	Abcavir	Warning and precautions
Deletion 5q	REVLIMID	Lenalidomide	Hematologic Toxicity: Clinical studies: Precautions: Adverse Reacions

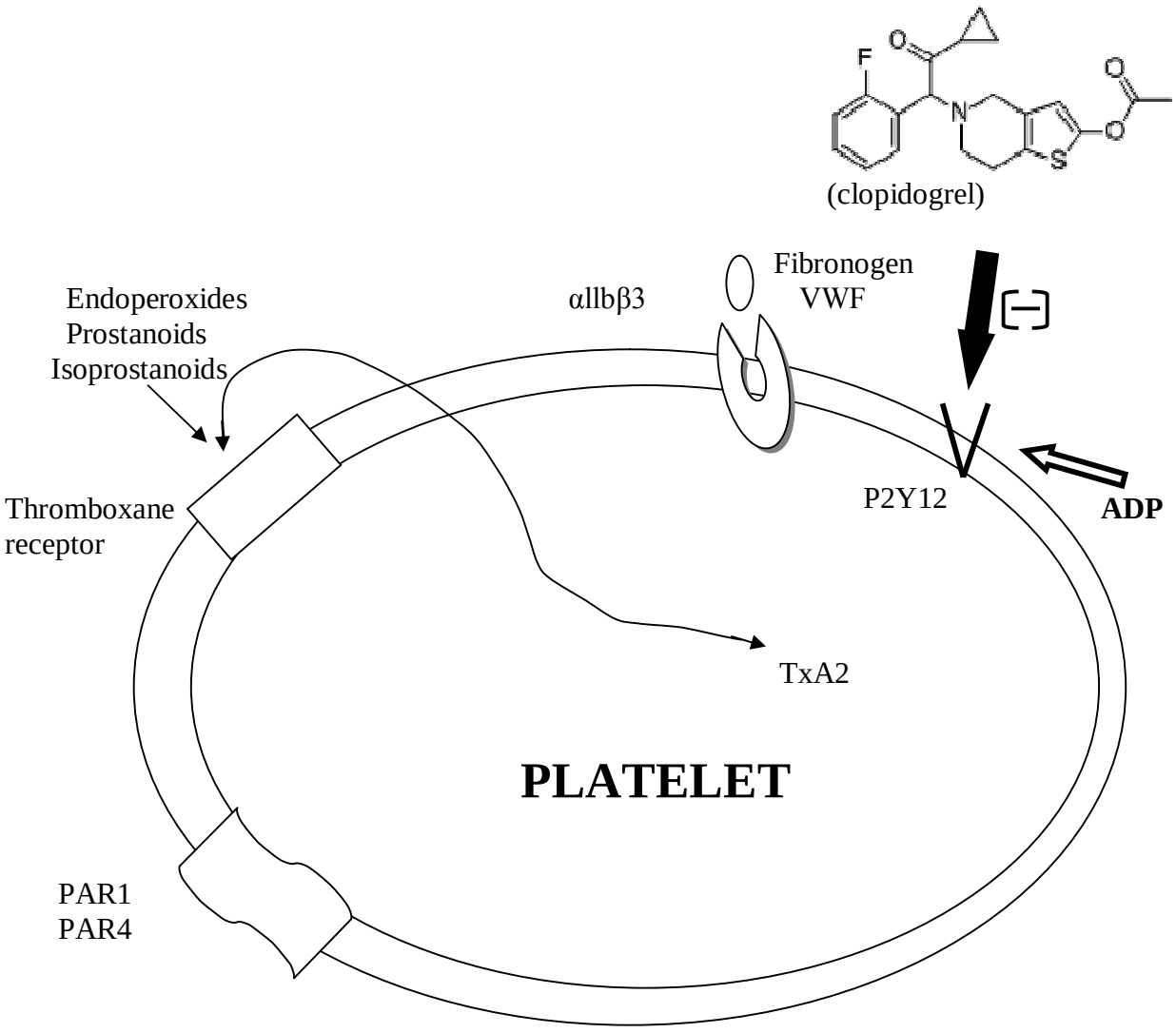


Figure 20: Clopidogrel mechanism of inhibition

3.6 The future

The use of molecular profiling throughout the drug discovery and development process is likely to increase dramatically over the next few years. This will be based on the clear advantages to multivariate biomarker approaches including: the ability to provide a broad view of the biological state of a cell or tissue; the increased predictive power of monitoring multiple parameters simultaneously; and the power of correlating specific molecular phenotypes to clinical, histopathological or disease model endpoints. In some cases, particularly in cancer, developing molecular profiling of genomic expression will likely also dovetail with basic discoveries about the mechanisms and characteristics of the disease.

It is possible that molecular profiling will eventually replace the current single target based dogma as the central paradigm of drug discovery (Figure 2). However, the question as to the speed with which this will occur is difficult to answer. There are many challenges to fulfilling the promise of molecular profiling including: limitations and costs associated with current technologies, the fact that validation of profiles can be time consuming and expensive and resistance to change in organizations – particularly large pharmaceutical companies – that are entrenched in the dogma of single target based drug discovery. Thus, if a paradigm shift occurs, it will likely involve significant upheaval (Dalton WS and Friend SH, 2006)

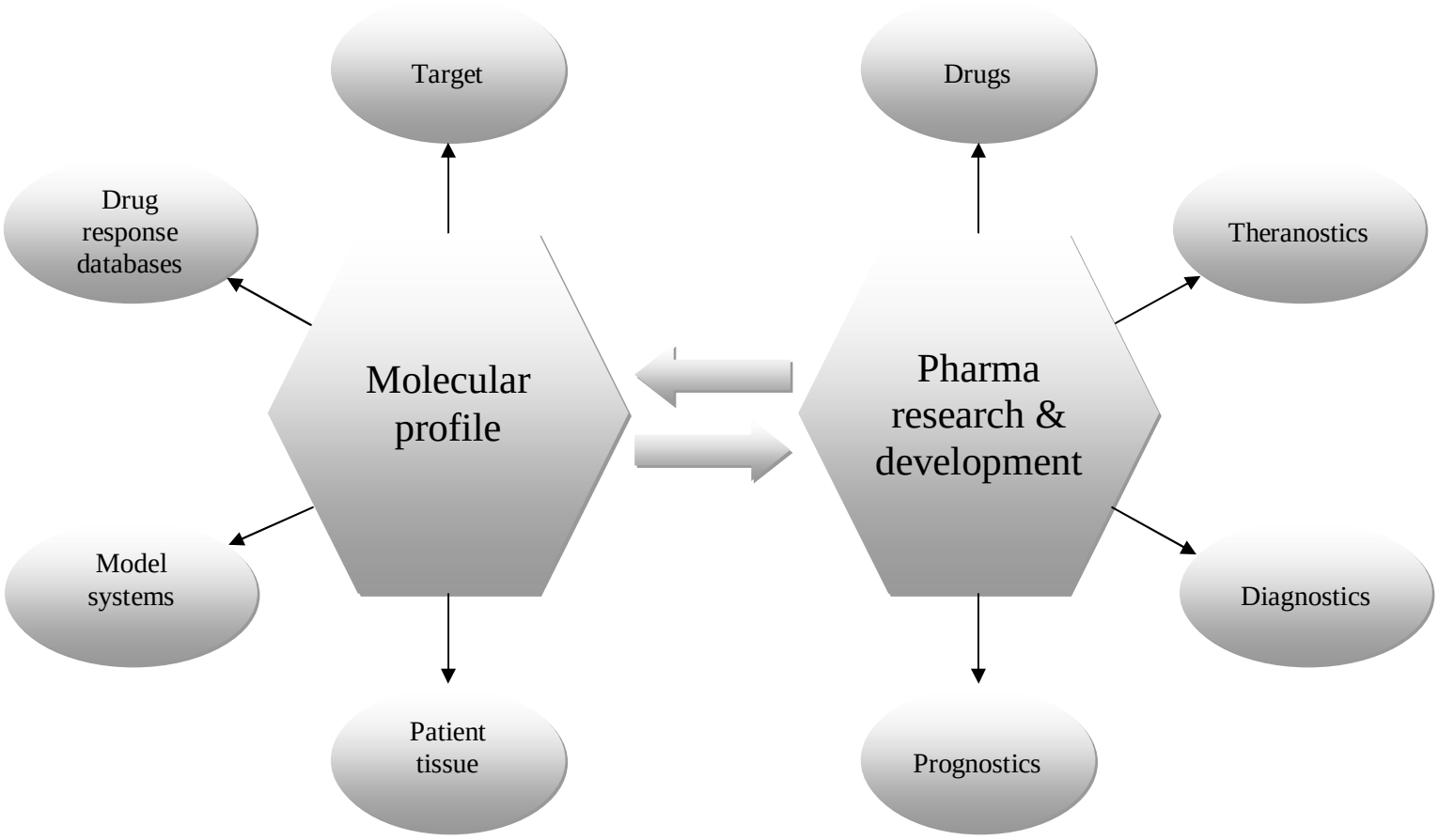


Figure 21: The future of drug discovery will likely see an increasingly central role for molecular profiling

Following are several examples of emerging strategies for the use of molecular profiling in drug discovery and development – mostly in the area of cancer. Cancer is essentially a disease of the genome in that one of the hallmarks of all cancer cells is the presence of profound alterations of their genomic DNA that includes mutations, deletions, amplifications and rearrangements. For this reason both academic and industrial efforts to develop better medicines for cancer have led to many examples of the use of genomic molecular profiling to improve the drug discovery and development process.

Until then, it is clear that the increased use of molecular profiling will continue to make an important contribution to drug discovery and development efforts worldwide and will hopefully lead to lower failure rates, faster progression through the development process, and increasingly precise tests to match the right medicine with the right patient

3.7 Pharmacogenomics

Pharmacogenomics has its roots in pharmacogenetics. Whereas pharmacogenetics is the study of the linkage between an individual's genotype and that individual's ability to metabolize a foreign compound, pharmacogenomics is quite broad in scope, and is similar to molecular medicine, aiming to detect, monitor and treat the molecular causes of disease. Pharmacogenomics involves the application of genomics technologies such as gene sequencing, statistical genetics and gene expression analysis to drugs in clinical development and trials. Since many diseases develop as a result of a network of genes failing to perform correctly, pharmacogenomics can identify the genes or loci which are involved in determining the responsiveness to a given drug. In this way, genetic characterization of patient populations is becoming an integral part of the drug discovery and development process. Pharmacogenomics may aim to capitalize on these new molecular insights to discover new therapeutic targets and interventions and to elucidate the constellation of genes that determine the efficacy and toxicity of specific medications.

Chapter Four

Conclusion

4.1 CONCLUSION

By contrast to the agents administered to patients in clinical wars, the process of drug discovery is not a prescriptive series of steps. The risks are high and there are long timelines to be endured before it is known whether a candidate drug will succeed or fail. At each step of the drug discovery process there is often scope for flexibility in interpretation, which over many steps is cumulative. The pharmaceutical companies most likely to succeed in this environment are those that are able to make informed accurate decisions within an accelerated process. New gene targets for therapeutic intervention only provide a starting point in the long and difficult process of drug discovery. However, genomics will have an important impact in the later stages of drug development, especially in providing an understanding of the molecular nature of diseases and of the responses, both desirable and adverse to drugs. Modern genetics will bring about significant improvements in the provision and practice of healthcare by redefining disease and targeting treatment.

Chapter Five

References

5.1 References

Abecasis,G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker,R.E., Kang, H.M., Marth, G.T., McVean, G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature, 491 (7422), 56–65.* Bibcode*:*2012Natur.491...56T. Doi: 10.1038/nature11632. PMC 3498066. PMID 23128226.

Achenbach, T., Brunner, B., Heermeier, K. (2003) Oligonucleotide-based knockdown technologies: antisense versus RNA interference. *J Chembiochem*, 4, 928-935.

Adib, T.R., Henderson, S., Perrett, C. et al. (2004) Predicting biomarkers for ovarian cancer using gene-expression microarrays. *Br J Cancer*, 90, 686–692.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier,P.H., Smith, A.J., Staden, R., Young, I.G., (1981) Sequence and organization of the human mitochondrial genome. *Nature ,**290*** (5806), 457–65.

Ankeny, R.A. (2003) Sequencing the genome from nematode to human: changing methods, changing science". *J Endeavour,* 27 (2), 87–92. Doi: 10.1016/S0160-9327(03)00061-9. PMID 12798815.

Anzick, S.L., Trent, J.M. (2002) Role of genomics in identifying new targets for cancer therapy. *Oncology (Williston Park), 16,* (5 suppl 4),*7*–13.

Asangani, I., Rasheed, S., Nikolova, D., Leupold, J., Colburn, N., Post, S., Allgayer, H. (2008) MicroRNA-21 (miR-21) post-transcriptionally down regulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Journal of Oncogene* ,27, 2128-2136.

Barnes, B., Dupré, J. (2008) Genomes and what to make of them. Chicago. *University of Chicago Press*,ISBN 978-0-226-17295-8.

Barrett, M.T., Scheffer,A., Ben-Dor, A., Sampas, N., Lipson,D., Kincaid,R., Tsang, P., Curry, B., Baird, K., Meltzer, P. et al. (2004) Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proceedings of the National Academy of Sciences of the United States of America,* 101, 17765-17770.

Baek, D., Villén,J., Shin, C., Camargo, F., Gygi, S., Bartel, D. (2008) The impact of microRNAs on protein output. *Nature* ,455,64-71.

Beckmann, J., Estivill, X., Antonarakis, S.E. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature reviews,* 8,639-645

Beck, S., Rakyan, V. (2008) The Methylome: Global approaches for DNA methylation profiling. *Trends in Genetics,* 24, 231-237.

Brenner, S.E., Levitt, M. (2000) Expectations from structural genomics. *J Protein Science*, **9** (1), 197–200. Doi: 10.1110/ps.9.1.197. PMC 2144435. PMID 10739263.

Brenner, S.E (2001) A tour of structural genomics. *Nature Reviews Genetics*, ***2*** (10), 801–9. Doi: 10.1038/35093574. PMID 11584296.

Botstein, D., Fink, G.R. (2011) Yeast: an experimental organism for 21st Century biology. Genetics. *Genetics Society of America*, 189(3), 695–704.

Carè, A., Catalucci, D., Felicetti, F., Bonci, D., Addario, A., Gallo, P., Bang, M., Segnalini, P., Gu, Y., Dalton, N. et al. (2007) MicroRNA-133 controls cardiac hypertrophy. *Nature Medicine*, 13, 613-618.

Castoldi, M., Schmidt, S., Benes, V., Noerholm, M., Kulozik, A., Hentze, M., Muckenthaler, M. (2006) A sensitive array for micro RNA expression profiling (miChip) based on locked nucleic acids (LNA). *RNA* 12, 913-920.

Castoldi, M., Schmidt, S., Benes, V., Hentze, M., Muckenthaler, M. (2008) miChip: an array-based method for microRNA expression profiling using locked nucleic acid capture probes. *Nature Protocols,* 3, 321-329.

Carvalho, B., Ouwerkerk, E., Meijer, G.A. et al. (2004) High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J Clin Pathol*, *57, 644*–666.

Chaston, J., Douglas, A.E. (2012) Making the most of "omics" for symbiosis research. *The Biological Bulletin*, ***223*** (1), 21–9. PMC 3491573. PMID 22983030.

Chen, J.J. (2007) Key aspects of analyzing microarray gene-expression data. *Pharmacogenomics*, ***8*** (5), 473–82. Doi: 10.2217/14622416.8.5.473. PMID 17465711.

Cheung, A.N. (2007) Molecular targets in gynaecological cancers. *Journel of Pathology,* **39** (1), 26–45. Doi: 10.1080/00313020601153273. PMID 17365821.

Christians, U., Schmitz, V., Schoning, W. et al. (2008) Toxicodynamic therapeutic drug monitoring of immunosuppressants: promises, reality, and challenges. *Ther Drug Monit*, 30(2), 151–158.

Chu, C.Y., Rana, T. (2006) Translation repression in human cells by microRNA-induced gene silencing requires RCK/p54. *Journal PLOS biology,* 4, 1122-1136

Concepts of genetics (10th ed.) (2012) San Francisco: *Pearson Education*, ISBN 9780321724120.

Cong, F., Cheung, A.K., Huang, S.M.A. (2012) Chemical genetics-based target identification in drug discovery. *Annu Rev Pharmacol Toxicol*, 52(1), 57–78.

Complete genomes: Viruses. NCBI(2011).

Corey, D. (2007) RNA learns from antisense. *Nature Chemical biology,* 3, 8-11.

Culver, K.W., Labow, M.A. (2002) Genomics In Robinson R. Genetics. Macmillan Science Library. *Macmillan Reference,* USA ISBN 0028656067.

Darden, L., James, Tabery., (2010) Molecular Biology In Zalta EN. *The Stanford Encyclopedia of Philosophy (Fall 2010 ed.*), *12-20.*

Dalton, W.S., Friend, S.H. (2006) Cancer biomarkers – an invitation to the table. *Science*, 312(5777) 116, 5-8.

Da. B., SA, C., JM, C. (1997). Yeast as a Model Organism. Science. *NIH Public* Access, 277 (5330) 12, 59–60.

Dielbandhoesing, S.K., Zhang, H., Caro, L.H., vander Vaart, J.M., Klis, F.M., Verrips, C.T. et al. (1998) Specific cell wall proteins confer resistance to nisin upon yeast cells. *Appl Environ Microbiol*, 64(10) 40, 47–52.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A. et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genetics,* 38, 1378-1385.

Egan, C., Sridhar, S., Wigler, M., Hall, I. (2007) Recurrent DNA copy number variation in the laboratory mouse. *Nature Genetics,* 39, 1384-1389

Egger, **G**., Liang, G., Aparicio, A., Jones, P. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429, 457-463.

Eisen, J.A. (2007) Environmental shotgun sequencing: its potential and challenges for studying the hidden world of microbes. *An intrnational journal PLoS Biology,* **5** (3), e82. Doi: 10.1371/journal.pbio.0050082. PMC 1821061. PMID 17355177.

Eisen, J.A. (2012) Badomics words and the power and peril of the ome-meme. *GigaScience* **1,** (1), 6. *Doi*: 10.1186/2047-217X-1-6.

Emilien, G., Ponchon, M., Caldas, C., Isacson, O., Maloteaux, J.M. (2000) Impact of genomics on drug discovery and clinical medicine. QJM: *An intrnational journal of medicine,* 93(7), 391-423.

Ericson, E., Hoon, S., St Onge, R.P., Giaever, G., Nislow, C. (2010) Exploring Gene Function and Drug Action Using Chemogenomic Dosage Assays. *Elsevier* pp, 233–55.

Ernst, R., Kueppers, P., Stindt, J., Kuchler, K., Schmitt, L. (2010) Multidrug efflux pumps: substrate selection in ATP-binding cassette multidrug efflux pumps–first come, first served? *FEBS J. Blackwell Publishing Ltd,* 277(3) 54, 0–9.

Fahrig, F., Steinkamp-Zucht, A. (1996) Induction or suppression of SV40 amplification by genotoxic carcinogens, non-genotoxic carcinogens or tumor promoters. *Mutation Research,* 356, 217-224.

Farmer, P., Bonnefoi, H., Becette, V. et al. (2005) Identification of molecular apocrine breast tumours by microarray analysis. *J Oncogene* , *24, 4660*–4671.

Fehlbaum, P., Guihal, C., Bracco, L., Cochet, O. (2005) A microarray configuration to quantify expression levels and relative abundance of splice variants. *Nucleic Acids Research,* 33, e47-e58.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert,J. et al. (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene". *Nature, **260*** (5551), 500–7.

Fiers,W., Contreras, R., Haegemann, G., Rogiers, R., Vande Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., Ysebaert, M. (1978) Complete nucleotide sequence of SV40 DNA. *Nature, **273*** (5658), 113–20.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R. et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science, **269*** (5223), 496–512. *Bibcode*:*1995Sci...269..496F*. Doi: 10.1126/science.7542800. PMID *7542800.*

Francis, R.C. (2011) Epigenetics, the ultimate mystery of inheritance. *New York: W.W. Norton*, ISBN 9780393070057.

Frederiksen, C.M., Knudsen, S., Laurberg, S. et al. (2003) Classification of Dukes' B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol,129, 263*–271.

Gabriely, G., Wurdinger, T., Kesari, S., Esau, C., Burchard, J., Linsley, P., Krichevsky, A., (2008) MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators. *Molecular and Cellular Biochemistry,* 28, 5369-5380.

Gant, T. (2007) Novel and future applications of microarrays in toxicological research. *Expert opinion in drug metabolism and toxolcology* 3, 599-608.

Geary, R., Henry, S., Grillone, L. (2002) Fomivirsen: clinical pharmacology and potential drug interactions.*Clinical Pharmacokinetics,*41, 255-260.

Genome Project Statistics (2011) Entrez Genome Project. Retrieved .

Genome, n. (2008) Oxford English Dictionary (Third ed.). Oxford University Press.

Ghoshal, K., Bai, S. (2007) DNA methyltransferase as targets for cancer therapy. *Drugs of Today,* 43,395.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G. (1996) Life with 6000 genes". *Science, **274*** (5287), 546, 563–7. *Bibcode*:*1996Sci...274..546G*. *Doi*: *10.1126/science.274.5287.546*. PMID 8849441.

Goossens, N., Nakagawa, S., Sun, X., Hoshida, Y. (2015) Can Cancer biomarker discovery and validation. *Transl Cancer Res,* 4 (3), 256-269. Doi: 10.3978/j.issn.2218-676X.

Grimm, D., Streetz, K., Jopling, C., Storm, T., Pandey, K., Davis, C., Marion, P., Salazar, F., Kay, M. (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature*, 441, 537-541.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Ronneau, S.V.E. et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature. Nature Publishing Group,* 418 (6896), 387–91.

Greiner, T.C. (2004) mRNA microarray analysis in lymphoma and leukemia. *Cancer Treat Res* 121, 1–*12.*

Hammond, S. (2006) MicroRNA therapeutics: a new niche for antisense nucleic acids. *Trends in Molecular Medicine*, 12, 99-101.

Hardiman, G. (2004) Microarray platforms–comparisons and contrasts. *Journal of Pharmacogenomics, 5, 487*–502.

Harkin, D.P. (2000) Uncovering functionally relevant signaling pathways using microarray-based expression profiling. *The Oncologist*,5, 501–*507.*

Heinloth, A.N., Irwin, R.D., Boorman, G.A., et al. (2004) Gene expression profiling of rat livers reveals indicators of potential adverse effects. *Toxicol Sci,* 80(1), 193–202.

He, L., Hannon, G. (2004) Micro RNAs Small RNAs with a big role in gene regulation. *Nature Genetics*, 5, 522-531.

Henry, G., Liddell, R. S. (2015) An Intermediate Greek-English Lexicon, *γίγνομαι*".

Ho, C.H., Piotrowski, J., Dixon, S.J., Baryshnikova, A., Costanzo, M., Boone, C. (2011) Combining functional genomics and chemical biology to identify targets of bioactive compounds. *Current opinion in chemical biology*,15(1), 66–78.

Holley, R.W., Everett, G.A., Madison, J.T., Zamir, A. (1965) Nucleotide sequences in the yeast alanine transfer ribonucleic acid. *The Journal of Biological Chemistry*,**240** (5), 2122–8. PMID 14299636.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., Zamir, A. (1965) Structure of a ribonucleic acid. *Science*, ***147*** (3664), 1462–5. Bibcode*:*1965Sci...147.1462H. Doi: 10.1126/science.147.3664.1462. PMID 14263761.

Hong, H., Goodsaid, F., Shi, L., Tong, W. (2010) Molecular biomarkers: a US FDA effort. *Biomarkers in Medicine*, 4, 215-225.

Hotz, R.L. (2012) Here"s an Omical Tale: Scientists Discover Spreading Suffix. *Wall Street Journal ISSN* , 0099-9660.

Hugenholtz, P., Goebel, B.M., Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology*, **180** (18), 4765–74. PMC 107498. PMID 9733676.

Human gene number slashed (2012) BBC.

Hyman, E., Kauraniemi, P., Hautaniemi, S., Wolf, M., Mousses, S., Rozenblum, E., Ringner, M., Sauter, G., Monni, O., Elkahloun, A. et al. (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Research* .62, 6240-6245.

Ioannidis, J.P.A., Castaldi, P., Evangelou, E. (2010) A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. *J Natl Cancer Inst*, 102, 846-858

Ivanov, M., Barragan, I. (2014) Epigenetic mechanisms of importance for drug treatment. *Trends in Pharmacological Sciences,* Volume 35, Issue 8, Pages 384-396

Jacinto, F., Ballestar, E., Esteller, M. (2008) Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *BioTechniques*, 44, 35-39.

Jaskowiak, Pablo, A., Campello, Ricardo, J.G.B., Costa, Ivan, G. (2014) On the selection of appropriate distances for gene expression data clustering*". BMC Bioinformatics*, ***15*** (Suppl 2), S2. Doi: 10.1186/1471-2105-15-S2-S2

Jones, P., Baylin, S. (2007) The epigenomics of cancer. *An international journal of Cell*, 128, 683-392.

Joosse, S., vanBeers, E., Nederlof, P., (2007) Automated array-CGH optimized for archival formalin-fixed, paraffin-embedded tumor material. *BMC Cancer*, 7, 43-54.

Kallioniemi, O.P. (2001) Biochip technologies in cancer research. *Ann Med* 33, 142–147.

Kawasaki, E.S. (2006*)* The end of the microarray Tower of Babel: will universal standards lead the way? *J Biomol Tech, **17** (3), 200–6*. PMC 2291790. PMID 16870711.

Kaiser, O., Bartels, D., Bekel, T., Goesmann, A., Kespohl, S., Pühler, A., Meyer, F. (2003) Whole genome shotgun sequencing guided by bioinformatics pipelines--an optimized approach for an established technique. *Journal of Biotechnology,* ***106*** (2-3), 121–33. Doi: 10.1016/j.jbiotec.2003.08.008. PMID 14651855.

Khan, A.S. (2004) Genomics and microarray for detection and diagnostics. *Acta Microbiol Immunol Hung, 51, 463*–467.

Khor, T.O., Ibrahim, S., Kong, A.N. (2006) Toxicogenomics in drug discovery and drug development: potential applications and future challenges. *Pharm Res*,23(8),1659–1664.

Kreiner, T., Buck, K.T. (2005) Moving toward whole-genome analysis: a technology perspective. *Am J Health Syst Pharm, 62, 296*–305.

Kolaczkowski, M. Kolaczowska, A. Luczynski, J. Witek, S. Goffeau, A. (1998) In vivo characterization of the drug resistance profile of the major ABC transporters and other components of the yeast pleiotropic drug resistance network. *Microb Drug Resist*, 4(3) 1, 43–58.

Laird, P.W. (2010) Principles and challenges of genomewide DNA methylation analysis. Nature Reviews. *Genetics,* **11** (3), 191–203. doi:10.1038/nrg2732. PMID 20125086.

Lamb, J. (2007) The Connectivity Map: a new tool for biomedical research. *Nature reviews Cancer,* 1, 54-60.

Lamb, J., Crawford, E., Peck, D., Modell, J., Blat, I., Wrobel, M., Lerner, J., Brunet, J., Subramanian, A., Ross, K. et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules,genes, and disease. *Science*, 313, 1929-1935

Lander,E.S. (2011) Initial impact of the sequencing of the human genome. *Nature,* 470, 187-213

Latronico, M., Catalucci, D., Condorelli, G. (2008) MicroRNA and cardiac pathologies. *Physiological Genomics* 34, 239-242.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S., Kim, V. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO Journal,* 23, 4051-4060.

Lewis, B., Shih, I.H., Jones-Rhoades, M., Bartel, D., Burge, C. (2003) Prediction of Mammalian MicroRNA Targets. *An intrnational journalof Cell,* 115, 787-798.

Lockhart, D., Winzeler, E. (2000) Genomics, gene expression and DNA arrays. *Nature*, 405, 827-836.

Magic, Z., Radulovic, S., Brankovic-Magic, M. (2007) cDNA microarrays: identification of gene signatures and their application in clinical practice. *J BUON, 12 Suppl* 1, S39–44. PMID 17935276.

Marco, D. (2010) Metagenomics: Theory, Methods and Applications. *Caister Academic Press*, ISBN 978-1-904455-54-7.

Marco, D. (2011) Metagenomics: Current Innovations and Future Trends. *Caister Academic Press*, ISBN 978-1-904455-87-5.

Marsden, R.L., Lewis, T.A., Orengo, C.A. (2007) Towards a comprehensive structural coverage of completed genomes: a structural genomics viewpoint". *BMC Bioinformatics*, **8**, 86. Doi: 10.1186/1471-2105-8-86. PMC 1829165. PMID 17349043.

McCutcheon, J.P., von Dohlen, C.D. (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs". *Current Biology*, 21 (16) 13, 66–72. Doi: 10.1016/j.cub.2011.06.051. PMC 3169327. PMID 21835622.

McElheny, V. (2010) Drawing the map of life : inside the Human Genome Project. *New York NY: Basic Books*. ISBN 9780465043330.

Maxam, A.M., Gilbert, W. (1977) A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, ***74***(2), 560–4. Bibcode:1977PNAS...74..560M. Doi: 10.1073/pnas.74.2.560. PMC 392330. PMID 265521.

Melamed, D., Arava, Y. (2007) Genome-wide analysis of mRNA polysomal profiles with spotted DNA microarrays. *Methods in Enzymology*, 431, 177-201

Micknich, S. (2006) The connectivity map. *Nature Chemical biology*, 2, 663-664.

Microarrays Factsheet ".Retrieved(2007).

Minjou, W., Haegeman, G., Ysebaert, M., Fiers, W. (1972) Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, ***237*** (5350), *82–8*. Bibcode: 1972Natur.237...82J. Doi: 10.1038/237082a0. PMID 4555447.

National Human Genome Research Institute (2004) Dog Genome Assembled: Canine Genome Now Available to Research Community Worldwide. *Genome.gov*. Retrieved 2012-01-20.

National Human Genome Research Institute (2010) FAQ About Genetic and Genomic Science. *Genome.gov*. Retrieved 2011-12-03.

National Human Genome Research Institute (2010) A Brief Guide to Genomics. *Genome.gov. Retrieved* 2011-12-03.

Nielsen, R. (2010) Genomics. In search of rare human variants. *Nature, **467*** (7319), 1050–1. *Bibcode, Natur.467.1050N*. Doi: 10.1038/4671050a. PMID 20981085

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., O'Neal, C., (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proceedings of the National Academy of Sciences of the United States of America, **53** (5) 116, 1–8. Bibcode:1965PNAS...53.1161N*. Doi: 10.1073/pnas.53.5.1161. PMC 301388. *PMID* 5330357.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S. et al. (1986) Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. *Nature 322,* (6079), 572. Bibcode: 1986Natur.322..572O. Doi: 10.1038/322572a0.

Pasquinelli, A., Hunter, S., Bracht, J. (2005) MicroRNAs: a developing story. *Current Opinion in Genetics and Development,* 15, 200-205.

Paules, R. (2003) Phenotypic anchoring: Linking cause and effect. *Environ Health Perspect*, 111 (6), A338–A339.

Petretto, E., Sarwar, R., Grieve, I,. Lu, H., Kumaran, M., Muckett, P., Mangion, J., Schroen, B., Benson, M., ,Punjabi, P. et al. (2008) Integrated genomic approaches implicate osteoglycin (Ogn) in the regulation of left ventricular mass. *Nature Genetics,* 40, 546-552.

Pevsner, J. (2009) Bioinformatics and functional genomics (2nd ed.). Hoboken, N.J: *Wiley-Blackwell*, ISBN 9780470085851.

Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y. et al. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genetics,* 20, 207-211.

Pinkel, D., Albertson, D. (2005) Comparative genomic hybridization. *Annual reviews in genomics and human genetic,s* 6, 331-354.

Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani, R., Botstein, D., Borresen-Dale, A.L., Brown, P.O. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proceedings of the National Academy of Sciences of the United States of America,* 99, 12963-12968.

Powell, C.L., Kosyk, O., Ross, P.K., et al. (2006) Phenotypic anchoring of acetaminophen-induced oxidative stress with gene expression profiles in rat liver. *Toxicol Sci.,* 93(1), 213–222.

Rayburn, E., Zahng, R., (2008) Antisense, RNAi, and gene silencing strategies for therapy: Mission possible or impossible? *Drug Discovery Today,* 13, 514-521

Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews, T., Fiegler, H., Shapero, M., Carson, A., Chen, W., et al. (2006) Global variation in copy number in the human genome. *Nature,* 444, 444-454.

Ridd, K., Zhang, S.D., Edwards, R.E. et al. (2006) Association of gene expression with sequential proliferation, differentiation and tumor formation in murine skin. *Carcinogenesis,* 27(8), 1556–1566.

Roden, D.M., Altman, R.B., Benowitz, N.L., Flockhart, D.A., Giacomini, K.M., Johnson, J.A., et al. (2006) Pharmacogenomics: Challenges and Opportunities. *Ann Intern Med.,* 145, 749-757. doi:10.7326/0003-4819-145-10-200611210-00007

Rogers, B., Decottignies, A., Kolaczkowski, M., Carvajal, E., Balzi, E., Goffeau, A., (2001) The pleitropic drug ABC transporters from Saccharomyces cerevisiae. *J Mol Microbiol Biotechnol,* 3(2) 2 , 07–14.

Rotger, M., Saumoy, M., Zhang, K., Flepp, M., Sahli, R., Decosterd, L., Telenti, A., Study, S.H.C. (2007) Partial deletion of CYP2B6 owing to unequal crossover with CYP2B7. *Journal of Pharmacogenetics and Genomics,* 17, 885-890.

Sams-Dodd, F., (2005) Target-based drug discovery: is something wrong? *Drug Discov Today,* 15 10 (2) , 139–47.

Sanger, F., Nicklen, S., Coulson, A.R., (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America, **74*** (12), 5463–7. *Bibcode*:*1977PNAS...74.5463S*. Doi: 10.1073/pnas.74.12.5463. PMC 431765. PMID 271968.

Sanger, F. (1980). "Nobel lecture: Determination of nucleotide sequences in DNA" (PDF). *Nobelprize.org*, Retrieved 2010-10-18.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., Smith, M., (1977) Nucleotide sequence of bacteriophage phi X174 DNA. *Nature, **265*** (5596) 6, 87–95. *Bibcode*:*Natur.265..687S*. Doi: 10.1038/265687a0. PMID 870828.

Sawan, C., Vaissiere, T., Murr, R., Herceg, Z. (2008). Epigenetic drivers and genetic passengers on the road to cancer. *Mutation Research,* 642, 1-13

Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E., Davis, R.W. (1998) Microarrays biotechnology's discovery platform for functional genomics. *J Trends in Biotechnology,* 16, 301-306.

Schilling, E., Rehli, M. (2007) Global, comparative analysis of tissue-specific promoter CpG methylation *Genomics* 90, 314-323.

Scudellari, Megan, (2011) Data Deluge. *The Scientis,*. Retrieved 2013-01-04.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature,* 455, 58-63.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T. et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *The EMBO Journal, **5*** (9), 2043–2049. PMC 1167080. PMID 16453699.

Sonkoly, E., Wei, T., Janson, P., Sääf, A., Lundeberg, L., Tengvall-Linder, M., Norstedt, G., Alenius, H., Homey, B., Scheynius, A., et al. (2007) MicroRNAs: novel regulators involved in the pathogenesis of Psoriasis? *The Plos one journal* 11, e610.

Sonkoly, E., Ståhle, M., Pivarcsi, A. (2008) MicroRNAs: novel regulators in skin inflammation. *Journal of Clinical and experimental dermatology,* 33, 312-315.

Stoilov, P., Lin, C.H., Damoiseaux, R., Nikolic, J., Black, D. (2000) A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. *Proceedings of the National Academy of Sciences of the United States of America,* 32, 11218-11223.

Stokes, M.E., and McCourt, P. (2014) Towards personalized agriculture: what chemical genomics can bring to plant biotechnology. *Front. Plant Sci.,* **5**, 344. Doi:10.3389/fpls.2014.00344

Stonge, R., Schlecht, U., Scharfe, C., Evangelista, M. (2012) Forward Chemical Genetics in Yeast for Discovery of Chemical Probes Targeting Metabolism. *Journal of Molecules, **17***, 13098-13115

Suter, L., Babiss, L.E., Wheeldon, E.B. (2004) Toxicogenomics in predictive toxicology in drug development". *Chem. Biol., **11** (2),* 161–71. Doi: 10.1016/j.chembiol.2004.02.003. PMID 15123278.

SV. (2013) Genomics and proteomics in solving brain Kadakkuzha BM, Puthanveettil complexity. *Molecular bioSystems,* **9** (7), 1807–21.

Swinney, D.C., Anthony, J. (2011) How were new medicines discovered? Nature Reviews Drug Discovery. *Nature Publishing Group*, 10 (7), 507–19.

Tai, I.T., Dai, M., Owen, D.A. et al. (2005) Genome-wide expression analysis of therapy-resistant tumors reveals SPARC as a novel target for cancer therapy. *J Clin Invest, 115, 1492*–1502.

Tamarin, R.H., (2004) Principles of genetics (7 ed.). London: McGraw Hill. *ISBN 9780071243209*

Taylor, E., Gant, T. (2008) Emerging fundamental roles for non-coding RNA species in toxicology. *Toxicology,* 249, 34-39.

Thompson, J., Parker, J., Perou, C.M., Hammond, S.M. (2004) A custom microarray platform for analysis of microRNA gene expression. *Nature Methods,* 1, 47-53.

Thompson, J., Parker, J., Hammond, S. (2007) Microarray analysis of miRNA gene expression. *Methods in Enzymology*, 427, 107-119.

Turton, N.J., Judah, D.J., Riley, J., Davies, R., Lipson, D., Styles, J.A., Smith, A.G., Gant, T.W. (2001) Gene expression and amplification in breast carcinoma cells with intrinsic and acquired doxorubicin resistance.*J Oncogene*, 20, 1300-1306.

Twyman, R.M. (2004) SNP discovery and typing technologies for pharmacogenomics. *Curr Top Med Chem, 4, 1423*–1431.

Vander, Aart, Q.J., Agostoni-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P., Benit, P. (1992). The complete DNA sequence of yeast chromosome III. *Nature,* ***357*** (6373), 38–46. *Bibcode*:*1992Natur.357...38O*. Doi: 10.1038/357038a0. PMID 1574125.

Wang, E.J., Snyder, R.D., Fielden, M.R., et al. (2008) Validation of putative genomic biomarkers of nephrotoxicity in rats.*J Toxicology*, 246(2–3), 91–100.

Wang, G., Cooper, T. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Revews Genetics*, 8, 749-761

Wang, Y., Jatkoe, T., Zhang, Y., et al. (2004) Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol, 22, 1564*–1571.

Workman, P. (2004) Drug discovery strategies: technologies to accelerate translation from target to drug. *J Chemother, 16* (suppl 4), *13*–15.

Workman, P. (2003) The opportunities and challenges of personalized genome-based molecular therapies for cancer: targets, technologies, and molecular chaperones. *J Cancer Chemother Pharmacol*, *52* (suppl 1), *S45*–S56.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J., Hooper, S.D., Pati, A., Lykidis, A., Spring, S., Anderson, I.J., D'haeseleer, P., Zemla, A., Singer, M., Lapidus, A., Nolan, M., Copeland, A., Han, C., Chen, F., Cheng, J.F., Lucas, S., Kerfeld, C., Lang, E., Gronow, S., Chain, P., Bruce, D., Rubin, E.M., Kyrpides, N.C., Klenk, H.P., Eisen, J.A. (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, ***462*** (7276), 1056–60.

Yadav, S.P., (2007) The wholeness in suffix -omics, -omes, and the word om. *Journal of Biomolecular Techniques,* 18 (5), 277. PMC 2392988. PMID 18166670.

Yan, Fen., Timothy, J., Mitchison., Andreas., Bender., Daniel, W., Young & John, A., Tallarico, (2009) Multi-parameter phenotypic profiling: using cellular effects to characterize small-molecule compounds. *Nature Reviews Drug Discovery,* 8, 567-578 | Doi: 10.1038/nrd2876

Yao, S.N., Wilson, J.M., Nabel, E.G., Kurachi, S., Hachiya, H.L., Kurachi, K. (1991) Expression of human factor IX in rat capillary endothelial cells: Toward somatic gene therapy for hemophilia B. *Proc Natl Acad Sci,* 88, 8101–5.

Yue, G.H., Lo, L.C., Zhu, Z.Y., Lin, G., Feng, F. (2006) The complete nucleotide sequence of the mitochondrial genome of Tetraodon nigroviridis. *J DNA Sequence, **17** (2), 115–21. doi:10.1080/10425170600700378.* PMID 17076253.

Zimmer, C., (2009) Scientists Start a Genomic Catalog of Earth's Abundant Microbes. *The New York Times, ISSN 0362-4331.*

Zhang, S.D., Gant, T. (2014) A statistical framework for the design of microarray experiments and effective detection of differential gene expression. *J Bioinformatics* ,20, 2821-2828.

Zhang, S., Gant, T. (2008) A simple and robust method for connecting small-molecule drugs using gene expression signatures. *Bmc Bioinformatics,* 9, 258-268