PR14_Pomen genomskih raziskav za medicino in farmacevtsko industrijo

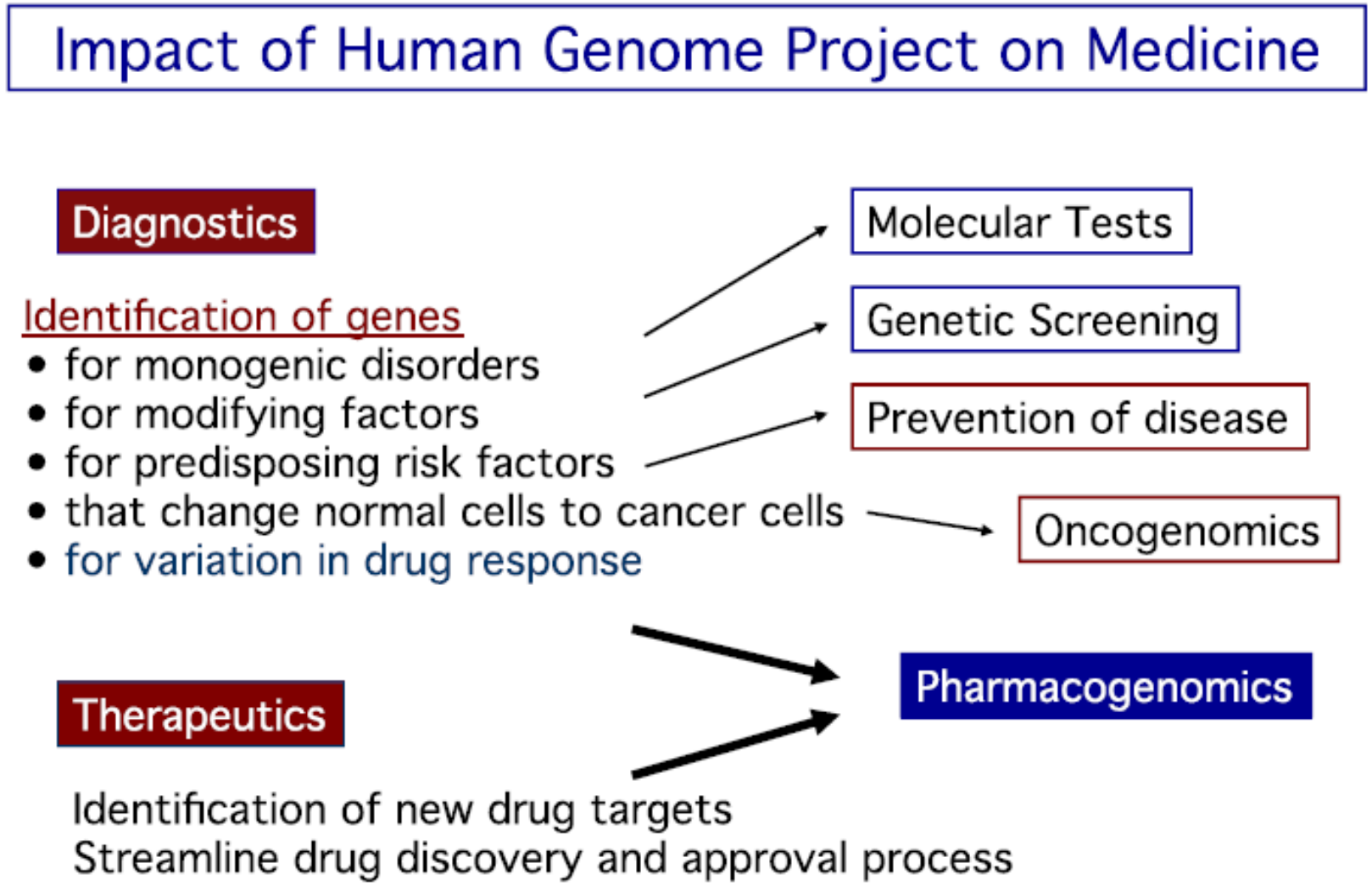


Figure 1. Expected impact of the human genome project on medicine, as envisioned in 2003.

Genomska medicina

Medicine in the post-genomic era

Genome Medicine publishes peer-reviewed research articles, new methods, software tools, reviews and comment articles in all areas of ***medicine studied from a post-genomic perspective***. Areas covered include, but are not limited to, ***disease genomics (including genome-wide association studies and sequencing-based studies)***, ***disease epigenomics***, ***pathogen and microbiome genomics***, ***immunogenomics***, ***translational genomics***, ***pharmacogenomics and personalized medicine***, ***proteomics and metabolomics in medicine***, ***systems medicine***, and ethical, legal and social issues.

BMC Medical Genomics is an open access, peer-reviewed journal that considers articles on all aspects of ***functional genomics***, ***genome structure***, ***genome-scale population genetics***, ***epigenomics***, ***proteomics***, ***systems analysis***, and ***pharmacogenomics*** in relation to human health and disease.

Genomic Medicine: New Frontiers and New Challenges

Maria D. Pasic,[1,2] Sara Samaan,[3,4] and George M. Yousef[1,3,4*]

Table 1. The potential applications of genomic medicine.

1. Examination of variation among healthy individuals
2. Understanding disease risk, susceptibility, and etiology
3. Disease prevention
4. Early diagnosis at the preclinical stage (to help modify disease course, e.g., diet and exercise in diabetes)
5. Identification of new disease-related mutations
6. Accurate diagnosis of challenging cases with inconclusive results for clinical parameters
7. Accurate disease classification based on molecular signature
8. Health management (prognosis and predictive markers)
9. Identification of personal drug-related profile
10. Selection of patients for clinical trials
11. Monitoring disease status (e.g., recurrence)
12. Monitoring tumor evolution in response to treatment
13. Developing new targeted therapies, especially for metastatic tumors refractory to treatment

Table 2. Strategies to increase the cost-effectiveness of genomic screening strategies.

1. Target screening to a high-risk population
2. Focus on chronic diseases
3. Focus screening on diseases with known interventions
4. Identify molecular changes directly related to short-term benefits
5. Limit screening to diseases with a higher prevalence
6. Interpret data in a clinical context (e.g., environmental factors)

Table 3. Potential challenges of the era of genomic medicine.

1. Data interpretation and extraction of actionable items
2. Guidelines for implementing new molecular testing
3. Cost-effectiveness of molecular testing
4. Patient heterogeneity and ethnic variation
5. Test optimization and standardization
6. The risk of incidental findings and false-positive results
7. Need for training and teamwork efforts
8. Between-platform variation in results

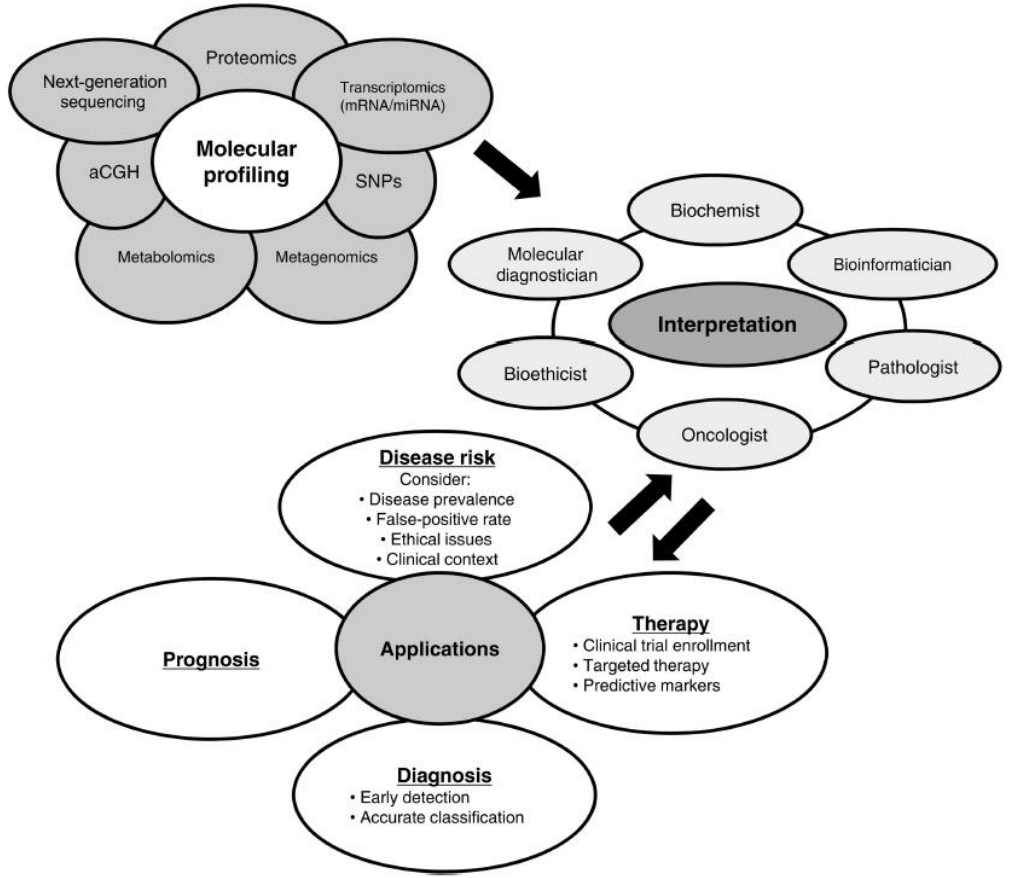


Fig. 1. Schematic showing the multistep process of translating genomics into the clinical setting.
Multilevel molecular analysis should be integrated and interpreted by a multidisciplinary team that will decide on the proper clinical implementation. Genomic markers can be used to assess disease risk or serve as diagnostic or prognostic markers; they can also have therapeutic applications. aCGH, array-comparative genomic hybridization.

New insights into the genetics of addiction

Ming D. Li and Margit Burmeister

Abstract | Drug addiction is a common brain disorder that is extremely costly to the individual and to society. Genetics contributes significantly to vulnerability to this disorder, but identification of susceptibility genes has been slow. Recent genome-wide linkage and association studies have implicated several regions and genes in addiction to various substances, including alcohol and, more recently, tobacco. Current efforts aim not only to replicate these findings in independent samples but also to determine the functional mechanisms of these genes and variants.

Table 1 | **Candidate genes associated with at least one drug addiction that are verified by meta-analysis**

Gene symbol	Gene name	Biological function	Chromosomal location	Drug (phenotype)	Evidence from knockout animal model	References for meta-analysis
5HTT (also known as *SERT*)	5-hydroxytryptamine transporter	Neurotransmitter transport	17q11.1–q12	Alcohol (i, d, c), cocaine (d, c), heroin (d), methamphetamine (d), nicotine (d)	Increased sensitivity to alcohol-induced sedation and hypnosis; motor-coordination deficits in response to alcohol; reduced gross alcohol intake; altered behavioural responses to cocaine and alcohol	93,94
CYP2A6	Cytochrome P450, family 2, subfamily A, polypeptide 6	Oxidation reduction	19q13.2	Alcohol (d), nicotine (i, d, c)	None	94
DAT1	Dopamine transporter	Neurotransmitter transport	5p15.3	Alcohol (d, c), cocaine (d), heroin (d), methamphetamine (d), nicotine (i, d, c)	Reduced alcohol preference in female mice; cocaine-induced stereotypy (repetitive behaviour)	94,95
DRD2	Dopamine receptor 2	Synaptic transmission, dopaminergic	11q23.1–q23.2	Alcohol (d, c), cocaine (d), heroin (d), nicotine (i, d, c)	Alcohol preference and alcohol-induced ataxia; reduced rate of high-dose self-administration of cocaine	96–99
IL10	Interleukin-10	Cytokine activity	1q31–q32	Alcohol (d)	None	100
BDNF	Brain-derived neurotropic factor	Regulation of synaptic plasticity	11p13	Alcohol (i, d, c), nicotine (d), cocaine (d), methamphetamine (d)	Increased alcohol intake; increased preference for cocaine	101

The genes reported here have one or more variants that have been associated with more than one addiction. This table may only represent a partial list of genes that have been investigated for association with addictions. c, cessation or withdrawal; d, abuse or dependence; i, initiation.

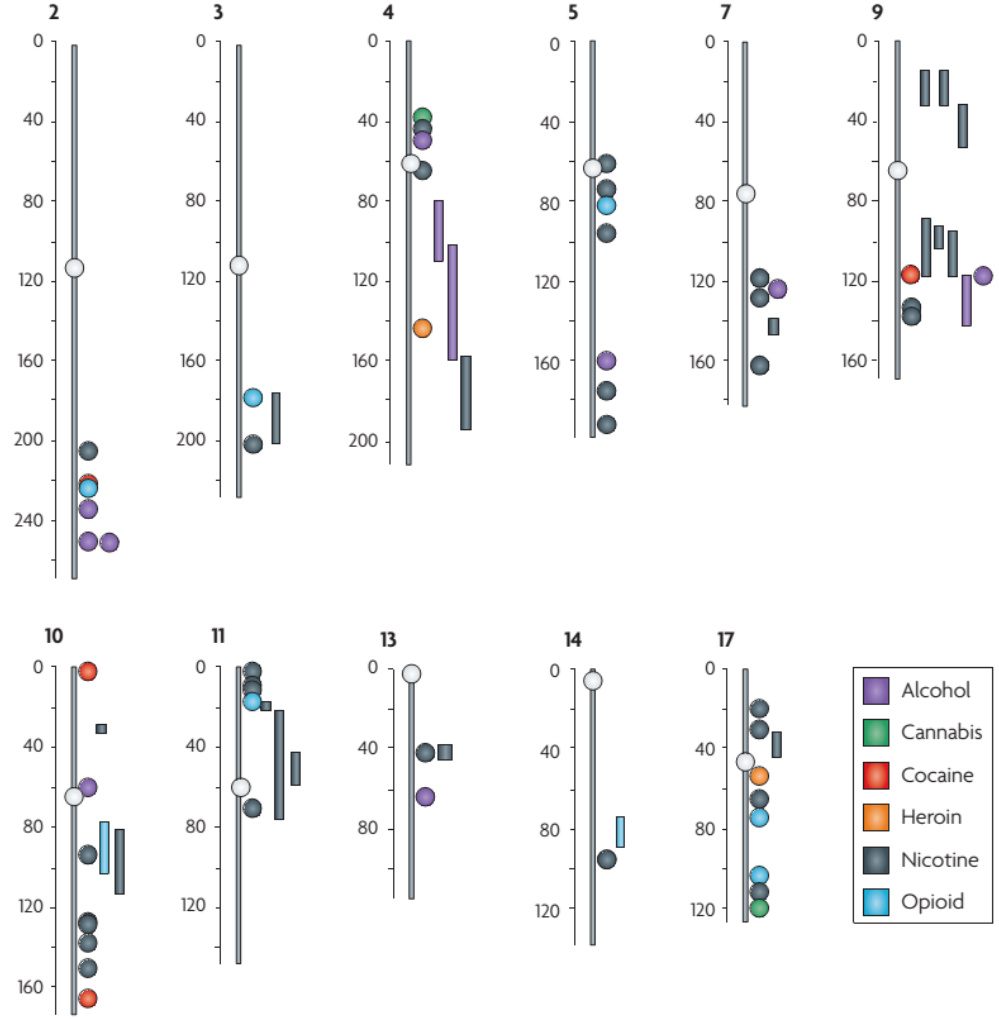


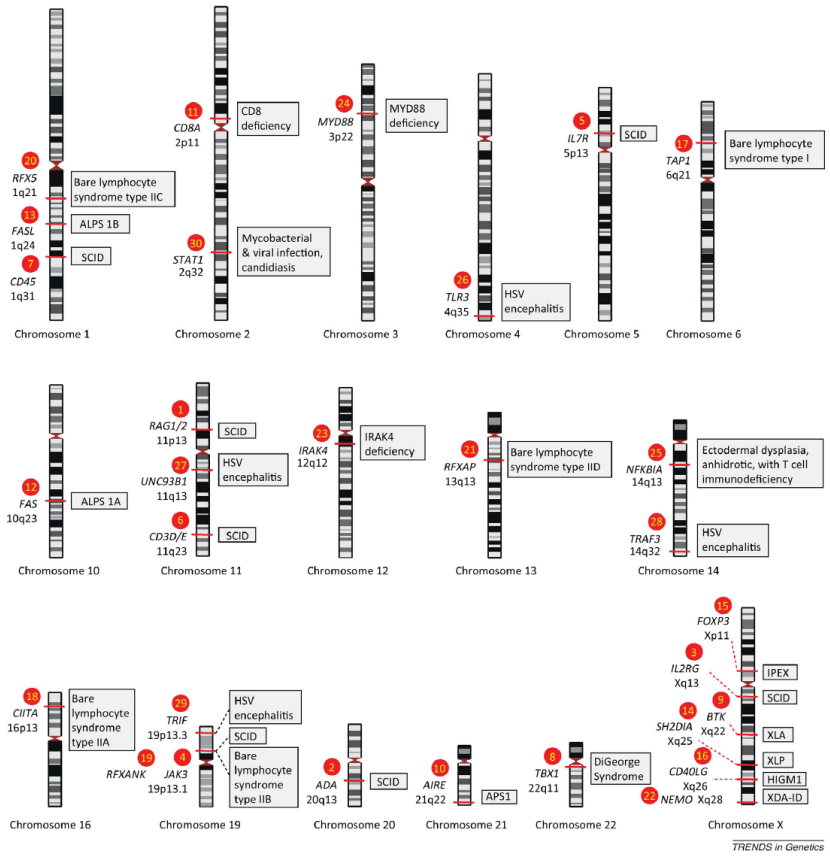
Figure 1 | **Chromosomal locations of peaks or intervals for addiction to multiple substances.** Summary of chromosomal locations of nominated peaks or intervals for addictions to alcohol, cannabis, cocaine, heroin, nicotine and opoids. Each linkage is shown with either a colour-filled circle or a rectangle, representing a reported linkage peak or region, respectively. 'Significant' or 'suggestive' linkage was determined by independent studies on at least two substances of abuse. The determination of significant or suggestive linkage at each linkage peak or interval was based on the rigorous criteria proposed by Lander and Kruglyak[34], which define a logarithm of odds (LOD) score greater than 3.6 or a *p* value less than 2.2×10^{-5} as a significant linkage, and an LOD score greater than 2.2 but less than 3.6 or a *p* value less than 7.4×10^{-4} as a suggestive linkage. Each colour represents a type of abused substance. The unit for each chromosome is cM.

Genomic modulators of the immune response☆

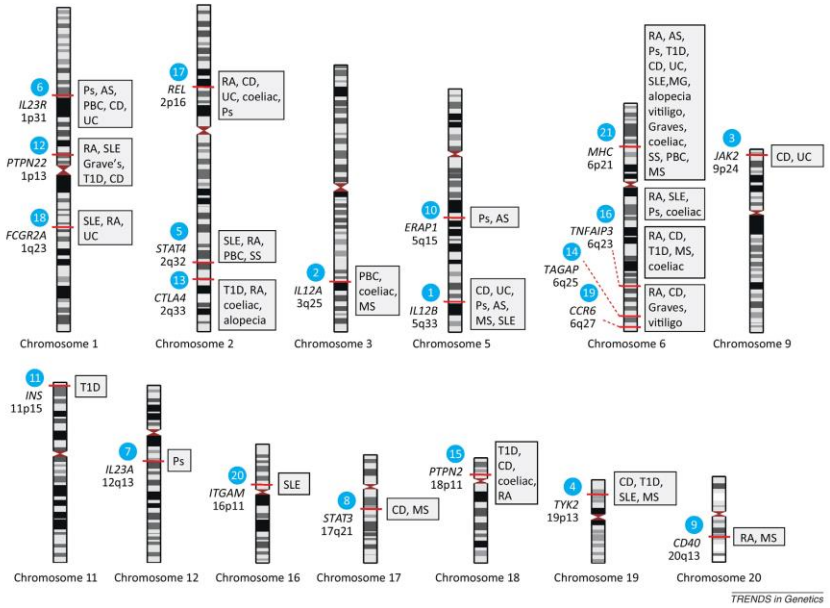
Julian C. Knight

Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK

Our understanding of immunity has historically been informed by studying heritable mutations in both the adaptive and innate immune responses, including primary immunodeficiency and autoimmune diseases. Recent advances achieved through the application of genomic and epigenomic approaches are reshaping the study of immune dysfunction and opening up new avenues for therapeutic interventions. Moreover, applying genomic techniques to resolve functionally important genetic variation between individuals is providing new insights into immune function in health. This review describes progress in the study of rare variants and primary immunodeficiency diseases arising from whole-exome sequencing (WES), and discusses the application, success, and challenges of applying genome-wide association studies (GWAS) to disorders of immune function and how they may inform more rational use of therapeutics. In addition, the application of expression quantitative-trait mapping to immune phenotypes, progress in understanding MHC disease associations, and insights into epigenetic mechanisms at the interface of immunity and the environment are reviewed.



Examples of Mendelian traits involving immune defects. Red filled numbered circles correspond to those shown in the overview of the immune system (Figure 1). The implicated gene (given in italics) and chromosomal position are shown together with the associated phenotype (grey shaded box). Note that the genetic variants involved range from point mutations (single-nucleotide variants) to large structural variants.
SCID: severe combined immunodeficiency syndrome



Examples of common autoimmune diseases where loci have been reported by GWAS.
Blue filled numbered circles correspond to those shown in the overview of the immune system. This is not an exhaustive list and illustrates the overlap seen at many associated loci for different autoimmune diseases and how genes encoding different components of pathways may be associated with a given disease using data from the National Human Genome Research Institute (NHGRI) Catalogue of Published Genome Wide Association Studies (GWAS) where full details can be found. It should also be noted that, in the majority of cases, the genes listed are candidates based on the genomic locus that was associated, and causality has not been established except in a small number of instances.

Vloga nekodirajočih RNA pri nastanku bolezni

Non-coding RNAs in human disease

Manel Esteller

Abstract | The relevance of the non-coding genome to human disease has mainly been studied in the context of the widespread disruption of microRNA (miRNA) expression and function that is seen in human cancer. However, we are only beginning to understand the nature and extent of the involvement of non-coding RNAs (ncRNAs) in disease. Other ncRNAs, such as PIWI-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs), transcribed ultraconserved regions (T-UCRs) and large intergenic non-coding RNAs (lincRNAs) are emerging as key elements of cellular homeostasis. Along with microRNAs, dysregulation of these ncRNAs is being found to have relevance not only to tumorigenesis, but also to neurological, cardiovascular, developmental and other diseases. There is great interest in therapeutic strategies to counteract these perturbations of ncRNAs.

Table 1 | **Types of ncRNAs***

Name	Size	Location	Number in humans	Functions	Illustrative examples	Refs
Short ncRNAs						
miRNAs	19–24 bp	Encoded at widespread locations	>1,424	Targeting of mRNAs and many others	miR-15/16, miR-124a, miR-34b/c, miR-200	3–8
piRNAs	26–31bp	Clusters, intragenic	23,439	Transposon repression, DNA methylation	piRNAs targeting *RASGRF1* and LINE1 and IAP elements	13–19
tiRNAs	17–18bp	Downstream of TSSs	>5,000	Regulation of transcription?	Associated with the *CAP1* gene	37
Mid-size ncRNAs						
snoRNAs	60–300 bp	Intronic	>300	rRNA modifications	U50, SNORD	20–22
PASRs	22–200 bp	5′ regions of protein-coding genes	>10,000	Unknown	Half of protein-coding genes	10
TSSa-RNAs	20–90 bp	–250 and +50 bp of TSSs	>10,000	Maintenance of transcription?	Associated with *RNF12* and *CCDC52* genes	35
PROMPTs	<200 bp	–205 bp and –5 kb of TSSs	Unknown	Activation of transcription?	Associated with *EXT1* and *RBM39* genes	36
Long ncRNAs						
lincRNAs	>200 bp	Widespread loci	>1,000	Examples include scaffold DNA–chromatin complexes	*HOTAIR*, *HOTTIP*, *lincRNA-p21*	2,28–30
T-UCRs	>200 bp	Widespread loci	>350	Regulation of miRNA and mRNA levels?	uc.283+, uc.338, uc160+	31–34
Other lncRNAs	>200 bp	Widespread loci	>3,000	Examples include X-chromosome inactivation, telomere regulation, imprinting	*XIST*, *TSIX*, TERRAs, *p15AS*, *H19*, *HYMAI*	2,23–25

*There is not necessarily a clear delineation between classes of non-coding RNA (ncRNA); for example, X-inactivation specific transcript (*XIST*) and its antisense transcript *TSIX* could be considered as large intergenic non-coding RNAs (lincRNAs). In the 'Location' column, '–' represents the number of base pairs upstream of the transcription start site (TSS) and '+' represents the number of base pairs downstream of the TSS. *CAP1*, CAP, adenylate cyclase-associated protein 1; *CCDC52*, coiled-coil domain containing 52 (also known as *SPICE1*); *EXT1*, exostosin 1; *HOTAIR*, homeobox (HOX) transcript antisense RNA; *HOTTIP*, HOXA distal transcript antisense RNA; *HYMAI*, hydatidiform mole associated and imprinted; IAP, intracisternal A-particle; lncRNA, long non-coding RNA; miRNAs, microRNAs; piRNAs, PIWI-interacting RNAs; PASRs, promoter-associated small RNAs; PROMPTs, promoter upstream transcripts; *RASGRF1*, RAS-protein-specific guanine nucleotide-releasing factor 1; *RBM39*, RNA-binding motif protein 39; *RNF12*, ring finger protein 12 (also known as *RLIM*); snoRNAs, small nucleolar RNAs; TERRAs, telomeric repeat containing RNAs; tiRNAs, transcription initiation RNAs; TSSa-RNAs, TSS-associated RNAs; T-UCRs, transcribed ultraconserved regions.

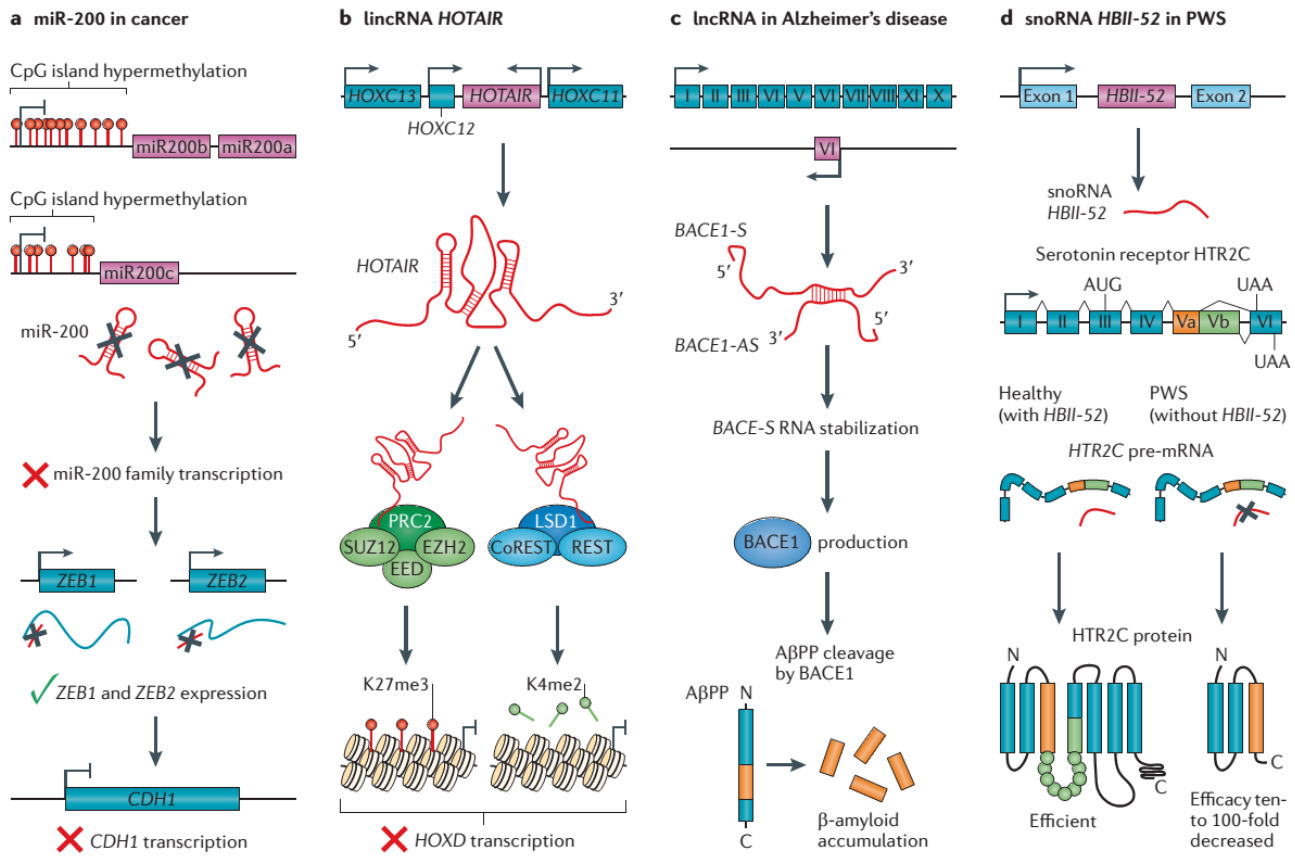


Figure 2 | **Examples of roles of ncRNAs in disease pathophysiology. a** | miR-200 is an example of a microRNA (miRNA) whose role in cancer is well characterized. Alterations in the epigenetic regulation of the miR-200 family are involved in epithelial-to-mesenchymal transition (EMT) in cancer. Specifically, CpG island hypermethylation-associated silencing of these miRNAs in human tumours causes an upregulation of the zinc finger E-box-binding homeobox (HOX) 1 (ZEB1) and ZEB2 transcriptional repressors, which, in turn, leads to a downregulation of E-cadherin (CDH1) — these are changes that promote EMT. **b** | The lincRNA HOX transcript antisense RNA (*HOTAIR*) is involved in polycomb retargeting across the genome. *HOTAIR* expression is increased in transformed cells and induces a genome-wide promoter re-occupancy by polycomb- and H3K27-trimethylation-associated silencing of target genes, such as the HOX protein *HOXD*. The effect of these changes is to increase cancer invasiveness. **c** | lncRNA targeting of β-secretase 1 (BACE1) has a role in the pathophysiology of Alzheimer's disease. An antisense lncRNA, *BACE1-AS*, regulates the expression of the sense *BACE1* gene (labelled *BACE1-S* in the figure) through the stabilization of its mRNA. *BACE1-AS* is elevated in Alzheimer's disease, increasing the amount of BACE1 protein and, subsequently, the production of β-amyloid peptide. **d** | The role of the snoRNA in Prader–Willi syndrome (PWS). The loss of the snoRNA in PWS changes the alternative splicing of the serotonin receptor *HTR2C* precursor mRNA (pre-mRNA), resulting in a protein with reduced function. AβPP, amyloid-β precursor protein; CoREST, REST corepressor.

Table 2 | **Illustrative list of ncRNAs disrupted by either genetic or epigenetic means in cancer**

Name	Class	Disruption	Consequence	Cancer type
miR-124a	miRNA	CpG island hypermethylation	CDK6 overexpression	Colon, gastric, haematological
miR-34b and miR-34c	miRNA	CpG island hypermethylation	Metastasis	Many different tumour types
miR-148a	miRNA	CpG island hypermethylation	Metastasis	Colon, melanoma, breast
miR-9	miRNA	CpG island hypermethylation	Metastasis	Colon, melanoma, head and neck
miR-200c	miRNA	CpG island hypermethylation	EMT	Colon, breast, lung
miR-141	miRNA	CpG island hypermethylation	EMT	Colon, breast, lung
miR-205	miRNA	CpG island hypermethylation	EMT	Bladder
miR-196b	miRNA	CpG island hypermethylation	Unknown	Gastric
miR-129-2	miRNA	CpG island hypermethylation	SOX2 overexpression	Colon, endometrial, gastric
miR-137	miRNA	CpG island hypermethylation	CDC42 overexpression	Colon, head and neck
Uc.160+	T-UCR	CpG island hypermethylation	Unknown	Colon, breast, lung
Uc.283+A	T-UCR	CpG island hypermethylation	Cell survival, mitosis	Colon, breast, lung
Uc.346+	T-UCR	CpG island hypermethylation	Unknown	Colon, breast, lung
Uc.21	T-UCR	Mutation	Unknown	Epithelial tumours, leukaemia
Uc.72	T-UCR	Mutation	Unknown	Epithelial tumours, leukaemia
miR-151	miRNA	Gene amplification	Metastasis	Hepatocellular carcinoma
miR-517c and miR-520g	miRNA	Gene amplification	WNT upregulation	Neuroectodermal brain tumors
miR-106b-25	miRNA	Gene amplification	p21 and BIM depletion	Oesophageal adenocarcinoma
miR-15 and miR-16	miRNA	Genomic deletion	BCL2 overexpression	Haematological
U50	snoRNA	Genomic deletion	Increase growth	Breast
Uc.159	T-UCR	Genomic deletion	Unknown	Epithelial tumours, leukaemia

BCL2, B cell CLL/lymphoma 2; BIM, also known as BCL2L11; CDC42, cell-division cycle 42; CDK6, cyclin-dependent kinase 6; EMT, epithelial-to-mesenchymal transition; miRNA, microRNA; snoRNA, small nucleolar RNA; ncRNA, non-coding RNA; T-UCR, transcribed ultraconserved region.

Table 3 | **Illustrative list of ncRNAs that are disrupted in non-tumoural disorders**

Disease	Involved ncRNAs	ncRNA type	Refs
Spinal motor neuron disease	miR-9	miRNA	87
Spinocerebellar ataxia type 1	miR-19, miR-101, miR-100	miRNA	88
Amyotropic lateral sclerosis	miR-206	miRNA	86
Arrhytmia and hypertension	miR-1	miRNA	98
Atheromatosis and atherosclerosis	miR-10a, miR-145, mR-143 and miR-126	miRNA	100–102
Atheromatosis and atherosclerosis	Circular ncRNA linked to the CDKN2A locus	lncRNA	119
Cardiac hypertrophy	miR-21	miRNA	144
Rett's syndrome	miR-146a, miR-146b, miR-29 and miR-382	miRNA	108,109
5q syndrome	miR-145 and miR-146a	miRNA	106
ICF syndrome	miR-34b, miR-34c, miR-99b, let-7e and miR-125a	miRNA	107
Crohn's disease	miR-196	miRNA	110
Prader–Willi and Angelman syndromes	snoRNA cluster at 15q11–q13 imprinted locus	snoRNA	114–116
Beckwith–Wiedeman syndrome	lncRNAs *H19* and *KCNQ1OT1*	lncRNA	145
Uniparental disomy 14	snoRNA cluster at 14q32.2 imprinted locus	snoRNA	145
Silver–Russell syndrome	lncRNA *H19*	lncRNA	145
Silver–Russell syndrome	miR-675	miRNA	145
McCune–Albright syndrome	lncRNA *NESP-AS*	lncRNA	145
Deafness	miR-96	miRNA	111
Alzheimer's disease	miR-29, miR-146 and miR-107	miRNA	89–91
Alzheimer's disease	ncRNA antisense transcript for *BACE1*	lncRNA	112
Parkinson's disease	miR-7, miR-184 and let-7	miRNA	82
Down's syndrome	mir-155 and miR-802	miRNA	83
Idiopathic neurodevelopmental disease	T-UCRs uc.195, uc.392, uc.46 and uc.222	T-UCR	113
Rheumatoid arthritis	miR-146a	miRNA	147
Transient neonatal diabetes mellitus	lncRNA *HYMAI*	lncRNA	148
Pseudohypoparathyroidism	lncRNA *NESP-AS*	lncRNA	146

BACE1, β-secretase 1; CDKN2A, cyclin-dependent kinase inhibitor 2A; *HYMAI*, hydatidiform mole associated and imprinted; ICF syndrome, immunodeficiency, centromeric region instability and facial anomalies syndrome; KCNQ1OT1, KCNQ1 opposite strand/antisense transcript 1; lncRNA, long non-coding; miRNA, microRNA; NESP, also known as GNAS; NESP-AS, NESP antisense; ncRNA, non-coding RNA; snoRNA, small nucleolar RNA; T-UCR, transcribed ultraconserved region.

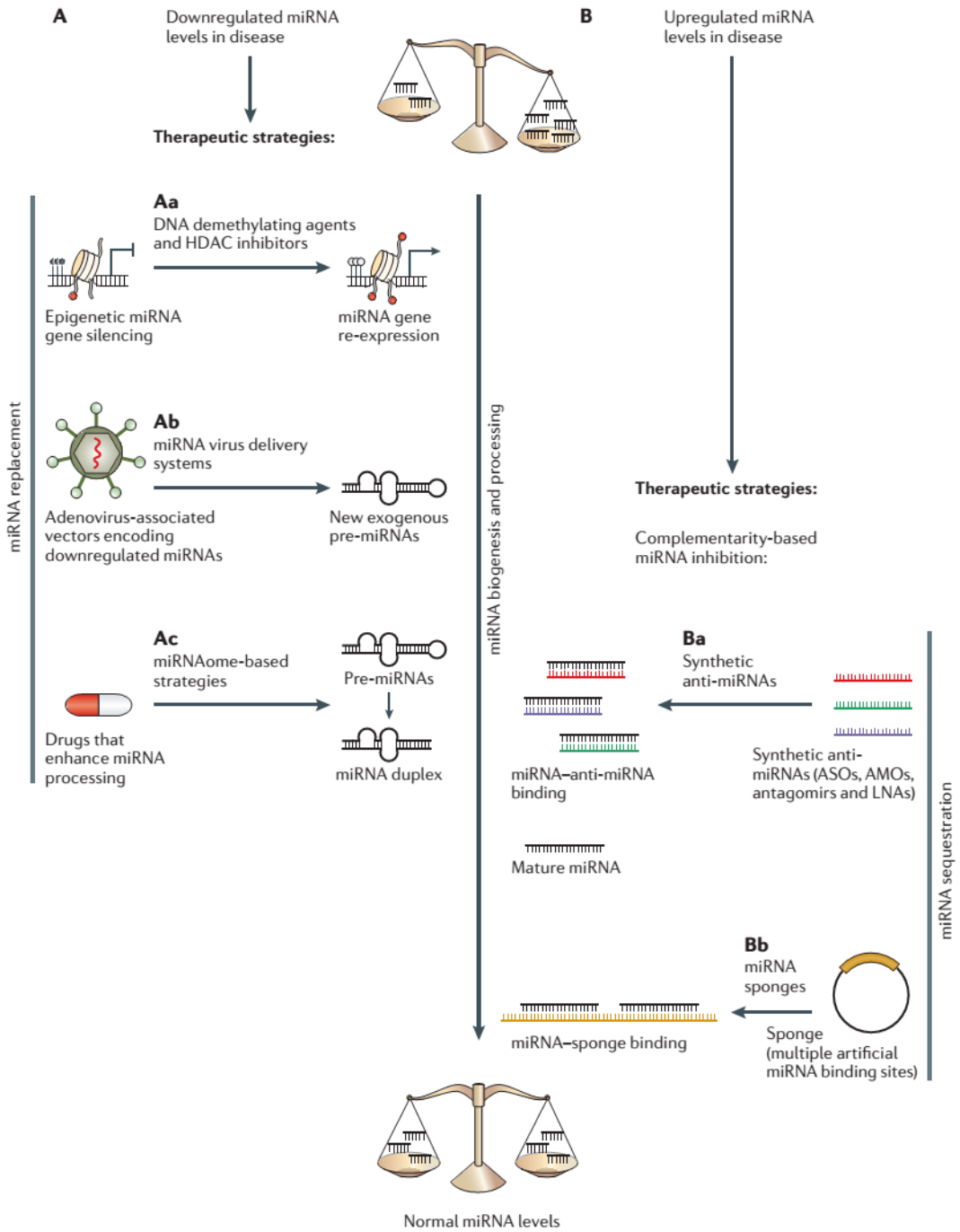


Figure 3 | **Therapies targeting miRNAs in human disease.** Aberrant microRNA (miRNA) levels are associated with multiple diseases. **A** | Strategies to increase miRNA levels include the following. **Aa** | Epigenetic drug treatments for reactivating the transcription of silenced miRNA genes, using DNA demethylating agents and histone deacetylase (HDAC) inhibitors. **Ab** | Replacement of miRNAs using virus delivery systems: for example, adenovirus-associated vectors (AVV) containing genes that code for downregulated miRNAs. **Ac** | miRNAome-based strategies based on drugs such as enoxacin, which enhance miRNA production by binding to TAR RNA-binding protein 2 (TARBP2), an integral component of a DICER1-containing miRNA-processing complex. **B** | Conversely, upregulated miRNAs can be sequestered by means of therapies based on base-pair complementarity, which include the following. **Ba** | Synthetic antisense oligonucleotides (ASOs) or modified ASOs that incorporate chemical groups to improve the stability and efficacy, such as anti-miRNA oligonucleotides (AMOs), antagomirs or locked nucleic acids (LNAs). **Bb** | miRNA sponges: vectors that contain multiple artificial miRNA binding sites that act as sponges for the cognate miRNA, preventing its association with endogenous targets.

Sistemska biologija (biološka omrežja)
bolezni pri človeku

Interactome Networks and Human Disease

Marc Vidal,[1,2,*] Michael E. Cusick,[1,2] and Albert-László Barabási[1,3,4,*]
[1]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[3]Center for Complex Network Research (CCNR) and Departments of Physics, Biology and Computer Science, Northeastern University, Boston, MA 02115, USA
[4]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: marc_vidal@dfci.harvard.edu (M.V.), alb@neu.edu (A.-L.B.)
DOI 10.1016/j.cell.2011.02.016

Complex biological systems and cellular networks may underlie most genotype to phenotype relationships. Here, we review basic concepts in network biology, discussing different types of interactome networks and the insights that can come from analyzing them. We elaborate on why interactome networks are important to consider in biology, how they can be mapped and integrated with each other, what global properties are starting to emerge from interactome network models, and how these properties may relate to human disease.

Global Disease Networks

One of the main predictions derived from the hypothesis that human disorders should be viewed as perturbations of highly interlinked cellular networks is that diseases should not be independent from each other, but should instead be themselves highly interconnected. Such potential cellular network-based dependencies between human diseases has led to the generation of various global disease network maps, which link disease phenotypes together if some molecular or phenotypic relationships exist between them. Such a map was built using known gene-disease associations collected in the OMIM database (Goh et al., 2007), where nodes are diseases and two diseases are linked by an edge if they share at least one common gene, mutations in which are associated with these diseases. In the obtained disease network more than 500 human genetic disorders belong to a single interconnected main giant component, consistent with the idea that human diseases are much more connected to each other than anticipated. The flipside of this representation of connectivity is a network of disease-associated genes linked together if mutations in these genes are known to be responsible for at least one common disorder. Providing

Network Perturbations by Pathogens

Pathogens, particularly viruses, have evolved sophisticated mechanisms to perturb the intracellular networks of their hosts to their advantage. As obligate intracellular pathogens, viruses must intimately rewire cellular pathways to their own ends to maintain infectivity. Since many virus-host interactions happen at the level of physical protein-protein interactions, systematic maps capturing viral-host physical protein-protein interactions, or "virhostome" maps, have been obtained using Y2H for Epstein-Barr virus (Calderwood et al., 2007), hepatitis C virus (de Chassey et al., 2008), several herpesviruses (Uetz et al., 2006), influenza virus (Shapira et al., 2009) and others (Mendez-Rios and Uetz, 2010), and by co-AP/MS methodologies for HIV (Jäger et al., 2010). An eminent goal is to find perturbations in network properties of the host network, properties that would not be made evident by small-scale investigations focused on one or a handful of viral proteins. For instance, it has been found several times now that viral proteins preferentially target hubs in host interactome networks (Calderwood et al., 2007; Shapira et al., 2009). The many host targets identified in virhostome screens are now getting biologically validated by RNAi knock-down and transcriptional profiling, leading to detailed maps of the interactions underlying viral-host relationships (Shapira et al., 2009).

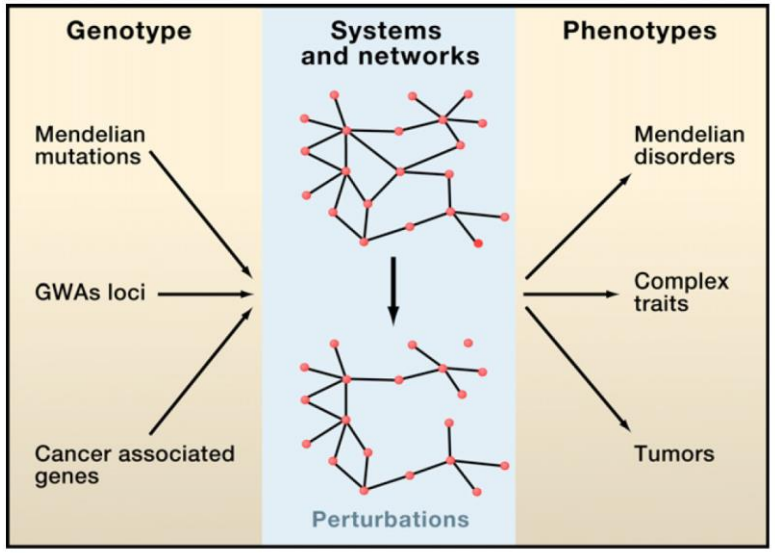


Figure 1. Perturbations in Biological Systems and Cellular Networks May Underlie Genotype-Phenotype Relationships
By interacting with each other, genes and their products form complex cellular networks. The link between perturbations in network and systems properties and phenotypes, such as Mendelian disorders, complex traits, and cancer, might be as important as that between genotypes and phenotypes.

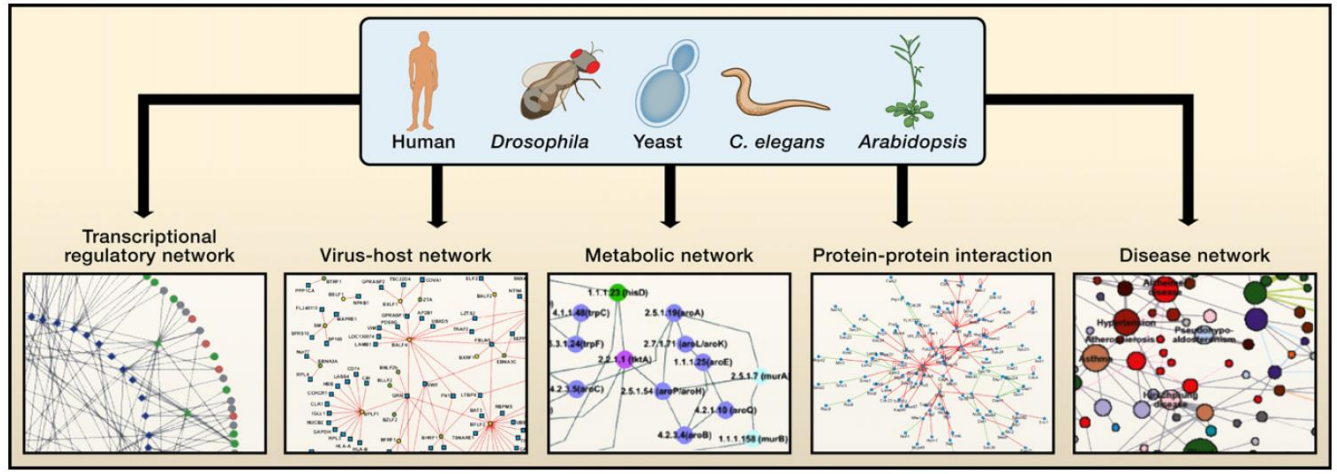


Figure 2. Networks in Cellular Systems
To date, cellular networks are most available for the "super-model" organisms (Davis, 2004) yeast, worm, fly, and plant. High-throughput interactome mapping relies upon genome-scale resources such as ORFeome resources. Several types of interactome networks discussed are depicted. In a protein interaction network, nodes represent proteins and edges represent physical interactions. In a transcriptional regulatory network, nodes represent transcription factors (circular nodes) or putative DNA regulatory elements (diamond nodes); and edges represent physical binding between the two. In a disease network, nodes represent diseases, and edges represent gene mutations of which are associated with the linked diseases. In a virus-host network, nodes represent viral proteins (square nodes) or host proteins (round nodes), and edges represent physical interactions between the two. In a metabolic network, nodes represent enzymes, and edges represent metabolites that are products or substrates of the enzymes. The network depictions seem dense, but they represent only small portions of available interactome network maps, which themselves constitute only a few percent of the complete interactomes within cells.

Network medicine: a network-based approach to human disease

*Albert-László Barabási**‡§, *Natali Gulbahce**‡|| *and Joseph Loscalzo*§

Abstract | Given the functional interdependencies between the molecular components in a human cell, a disease is rarely a consequence of an abnormality in a single gene, but reflects the perturbations of the complex intracellular and intercellular network that links tissue and organ systems. The emerging tools of network medicine offer a platform to explore systematically not only the molecular complexity of a particular disease, leading to the identification of disease modules and pathways, but also the molecular relationships among apparently distinct (patho)phenotypes. Advances in this direction are essential for identifying new disease genes, for uncovering the biological significance of disease-associated mutations identified by genome-wide association studies and full-genome sequencing, and for identifying drug targets and biomarkers for complex diseases.

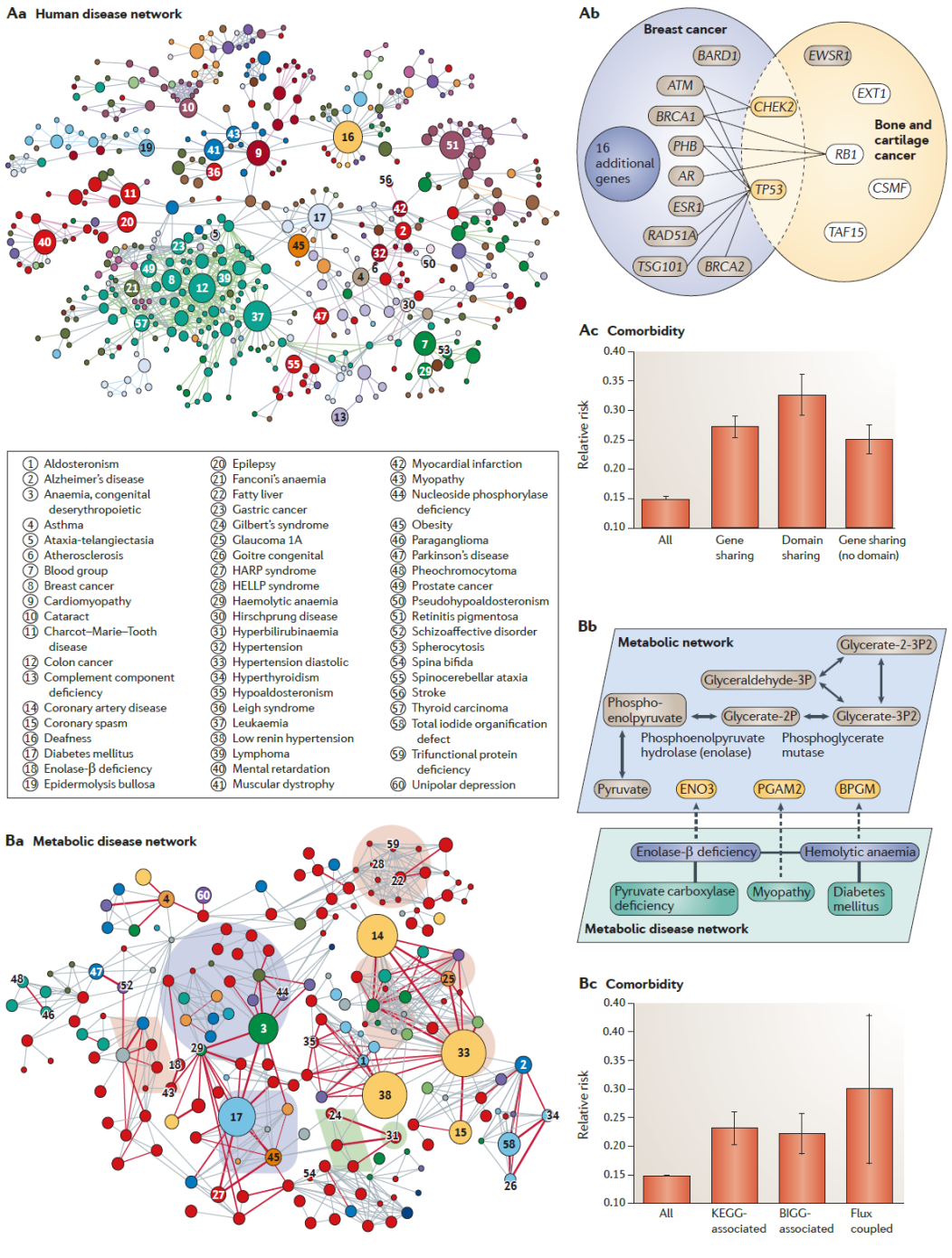


Figure 5 | **Disease networks. A** | An example of a human disease network (HDN), in which nodes represent diseases. The large panel shows the giant cluster of the obtained disease network (**Aa**). Small clusters of isolated diseases are not shown[42]. Two diseases are linked if they share one or several disease-associated genes, as shown in part **Ab**, involving breast cancer and bone and cartilage cancer[90]. The node colours reflect the class of the diseases that correspond to that node. Cancers appear as blue nodes and neurological diseases appear as red nodes. The node sizes correlate with the number of genes that are known to be associated with the corresponding disease (after REF. 42). Part **Ac** shows the comorbidity between diseases linked in the HDN as measured by the logarithm of relative risk, indicating that, if the disease-causing mutations affect the same module of the shared disease protein, then the comorbidity is higher[90]. **B** | A metabolic disease network (part **Ba**, with an example shown part **Bb**), which links two diseases if they are both associated with enzymes and if these enzymes catalyse reactions that share a metabolite (after REF. 93). Part **Bc** shows that comorbidity between metabolically linked diseases is higher than between those that are not connected, and diseases whose enzymes catalyse reactions that are coupled with each other at the flux level show even higher comorbidity. *AR*, androgen receptor; *ATM*, ataxia telangiectasia mutated; *BRCA*, breast cancer associated; *BARD1*, BRCA1 associated RING domain 1; BIGG, Biochemical Genetic and Genomics knowledgebase; BPGM, 2,3-bisphosphoglycerate mutase; *CHEK2*, CHK2 checkpoint homologue; *CSMF*, chondrosarcoma, extraskeletal myxoid, fused to EWS (also known as *NR4A3*); ENO3, enolase 3 (beta, muscle); *EWSR1*, Ewing sarcoma breakpoint region 1; *ESR1*, oestrogen receptor 1; *EXT1*, exostosin 1; HARP, hypoprebetalipoproteinaemia, acanthocytosis, retinitis pigmentosa and pallidal degeneration; HELLP, haemolytic anaemia, elevated liver enzymes and low platelet count; KEGG, Kyoto Encyclopedia of Genes and Genomes; PGAM2, phosphoglycerate mutase 2 (muscle); *PHB*, prohibitin; *RB1*, retinoblastoma 1; *TAF15*, TATA box-binding protein (TBP)-associated factor, RNA polymerase II; *TSG101*, tumour-susceptibility gene 101; *TP53*, tumour protein 53. Part **Aa** is reproduced, with permission, from REF. 42 © (2007) National Academy of Sciences. Part **Ab** is reproduced, with permission, from REF. 90 © (2009) Macmillan Publishers Ltd. All rights reserved. Part **Bb** is reproduced, with permission, from REF. 93 © (2008) National Academy of Sciences.

Genomika raka

Cancer is a disease of the genome

From human genome to cancer genome: The first decade

David A. Wheeler[1] and Linghua Wang

Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA

The realization that cancer progression required the participation of cellular genes provided one of several key rationales, in 1986, for embarking on the human genome project. Only with a reference genome sequence could the full spectrum of somatic changes leading to cancer be understood. Since its completion in 2003, the human reference genome sequence has fulfilled its promise as a foundational tool to illuminate the pathogenesis of cancer. Herein, we review the key historical milestones in cancer genomics since the completion of the genome, and some of the novel discoveries that are shaping our current understanding of cancer.

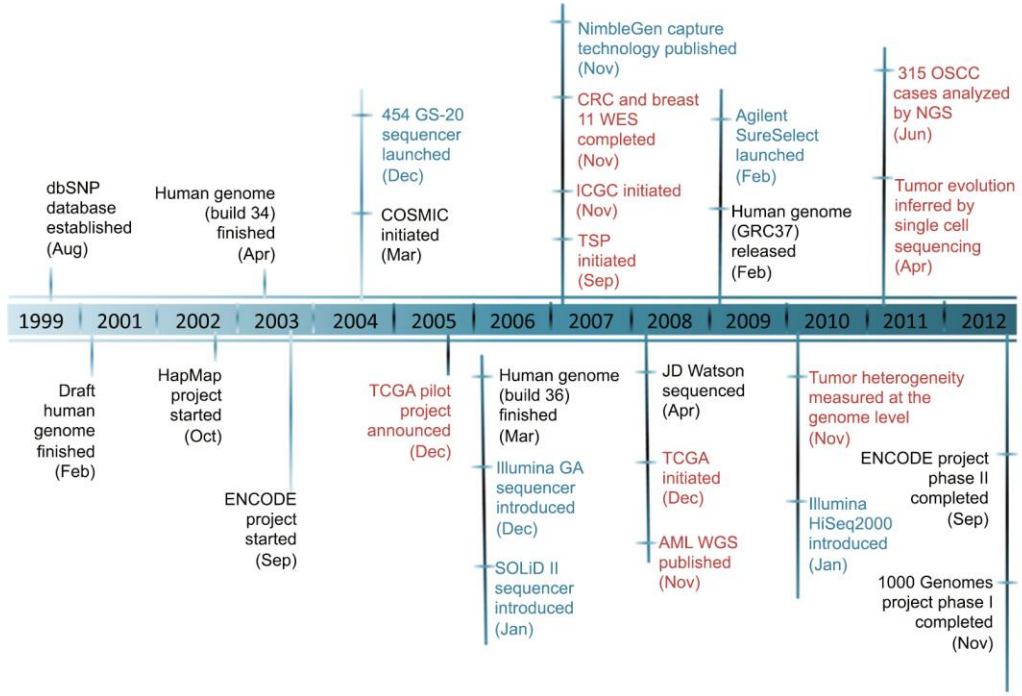
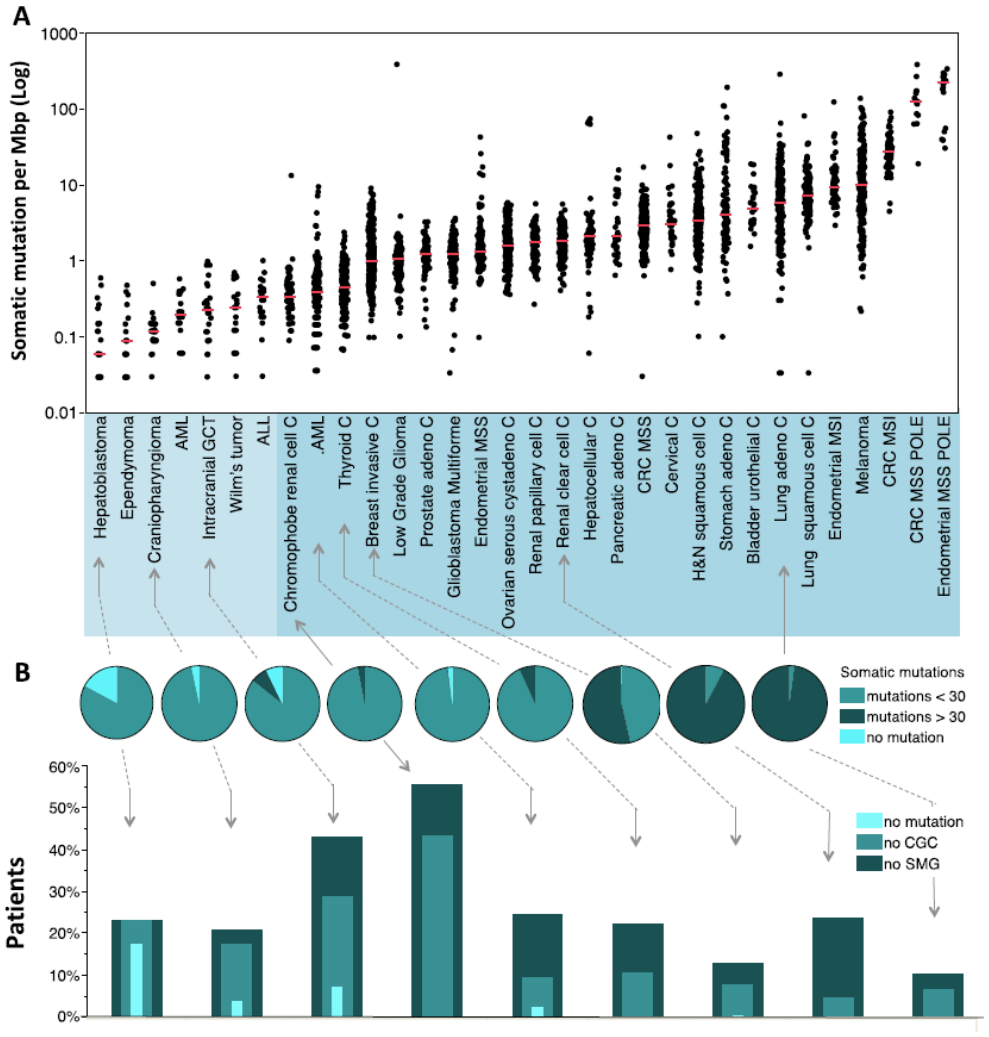


Figure 1. Major events in a decade of cancer genomics. (Dark blue) Major advances in massively parallel sequencing platforms and targeted enrichment technologies; (black) major large-scale projects designed to catalog genomic variations of normal human individuals; (red) cancer genomics. (dbSNP) Database of single nucleotide polymorphism; (HapMap) haplotype map of the human genome; (ENCODE) Encyclopedia of DNA Elements; (COSMIC) Catalog of Somatic Mutations in Cancer; (TCGA) The Cancer Genome Atlas; (GA) genome analyzer; (CRC) colorectal carcinoma; (WES) whole-exome sequencing; (ICGC) International Cancer Genome Consortium; (TSP) tumor sequencing project; (AML) acute myeloid leukemia; (WGS) whole-genome sequencing; (OSCC) ovarian small cell carcinoma.



Frequencies of somatic mutations in cancer patients. All data represents primary tumors. Only nonsilent mutations (missense, nonsense, frameshift, and splice site) were counted.

Lessons from the Cancer Genome

Systematic studies of the cancer genome have exploded in recent years. These studies have revealed scores of new cancer genes, including many in processes not previously known to be causal targets in cancer. The genes affect cell signaling, chromatin, and epigenomic regulation; RNA splicing; protein homeostasis; metabolism; and lineage maturation. Still, cancer genomics is in its infancy. Much work remains to complete the mutational catalog in primary tumors and across the natural history of cancer, to connect recurrent genomic alterations to altered pathways and acquired cellular vulnerabilities, and to use this information to guide the development and application of therapies.

Table 1. Current Large-Scale Cancer Genome Projects[a]

Anatomic Site	Tumor Type
Brain/Central nervous system	glioblastoma multiforme
	low-grade glioma
	pediatric: medulloblastoma
	pediatric: pilocytic astrocytoma
Head and neck	head/neck squamous cell cancer
	thyroid carcinoma
Thoracic	lung adenocarcinoma
	lung squamous cell carcinoma
Breast	breast lobular carcinoma
	breast ductal carcinoma
	breast triple-negative
	breast HER-2 positive
	breast ER positive vs. negative
Gastrointestinal	esophageal adenocarcinoma
	esophageal squamous carcinoma
	gastric adenocarcinoma
	gastric (intestinal/diffuse)
	hepatocellular (alcohol/adiposity)
	hepatocellular (virus)
	hepatocellular (general)
	pancreatic adenocarcinoma
	colorectal adenocarcinoma
	colon cancer (non-Western)
Gynecologic	ovarian serous cystadenocarcinoma
	endometrial carcinoma
	cervical cancer (squamous + adeno)
Urologic	renal: clear cell carcinoma
	renal: papillary carcinoma
	renal: chromophobe carcinoma
	bladder cancer
	prostate adenocarcinoma
	prostate adenocarcinoma, early onset
Skin	melanoma, cutaneous
Soft tissue (Sarcoma)	solitary fibrous tumors
	desmoid tumors
	angiorsarcomas
	leiomyosarcomas
	extraskeletal myxoid chondrosarcomas
Hematologic	acute myeloid leukemia
	lymphoma: chronic lymphocytic leuk.
	lymphoma: germinal B cell
	lymphoma: diffuse large B cell
	chronic myeloid disorders

[a]In conjunction with The Cancer Genome Atlas, International Cancer Genome Consortium, and Slim Initiative for Genomic Medicine.

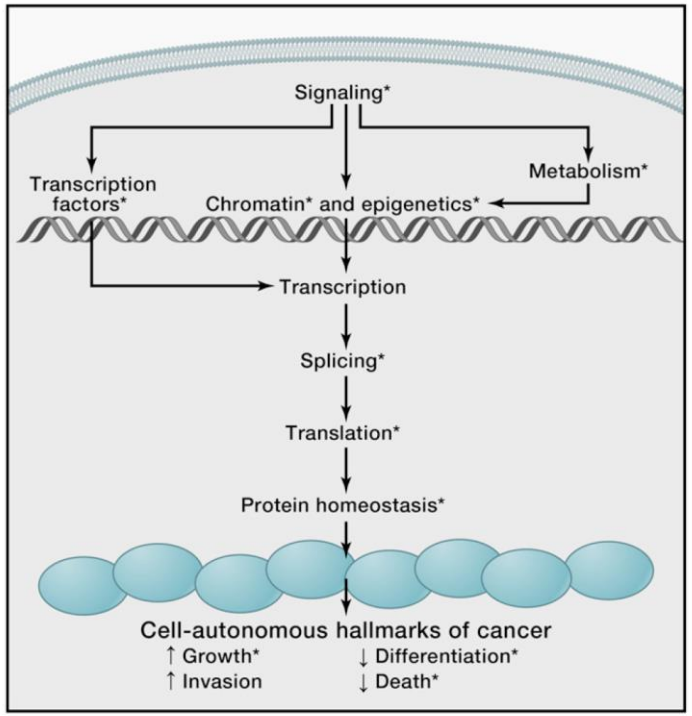


Figure 5. Genetic Alterations Disrupt Multiple Cellular Processes
Alterations in a range of cellular processes presumably contribute to cancer through their action on one or more target genes, mRNAs, or proteins, although the precise targets remain unknown in many cases (illustrated by shaded ovals). Even in advance of such knowledge, many mutations suggest potential targets for therapeutic development and allow stratification for clinical trials of targeted drugs.

Table 2. Discoveries from Cancer Genome Characterization

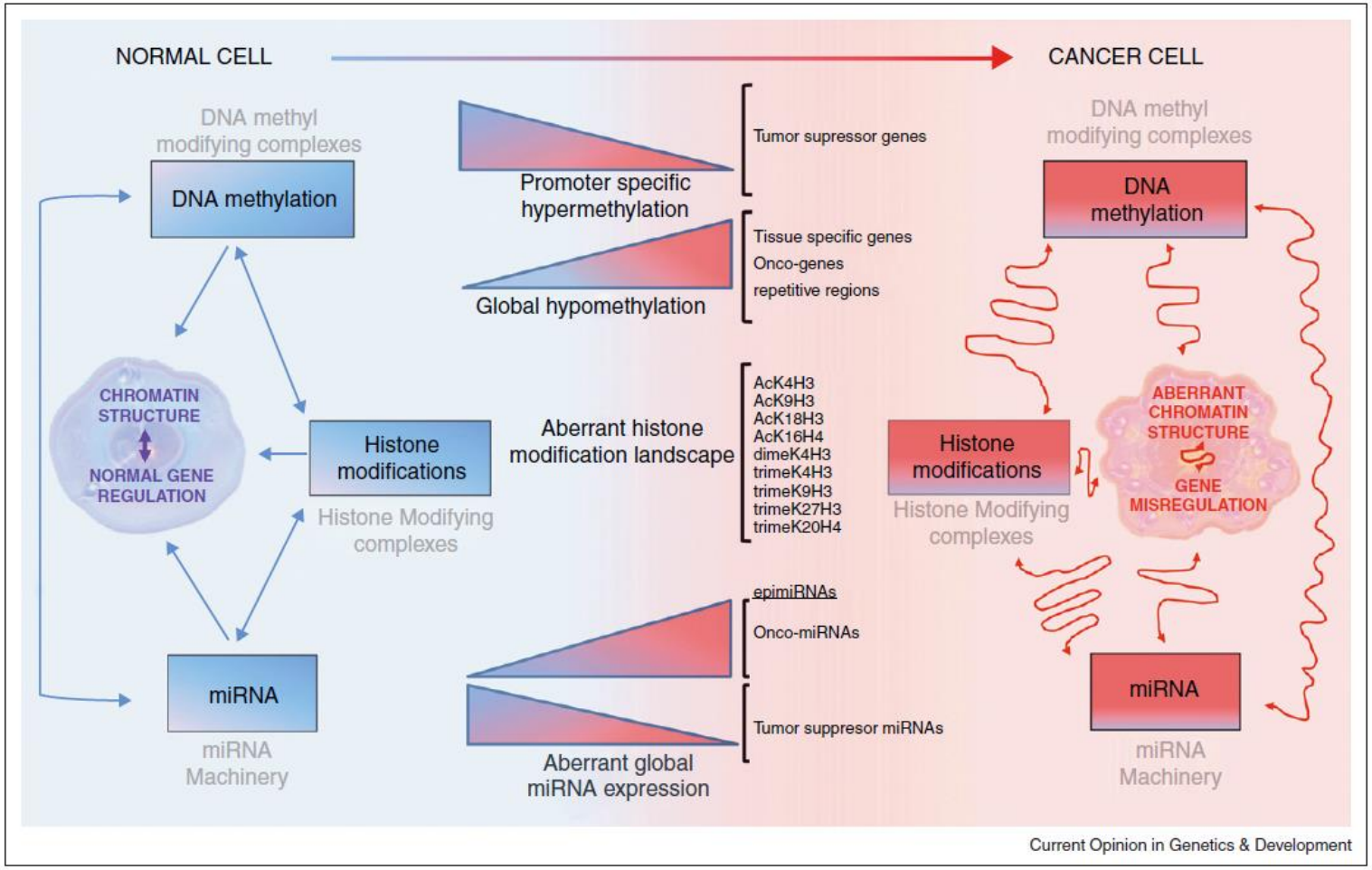
Cellular Process Altered by Genomic Alterations	Examples of Cancer Genes Discovered (or Extended to New Cancers*) by Genomics
RTK signaling	*EGFR*,[a] *ERBB2*,[*,a] *MET*,[*,a] *ALK*,[*,a] *JAK2*,[a] *RET*,[*,a] *ROS*,[*,a] *FGFR1*,[*,a] *FGFR2*,[a] *PDGFRA*,[*,a] and *CRKL*[a]
MAPK signaling (oncogenes)	*KRAS*,[*,a] *NRAS*,[*,a] *BRAF*, [a] and *MAP2K1*[a]
MAPK signaling (TSG)	*NF1*[*,b]
PI3K signaling (oncogenes)	*PIK3CA*,[a] *AKT1*,[a] and *AKT3*[a]
PI3K signaling (TSG)	*PTEN*[*,b] and *PIK3R1*[b]
Notch signaling (oncogene or TSG)	*NOTCH1*,[c] *NOTCH2*,[c] and *NOTCH3*[b]
TOR signaling (TSG)	*STK11*,[*,b] *TSC1*,[*,b] and *TSC2*[*,b]
Wnt/β-catenin signaling (TSG)	*APC*[*,b] and *CTNNB1*[*,a]
TGF-β signaling (TSG)	*SMAD2*,[*,b] *SMAD4*,[*,b] and *TGFBR2*[b]
NF-κB signaling (oncogene)	*MYD88*[a]
Other signaling	*RAC1*,[a] *RAC2*,[a] *CDC42*,[a] *KEAP1*,[b] *MAP3K1*,[b] *MAP2K4*,[b] *ROBO1*,[b] *ROBO2*,[b] *SLIT2*,[b] *SEMA3A*,[b] *SEMA3E*,[b] *ELMO1*,[d] and *DOCK2*[d]
Epigenetics DNA methylation	*DNMT3A*[b]
Epigenetics DNA hydroxymethylation	*TET2*[b]
Chromatin histone methyltransferases	*MLL*,[*,b] *MLL2*,[b] *MLL3*,[b] *EZH2*,[c] *NSD1*,[b] and *NSD3*[b]
Chromatin histone demethylases	*JARID1A*,[b] *UTX*,[b] *KDM5A*,[b] and *KDM5C*[b]
Chromatin histone acetyltransferases	*CREBP* [b] and *EP300*[b]
Chromatin SWI/SNF complex	*SMARCA1*,[*,b] *SMARCA4*,[b] *ARID1A*,[b] *ARID2*,[b] *ARID1B*,[b] and *PBRM1*[b]
Chromatin other	*CHD1*,[b] *CHD2*,[b] and *CHD4*[b]
Transcription factor lineage dependency or oncogene	*MITF*,[a] *NKX2-1*,[a] *SOX-2*,[a] *ERG*,[a] *ETV1*,[a] and *CDX2*[a]
Transcription factor other	*MYC*,[*,a] *RUNX1*,[b] *GATA3*,[b] *FOXA1*,[b] *NKX3.1*,[b] *SOX9*,[a] *NFE2L2*,[a] and *MED12*[d]
Splicing	*SF3B1*,[d] *U2AF1*,[d] *SFRS1*,[d] *SFRS7*,[d] *SF3A1*,[d] *ZRSR2*,[b] *SRSF2*,[d] *U2AF2*,[d] and *PRPF40B*[d]
RNA abundance	*DIS3*[d]
Translation/protein homeostasis/ubiquitination	*SPOP*,[d] *FBXW7*,[*,b] *WWP1*,[*,b] *FAM46C*,[d] and *XBP1*[d]
Metabolism	*IDH1*[a] and *IDH2* [a]
Genome integrity	*TP53*,[*,b] *MDM2*,[a] *MSH*,[*,b] *MLH*,[*,b] and *ATM*[*,b]
Telomere stability	*TERT* promoter mutations[a]
Cell cycle (oncogene)	*CCND1*[*,a] and *CCNE1*[*,a]
Cell cycle (TSG)	*CDKN2A*,[*,b] *CDKN2B*,[*,b] and *CDKN1B*[b]
Apoptosis regulation	*MCL1*,[a] *BCL2A1*,[a] and *BCL2L1*[a]

[a]Activating mutation or amplification.
[b]Inactivating mutation or deletion.
[c]Both activating and inactivating genomic events observed.
[d]Effect of mutations on protein function unknown.

Cancer epigenomics: beyond genomics

Juan Sandoval[1] and Manel Esteller[1,2,3]

For many years cancer research has focused on genetic defects, but during the last decade epigenetic deregulation has been increasingly recognized as a hallmark of cancer. The advent of genome-scale analysis techniques, including the recently developed next-generation sequencing, has enabled an invaluable advance in the molecular mechanisms underlying tumor initiation, progression, and expansion. In this review we describe recent advances in the field of cancer epigenomics concerning DNA methylation, histone modifications, and miRNAs. In the near future, this information will be used to generate novel biomarkers of relevance to diagnosis, prognosis, and chemotherapeutic response.



Global depiction of epigenomic alterations during oncogenesis. In conjunction with accumulation of genetic lesions, there is an aberrant pattern for the different epigenetic effectors: DNA methylation, histone modifications, and miRNAs. In normal cells, the interplay between the epigenetic factors and the chromatin structure leads to a tuned gene regulation. However, in cancer cells tumor suppressor genes promoters become hypermethylated and with an altered global pattern of histone modifications resulting in aberrant gene silencing. Moreover, global hypomethylation leads to chromosome instability and fragility. Epigenetic changes, including DNA methylation and histone modifications are responsible for abnormal mRNA and miRNA expression producing altered activation of oncogenes and silencing of tumor suppressor genes.

Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry

María Berdasco[1] and Manel Esteller[1,2,*]
[1]Cancer Epigenetics Group, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet de Llobregat, 08907 Barcelona, Catalonia, Spain
[2]Institució Catalana de Recerca I Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain
*Correspondence: mesteller@iconcologia.net
DOI 10.1016/j.devcel.2010.10.005

Appropriate patterns of DNA methylation and histone modifications are required to assure cell identity, and their deregulation can contribute to human diseases, such as cancer. Our aim here is to provide an overview of how epigenetic factors, including genomic DNA methylation, histone modifications, and microRNA regulation, contribute to normal development, paying special attention to their role in regulating tissue-specific genes. In addition, we summarize how these epigenetic patterns go awry during human cancer development. The possibility of "resetting" the abnormal cancer epigenome by applying pharmacological or genetic strategies is also discussed.

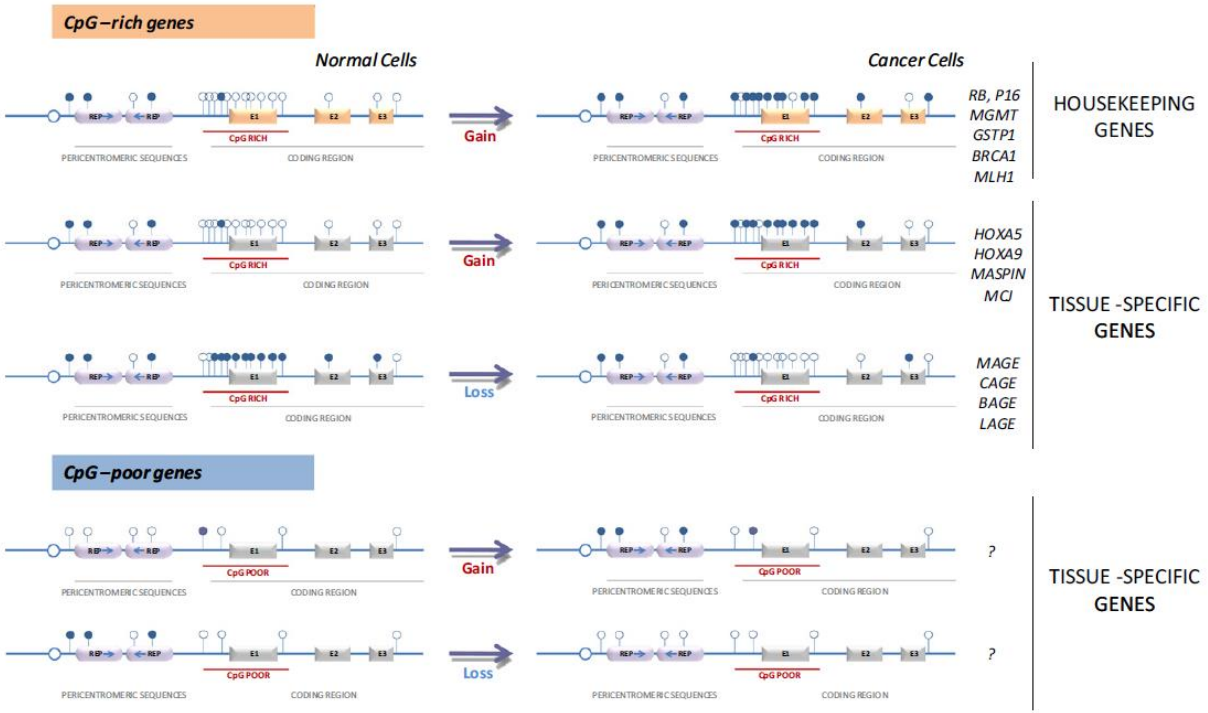


Figure 2. Distribution of CpG Islands in Promoters of Housekeeping and Tissue-Specific Genes
Promoters associated with CpG islands are found in all known housekeeping genes and half of all tissue-specific genes. Generally, housekeeping genes are unmethylated in normal cells, while tissue-specific genes may be unmethylated or methylated, depending on their requirement for lineage commitment. In a cancer cell, some of the housekeeping genes become aberrantly hypermethylated. Gains and losses of CpG hypermethylation can be observed for tissue-specific genes that are aberrantly expressed in cancer. CpG-poor regions are always found in tissue-specific genes, where expression is independent of CpG methylation status. Some examples of genes relevant to each of these circumstances are illustrated.

Table 1. A Representative List of Histone Modifiers Disrupted in Cancers

Gene Name	Substrate Specificity	Genetic Defect	Gain or Loss of Function	Tumor Type	References	PubMed ID Numbers
Histone acetyltransferases (HATs)						
*CBP (KAT3A)	H2AK5, H2BK12, H2BK15, H3K14, H3K18, H4K5, H4K8	deletion	loss	ALL; lung	Shigeno et al. (2004); Kishimoto et al. (2005)	15312679; 15701835
CBP (KAT3A)	H2AK5, H2BK12, H2BK15, H3KI14, H3K18, H4K5, H4K8	mutation	loss	lung; MSI+	Kishimoto et al. (2005); Ionov et al. (2004)	15701835; 14732695
*CBP (KAT3A)	H2AK5, H2BK12, H2BK15, H3KI14, H3K18, H4K5, H4K8	translocation	loss	AML	Panagopoulos et al. (2001, 2003)	11157802; 12461753
*p300 (KAT3B)	H2AK5, H2BK12, H2BK15	deletion	loss	cervix; ALL	Ohshima et al. (2001); Shigeno et al. (2004)	11181085; 15312679
p300 (KAT3B)	H2AK5, H2BK12, H2BK15	mutation	loss	breast; CRC	Gayther et al. (2000)	10700188
*p300 (KAT3B)	H2AK5, H2BK12, H2BK15	translocation	loss	AML	Ida et al. (1997); Chaffanet et al. (2000)	9389684; 10824998
pCAF (KAT2B)	H3K9, H3K14, H3K18; H2B	mutation	loss	epithelial cancer	Ozdağ et al. (2002); Zhu et al. (2009)	12402157; 19525977
*MORF (KAT6B)	H3K14; H4K16	translocation	loss	AML	Panagopoulos et al. (2001)	11157802
*MOZ (KAT6A)	H3K14; H4K16	translocation	loss	AML	Chaffanet et al. (2000); Panagopoulos et al. (2003)	10824998; 12461753
Histone Methyltransferases (HMTs)						
*DOT1L (KMT4)	H3K79	translocation	loss	AML	Okada et al. (2005)	15851025
EZH2 (KMT6)	H3K27	amplification	gain	prostate	Bracken et al. (2003)	14532106
*EZH2 (KMT6)	H3K27	mutation	loss	lymphoma	Morin et al. (2010)	20081860
G9a (KMT1C)	H3K9	overexpression	gain	HCC	Kondo et al. (2000)	11050047
*MLL1 (KMT2A)	H3K4	translocation	loss	AML, ALL	reviewed in Miremadi et al. (2007)	17613546
*MLL3 (KMT2C)	H3K4	deletion	loss	leukemia	Tan and Chow (2001)	11718452
NSD1 (KMT3B)	H3K36, H4K20	CpG hypermethylation	loss	neuroblastoma, glioma	Berdasco et al. (2009)	20018718
*NSD1 (KMT3B)	H3K36, H4K20	translocation	loss	AML	Jaju et al. (2001)	11493482
NSD3	H3K4, H3K27	amplification	gain	breast	Angrand et al. (2001)	11374904
RIZ1 (KMT8)	H3K9	CpG hypermethylation	loss	breast, liver	Du et al. (2001)	11719434
SMYD2 (KMT3C)	H3K36	amplification	gain	ESCS	Komatsu et al. (2009)	19423649
SUZ12 (HMT complex)	H3K9, H3K27	translocation	loss	ESS	Li et al. (2007); Panagopoulos et al. (2008)	18077430, 18722875
Histone deacetylases (HDACs)						
HDAC2	Many acetyl residues (except H4K16)	mutation	loss	MSI+	Ropero et al. (2006); Hanigan et al. (2008)	16642021; 18834886
Histone demethylase (HDMTs)						
GASC1 (KDM4C)	H3K9, H3K36	amplification	gain	ESCS; lung; breast	Cloos et al. (2006); Italiano et al. (2006); Liu et al. (2009)	16732293; 16737911; 19784073
LSD1 ((KDM1)	H3K4, H3K9	amplification	gain	prostate; bladder; lung; CRC	Kahl et al. (2006); Hayami et al. (2010)	17145880; 20333681
UTX (KDM6A)	H3K27	mutation	loss	multiple types	van Haaften et al. (2009)	19330029

Enzymes are grouped according to their catalytic activity, including histone acetyltransferases (HATs), histone methyltransferases (HMTs), histone deacetylases (HDACs) and histone demethylases (HDMTs). Hematological malignancies commonly exhibit chromosomal translocations; while solid tumors are more often affected by different genetic and epigenetic alterations, such as CpG promoter hypermethylation, deletions, point mutations or gene amplification. ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CRC, colorectal carcinoma; ESCS, esophageal squamous cell carcinoma; ESS, endometrial stromal sarcomas; HCC, hepatocellular carcinoma; MSI+, colorectal carcinoma with microsatellite instability. Genes affected in leukemia are marked with an asterisk (*).

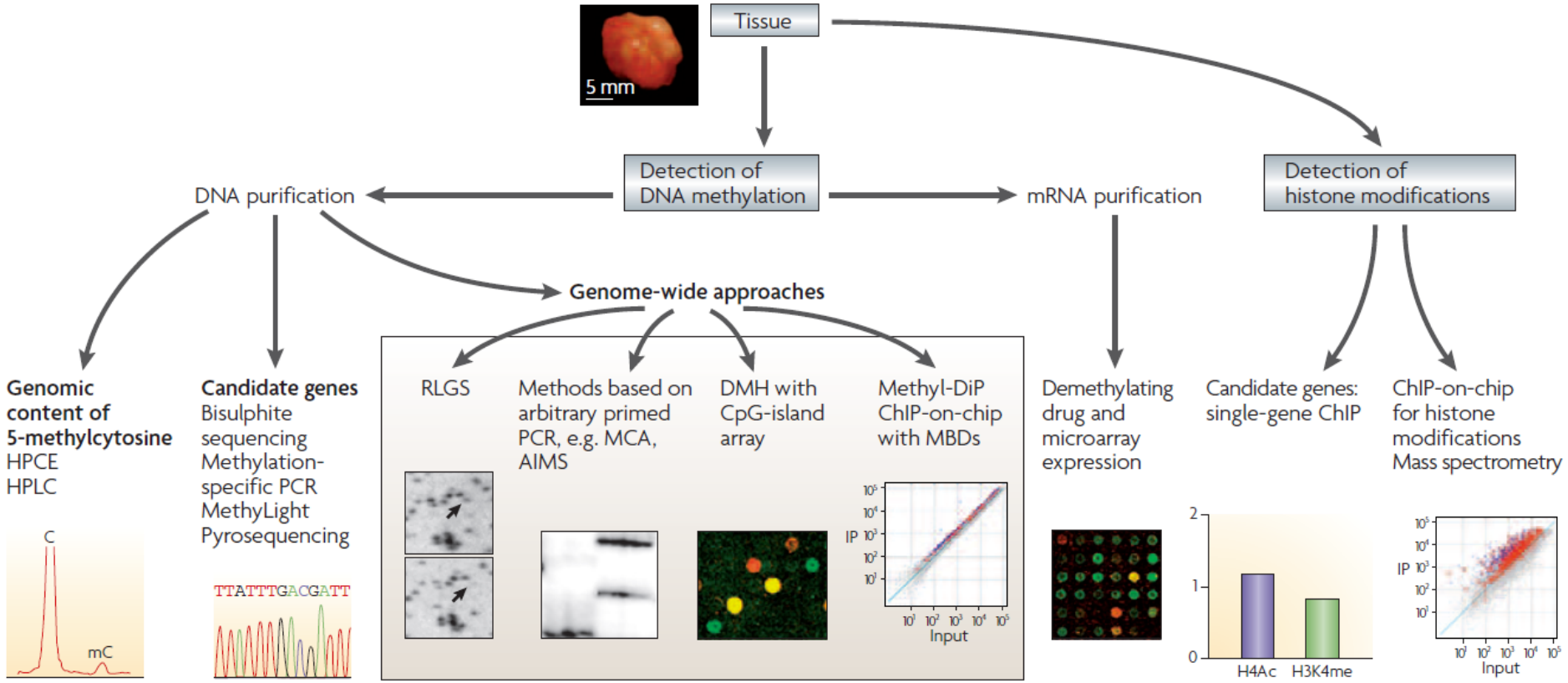


Figure 2 | **Techniques for studying epigenetic changes in cancer.** Most approaches to detecting DNA methylation start with the purification of DNA from cell samples. Subsequently, the overall DNA 5-methylcytosine content can be determined using high-performance capillary electrophoresis (HPCE) or high-performance liquid chromatography (HPLC), or the DNA methylation of specific candidate genes can be detected with methylation-sensitive methods. Recently, several genome-wide approaches to detecting DNA methylation have been developed (described in more detail in FIG. 3). An alternative approach for profiling DNA-methylation patterns is based on the extraction of mRNA, followed by microarray expression analysis. Analysis of the same sample in the presence or absence of a demethylating agent indicates genes that show increased expression owing to the removal of DNA methylation marks. For profiling histone modifications, marks at candidate genes can be detected using mass spectrometry or single-gene chromatin immunoprecipitation (ChIP) using antibodies against specific histone modifications. For global profiling, ChIP is combined with DNA arrays (ChIP-on-chip) to detect patterns across the genome. AIMS, amplification of intermethylated sites; ChIP, chromatin immunoprecipitation; DMH, differential methylation hybridization; MBDs, methyl-CpG-binding domain proteins; MCA, methylated CpG-island amplification; Methyl-DIP, methyl-DNA immunoprecipitation; RLGS, restriction landmark genomic scanning.

Cancer Epigenetics: From Mechanism to Therapy

Mark A. Dawson[1,2] and Tony Kouzarides[1,*]
[1]Gurdon Institute and Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK
[2]Department of Haematology, Cambridge Institute for Medical Research and Addenbrooke's Hospital, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK
*Correspondence: t.kouzarides@gurdon.cam.ac.uk
http://dx.doi.org/10.1016/j.cell.2012.06.013

The epigenetic regulation of DNA-templated processes has been intensely studied over the last 15 years. DNA methylation, histone modification, nucleosome remodeling, and RNA-mediated targeting regulate many biological processes that are fundamental to the genesis of cancer. Here, we present the basic principles behind these epigenetic pathways and highlight the evidence suggesting that their misregulation can culminate in cancer. This information, along with the promising clinical and preclinical results seen with epigenetic drugs against chromatin regulators, signifies that it is time to embrace the central role of epigenetics in cancer.

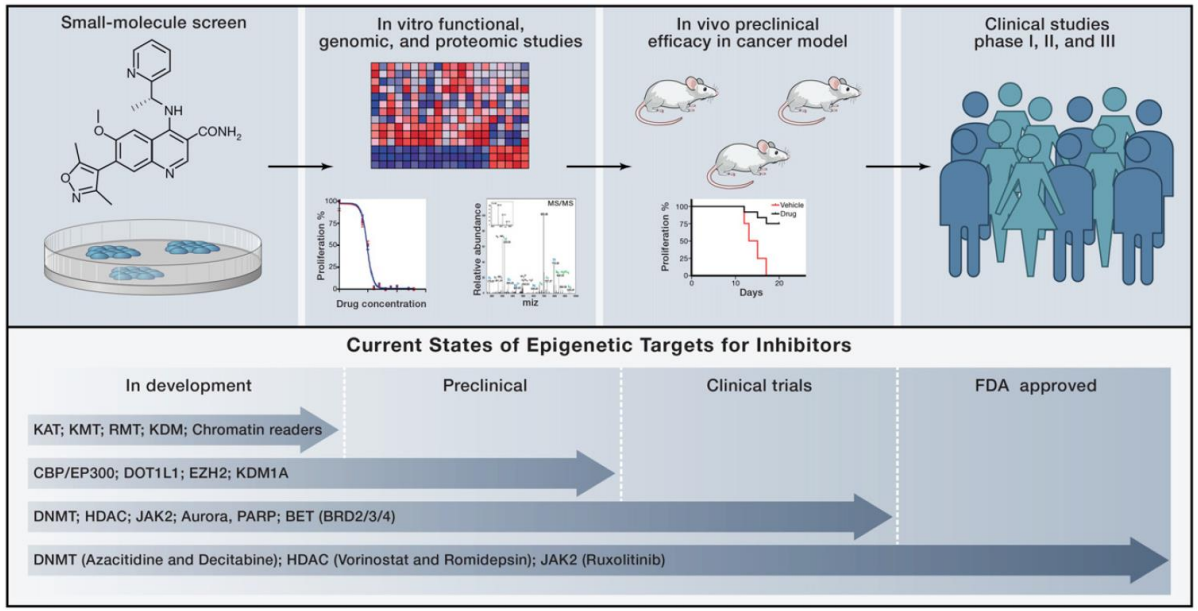


Figure 1. Epigenetic Inhibitors as Cancer Therapies
This schematic depicts the process for epigenetic drug development and the current status of various epigenetic therapies. Candidate small molecules are first tested in vitro in malignant cell lines for specificity and phenotypic response. These may, in the first instance, assess the inhibition of proliferation, induction of apoptosis, or cell-cycle arrest. These phenotypic assays are often coupled to genomic and proteomic methods to identify potential molecular mechanisms for the observed response. Inhibitors that demonstrate potential in vitro are then tested in vivo in animal models of cancer to ascertain whether they may provide therapeutic benefit in terms of survival. Animal studies also provide valuable information regarding the toxicity and pharmacokinetic properties of the drug. Based on these preclinical studies, candidate molecules may be taken forward into the clinical setting. When new drugs prove beneficial in well-conducted clinical trials, they are approved for routine clinical use by regulatory authorities such as the FDA. KAT, histone lysine acetyltransferase; KMT, histone lysine methyltransferase; RMT, histone arginine methyltransferase; and PARP, poly ADP ribose polymerase.

Acetyltransferases

Enzyme	Mutation	Tumor
KAT3A (CBP)*	T, N, F, M	AML, ALL, DLBCL, B-NHL, TCC
KAT3B (p300)*	T, N, F, M	AML, ALL, DLBCL, TCC, Colorectal, Breast, Pancreatic
KAT6A (MOZ)*	T	AML, MDS
KAT6B (MORF)+	T	AML, Uterine leiomyoma

Readers

Reader	Mutation	Tumor
BRD1*+	T	ALL
BRD3*	T	Midline carcinoma
BRD4*	T	Midline carcinoma
TRIM33*+	T	Papillary thyroid
PBRM1*	N, F, M, S, D	Renal, Breast

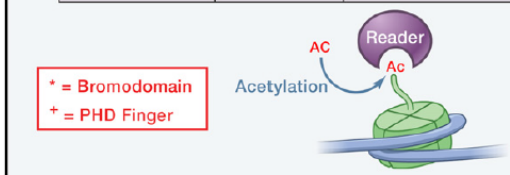


Figure 3. Cancer Mutations Affecting Epigenetic Regulators Involved in Histone Acetylation
These tables provide somatic cancer-associated mutations identified in histone acetyltransferases and proteins that contain bromodomains (which recognize and bind acetylated histones). Several histone acetyltransferases possess chromatin-reader motifs and, thus, mutations in the proteins may alter both their catalytic activities as well as the ability of these proteins to scaffold multiprotein complexes to chromatin. Interestingly, sequencing of cancer genomes to date has not identified any recurrent somatic mutations in histone deacetylase enzymes. Abbreviations for the cancers are as follows: AML, acute myeloid leukemia; ALL, acute lymphoid leukemia; B-NHL, B-cell non-Hodgkin's lymphoma; DLBCL, diffuse large B-cell lymphoma; and TCC, transitional cell carcinoma of the urinary bladder. Mutation types are as follows: M, missense; F, frameshift; N, nonsense; S, splice site mutation; T, translocation; and D, deletion.

SWI/SNF

Gene	Mutation	Tumor
BRG1*	N, M, F, D	Lung, Rhabdoid, Medulloblastoma, Breast, Prostate, Pancreas, HNSCC
BRM*	N, M,F	HNSCC
ARID1A	N, F, M, T	OCC, Endometroid, Renal, Gastric, Breast, Medulloblastoma, TCC
ARID1B	F, M, D	Breast
ARID2	N, F, S	Hepatocellular carcinoma
SNF5	D, N, F, S, T	Rhabdoid, Familial Schwannomatosis, Chondrosarcoma, Epethioloid sarcoma, Meningioma, Chordoma, Undifferentiated sarcoma
PBRM1*	N, F, M, S, D	Renal, Breast
BCL7A	T, M	B-NHL, Multiple myeloma
BAF60A	M	Breast

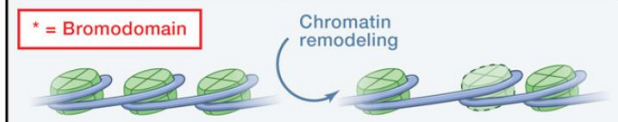


Figure 6. Cancer Mutations Affecting Members of the SWI/SNF Chromatin-Remodeling Complex
SWI/SNF is a multisubunit complex that binds chromatin and disrupts histone-DNA contacts. The SWI/SNF complex alters nucleosome positioning and structure by sliding and evicting nucleosomes to make the DNA more accessible to transcription factors and other chromatin regulators. Recurrent mutations in several members of the SWI/SNF complex have been identified in a large number of cancers. Abbreviations for the cancers are as follows: B-NHL, B-cell non-Hodgkin's lymphoma; HNSCC, head and neck squamous cell carcinoma; OCC, ovarian clear cell carcinoma; and TCC, transitional cell carcinoma of the urinary bladder. Mutation types are as follows: M, missense; F, frameshift; N, nonsense; S, splice site mutation; T, translocation; and D, deletion.

Genomika infekcijskih bolezni - patogenomika

Bacterial pathogenomics

Mark J. Pallen[1] & Brendan W. Wren[2]

Genomes from all of the crucial bacterial pathogens of humans, plants and animals have now been sequenced, as have genomes from many of the important commensal, symbiotic and environmental microorganisms. Analysis of these sequences has revealed the forces that shape pathogen evolution and has brought to light unexpected aspects of pathogen biology. The finding that horizontal gene transfer and genome decay have key roles in the evolution of bacterial pathogens was particularly surprising. It has also become evident that even the definitions for 'pathogen' and 'virulence factor' need to be re-evaluated.

Table 1 | Examples of mobile genetic elements that encode virulence factors and are present in human pathogens

Type of mobile element	Pathogen	Virulence factor
Plasmid	*Bacillus anthracis*	Anthrax toxin
	Clostridium tetani	Tetanus toxin
	Enterotoxigenic *Escherichia coli*	Heat-stabile toxin, heat-labile toxin and fimbriae
	Mycobacterium ulcerans	Polyketide toxin
	Salmonella enterica serovar Typhimurium	SpvR, SpvA, SpvB, SpvC and SpvD proteins*
	Shigella spp.	Type III secretion system
	Staphylococcus aureus	Exfoliatin B
	Pathogenic *Yersinia* spp.	Type III secretion system
Prophage	*Corynebacterium diphtheriae*	Diphtheria toxin
	Enterohaemorrhagic *E. coli*	Shiga toxin and type III secretion effectors
	S. aureus	Staphylococcal enterotoxin A, exfoliatin A and Panton–Valentine leukocidin
	Streptococcus pyogenes	Streptococcal pyrogenic exotoxins, DNases and streptococcal phospholipase A_2 (Sla)
	Vibrio cholerae	Cholera toxin
Pathogenicity island	*Clostridium difficile*	Clostridial enterotoxin and clostridial cytotoxin
	Enteropathogenic and enterohaemorrhagic *E. coli*	Type III secretion system
	Uropathogenic *E. coli*	Fimbriae, iron-uptake systems, the capsular polysaccharide and α-haemolysin
	Helicobacter pylori	Cag antigen
	S. enterica	Type III secretion systems
	S. aureus	Toxic-shock toxin, staphylococcal enterotoxin B, enterotoxin C, enterotoxin K and enterotoxin L

*Involved in intracellular survival.

Table 2 | Examples of 'virulence systems' encoded in the genomes of both pathogenic and non-pathogenic bacteria

Virulence factor	Role in virulence	Homologues in non-pathogens	Potential explanation for presence in non-pathogens	References
Type III secretion system	Role in infection with many human pathogens, including chlamydiae, salmonellae, shigellae and yersiniae	Remnants of type III secretion systems and effectors in commensal strains of *Escherichia coli*, including the laboratory strain *E. coli* K-12	Had a role in a former niche (a degenerate system)	29, 36
		Type III secretion systems in environmental bacteria: for example, *Myxococcus xanthus*, *Verrucomicrobium spinosum*, *Desulfovibrio vulgaris* and non-pathogenic *Yersinia* spp.	Mediate uncharacterized interactions with nematodes, and amoebae and other microscopic eukaryotes in terrestrial and aquatic environments	67
		Type III secretion systems in symbiotic bacteria: for example, *Photorhabdus luminescens*, '*Hamiltonella defensa*', *Aeromonas veronii*, *Sodalis glossinidius* and *Protochlamydia amoebophila*	Mediate symbiosis with plants, nematodes, leeches, insects and amoebae	67–73
Type VI secretion system	Role in infection with *Vibrio cholerae* or *Pseudomonas aeruginosa*	Type VI secretion systems in environmental bacteria: for example, *Rhodopirellula baltica*, *Hahella chejuensis* and *Oceanobacter* sp.	Mediate uncharacterized interactions with nematodes, and amoebae and other microscopic eukaryotes in aquatic environments	74, 75
ESAT6 and associated Esx secretion system	Key virulence determinant of *Mycobacterium tuberculosis* and *Staphylococcus aureus*, and major attenuating factor in the bacillus Calmette–Guérin (BCG) vaccine against tuberculosis	Esx gene clusters in *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus halodurans*, *Clostridium acetobutylicum*, *Listeria innocua* and *Streptomyces coelicolor*	Mediate uncharacterized interactions with nematodes, and amoebae and other microscopic eukaryotes in terrestrial and aquatic environments, or involved in conjugative transfer of plasmids	76–81
Specific invasion genes (for example, *yijP*, *ibeB* and *ompA*)	Contribute to invasion of *E. coli* in animal models of meningitis	Invasion genes in commensal strains of *E. coli*, including the laboratory strain *E. coli* K-12	Are a short-sighted local adaptation that does not contribute to transmission (a dead-end trait)	82–84

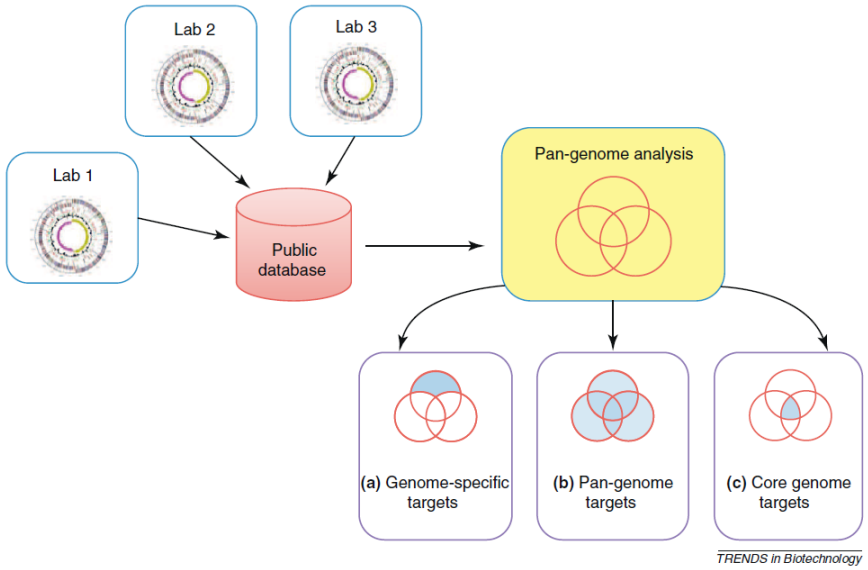
An 'eco–evo' perspective on host–pathogen interactions

Similarly, the readily available bacterial genome-sequence data have challenged the simplistic views that a bacterial pathogen can be understood solely by identifying its virulence factors and that pathogens always evolve from non-pathogens by acquiring virulence genes on plasmids, bacteriophages or pathogenicity islands. Instead, genomics has helped to blur the distinction between pathogens and non-pathogens and between virulence factors and colonization factors. And it has catalysed a copernican shift in how host–pathogen interactions are viewed, a shift away from an anthropocentric focus towards a broader perspective that places interactions between eukaryophilic bacteria and eukaryotes in a wider ecological and evolutionary context (Fig. 2). Inherent in this 'eco–evo' perspective is the need to identify the selective advantages of virulence factors in the broader lifestyle of the pathogen. In addition, 'evolutionary narratives' that interweave genomic changes with ecological shifts can now be constructed. For example, genomic comparisons allow a reconstruction of how the plague bacillus, *Y. pestis* (a rodent and flea pathogen that is occasionally transmitted to humans), evolved from a gastrointestinal pathogen (*Yersinia pseudotuberculosis*) in an evolutionary blink of an eye (about 10,000 years), through the processes of gene gain, loss and rearrangement[8,53,54].

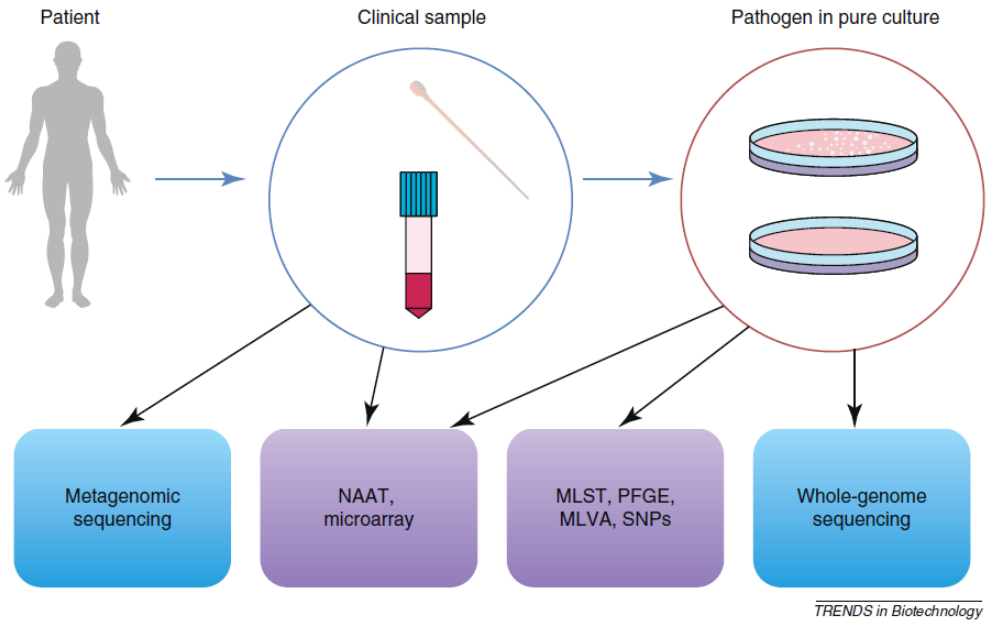
Bacterial population genomics and infectious disease diagnostics

Sandeep J. Joseph[1] and Timothy D. Read[1,2]

New sequencing technologies have made the production of bacterial genome sequences increasingly easy, and it can be confidently forecasted that vast genomic databases will be generated in the next few years. Here, we detail how collections of bacterial genomes from a particular species (population genomics libraries) have already been used to improve the design of several diagnostic assays for bacterial pathogens. Genome sequencing itself is also becoming more commonly used for epidemiological, forensic and clinical investigations. There is an opportunity for the further development of bioinformatic tools to bring even further value to bacterial diagnostic genomics.



Using population genomics to design genetic assays. Population genomics libraries are constructed from multiple individual genome sequences available through accession to public databases from laboratories around the world. ***Pan-genome analysis can be used to design and validate new genetic tests for clinical use.*** Such analysis of the library can be conceptualized as a Venn diagram. The figure shows the **intersection of the gene contents of three genomes from a species** (in reality, many more than three genomes are now available; Table 1). **Orthologous genes** fall into the area where the circles intersect. The bioinformatic strategy will depend on the type of genetic assay being developed, with specific areas shown in blue shading: (a) some assays target specific genes only present in one strain or a subset; (b) pan-genome microarrays target a non-redundant set of probes that encapsulates all known variation; and (c) some genetic assays attempt to identify all members of a species, with core genes as the target.



Using population genomics in clinical diagnosis. Genetic tests are used to identify and subtype pathogens from either clinical samples taken directly from the patient (e.g. blood and urine) or isolated bacteria grown in pure culture. Most traditional genetic assays (purple boxes) can be used on DNA harvested from pure cultures. Querying clinical samples has been generally limited to NAATs or microarray-based assays. Newly emerging whole-genome and metagenomic sequencing methods are shown in blue boxes. Pure cultures of bacteria and viruses can also be sequenced and compared with a population genomics library for that species (Figure 2).

Table 1. Number of publicly available genome sequences for selected bacterial pathogens[a]

Species	Number of draft (or "in progress") genomes	Number of finished genomes
B. anthracis	20	6
Bacillus cereus	43	9
Borrelia burgdorferi	12	5
C. trachomatis	23	20
C. botulinum	8	11
E. coli	270	38
Haemophilus influenzae	25	4
Listeria monocytogenes	24	6
M. tuberculosis	71	5
Salmonella enterica	62	18
S. aureus	69	21
Streptococcus pneumoniae	186	12

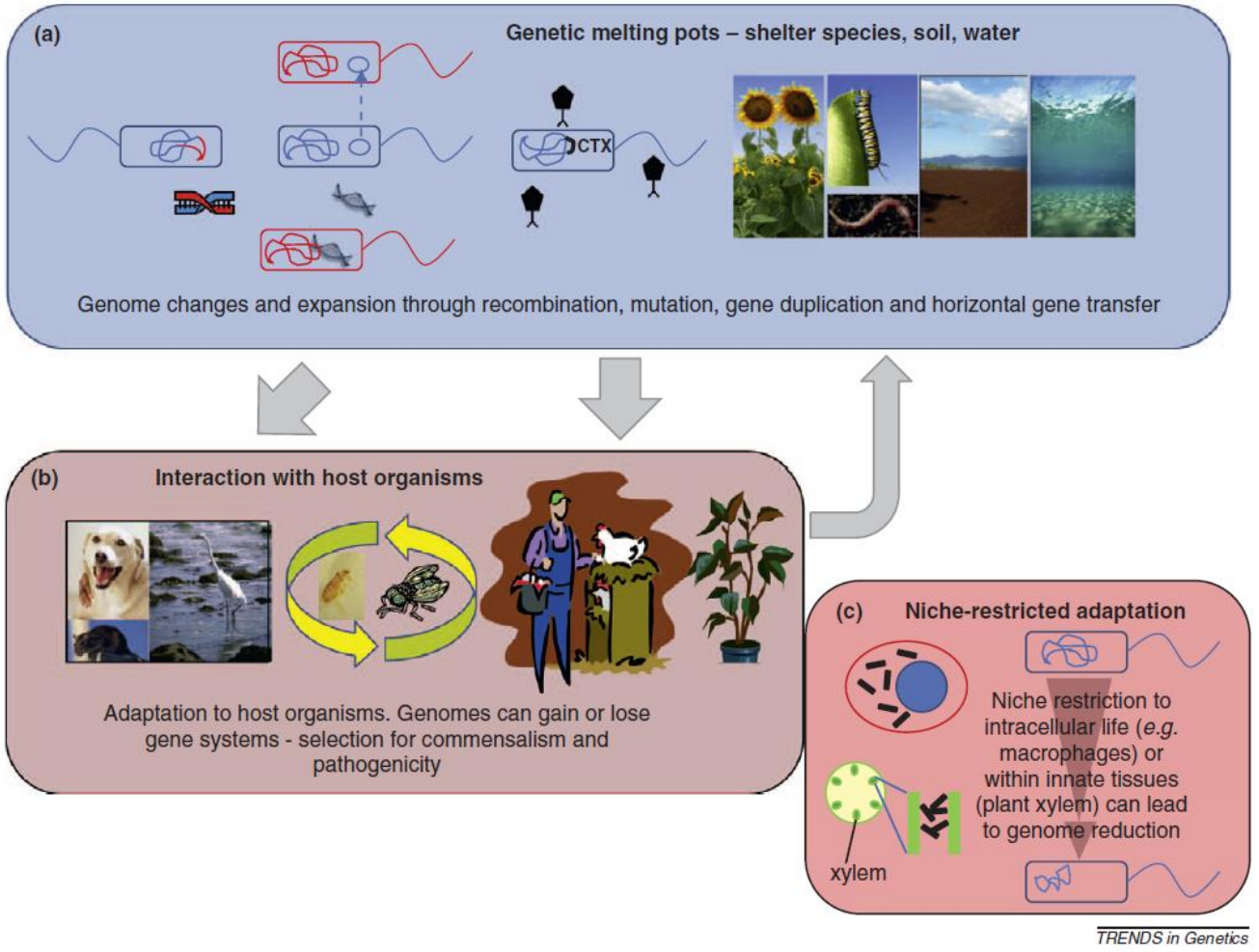
[a]Based on the NCBI Genome Project database (http://www.ncbi.nlm.nih.gov/), last accessed August 2010.

Bacterial pathogen evolution: breaking news

Robert W. Jackson[1], Louise J. Johnson[1], Simon R. Clarke[1] and Dawn L. Arnold[2]

The immense social and economic impact of bacterial pathogens, from drug-resistant infections in hospitals to the devastation of agricultural resources, has resulted in major investment to understand the causes and consequences of pathogen evolution. Recent genome sequencing projects have provided insight into the evolution of bacterial genome structures; revealing the impact of mobile DNA on genome restructuring and pathogenicity. Sequencing of multiple genomes of related strains has enabled the delineation of pathogen evolution and facilitated the tracking of bacterial pathogens globally. Other recent theoretical and empirical studies have shown that pathogen evolution is significantly influenced by ecological factors, such as the distribution of hosts within the environment and the effects of co-infection. We suggest that the time is ripe for experimentalists to use genomics in conjunction with evolutionary ecology experiments to further understanding of how bacterial pathogens evolve.

In short, pathogenicity depends not only on the details of interacting molecules, but also on the ecology of both pathogen and host. The short generation times, fluid genomes and vast population sizes of bacteria ensure they respond rapidly to selection, so evolutionary considerations are a vital part of long-term strategies to control or exploit pathogens.



Overview of bacterial pathogen evolution. Bacteria can evolve from non-pathogens to become pathogens through acquisition of new genetic material that enables them to colonise a host species and have detrimental effects on the host. Pathogenic bacteria can undergo further genetic modification that leads to altered virulence and changes in their genome.

Table 1. A selection of recent outbreaks*

Features	Disease or pathogen	When	Where	Scale	Comments
Airborne, point source	Legionnaire's disease	July 2012	Stoke on Trent, UK	<10 cases	Likely source a hot tub
Airborne, propagated human-to-human	Measles	2012 to now	South Wales, UK	>500 cases	Subsequent to poor take-up of measles, mumps and rubella (MMR) vaccine
Airborne, propagated human-to-human	*Bordetella pertussis*	2011 to now	England and Wales, UK	>2,000 cases	Perhaps related to waning immunity in adults
Airborne, propagated human-to-human	Bovine tuberculosis	2006	Birmingham, UK	<10 cases	Spread through social links, including nightclub
Blood-borne	Hepatitis B	2011	Swansea, UK	≥4 cases	Link between cases unclear
Bloodstream infection, common source	Anthrax	2009 to 2012	Europe, including UK	100s of cases	Thought be associated with contaminated batch of heroin
Exposure to animal feces	*E. coli* O157	September 2012	Sutton Coldfield, UK	<10 cases	Contact between humans and animals in suburban park
Food-borne, point source	*Salmonella* Newport	Early 2012	England, UK	>35 cases	Linked to consumption of watermelon
Hospital-acquired	*Pseudomonas aeruginosa*	Late 2011 to early 2012	Northern Ireland, UK	4 babies	Associated with contaminated hospital water supplies
Waterborne	Cholera	2010 to now	Haiti		Occurred 10 months after powerful earthquake
Waterborne	Cholera	2008 to now	Zimbabwe		Exacerbated by consequences of economic collapse, including poor water sanitation
Zoonotic, animal-to-human spread	Influenza H7N9	April 2013	China	>11 cases	Virus type known to be circulating in birds

*This list is drawn largely from the BBC news website [58] and is illustrative rather than exhaustive.

Table 2. How whole-genome sequencing contributes to each step in outbreak investigation

Step	Contribution of whole-genome sequencing (WGS)	References
Confirming the existence of an outbreak	Bench-top sequencing of whole bacterial genomes in near real time to confirm or refute the existence of outbreaks of MRSA or *C. difficile*	[25]
	Open-ended diagnostic metagenomics to identify and characterize outbreak strain	[30]
Case definition	WGS and/or metagenomics leads to the development of diagnostic reagents then used in defining cases within an outbreak	[3,31,32]
Descriptive study: collecting data and generating hypotheses	Integration of WGS with geographical data to uncover modes of spread of typhoid	[38]
	Reconstruction of routes of transmission, including hidden transmission events	[25,45,59,60]
	Identification of virulence factors and antimicrobial resistance	[26,34,36]
Analysis and hypothesis testing	Iterative refinements to assumptions and models	[25,27,36,41-47]
Institution and verification of control measures	Documenting effects of vaccination on pathogen populations	[48,49]
	Confirmation that infections are imported rather than locally transmitted	[25,27,50]
Communication	Need for user-friendly digital output easily transferred between laboratories and expert advice of clinical academics at home in research and clinical environments	

Table 3. Whole-genome sequencing in outbreak investigations: opportunities and challenges

Feature	Opportunities	Challenges
Sequence generation	Provision of data on a timescale that allows clinical interventions	Chasing a moving target: difficult to devise stable and agreed standard operating procedures in the face of relentless technical innovation
	Costs now comparable to those of other clinically relevant expenditure (such as of antibiotic treatment or bed occupancy)	Proof needed that WGS cost-effective across a range of clinical applications
	Use now comparable to that of other automated laboratory systems	Difficulties in predicting phenotype from genotype
	Delivers far richer data than any previous method	Still sufficiently technically demanding to require input of skilled staff
	Potential for open-ended one-size-fits-all culture-independent workflow	Resistance to adoption of potentially disruptive technology
Data handling	Provides portable, digital, library-based approach	Large datasets require significant hardware for storage and analysis
		Need for standardized, robust, user-friendly analysis pipelines
		Issues over data storage, ownership and presentation need to be resolved
		Integration with healthcare informatics systems to allow easy communication with clinicians
Epidemiological analysis	WGS provides highest possible resolution	Need to move beyond SNP typing of draft genomes of colony-purified isolates to embrace full range of genome variation, including within-patient variation
	Potential to link pathogen discovery, biology and evolution with phylogeny and epidemiology to facilitate iterative hypothesis generation, testing and refinement	Better integration with conventional epidemiology required to place data in context and evaluate hypothesized routes of transmissions
		Acquiring clinical metadata often remains a bottleneck

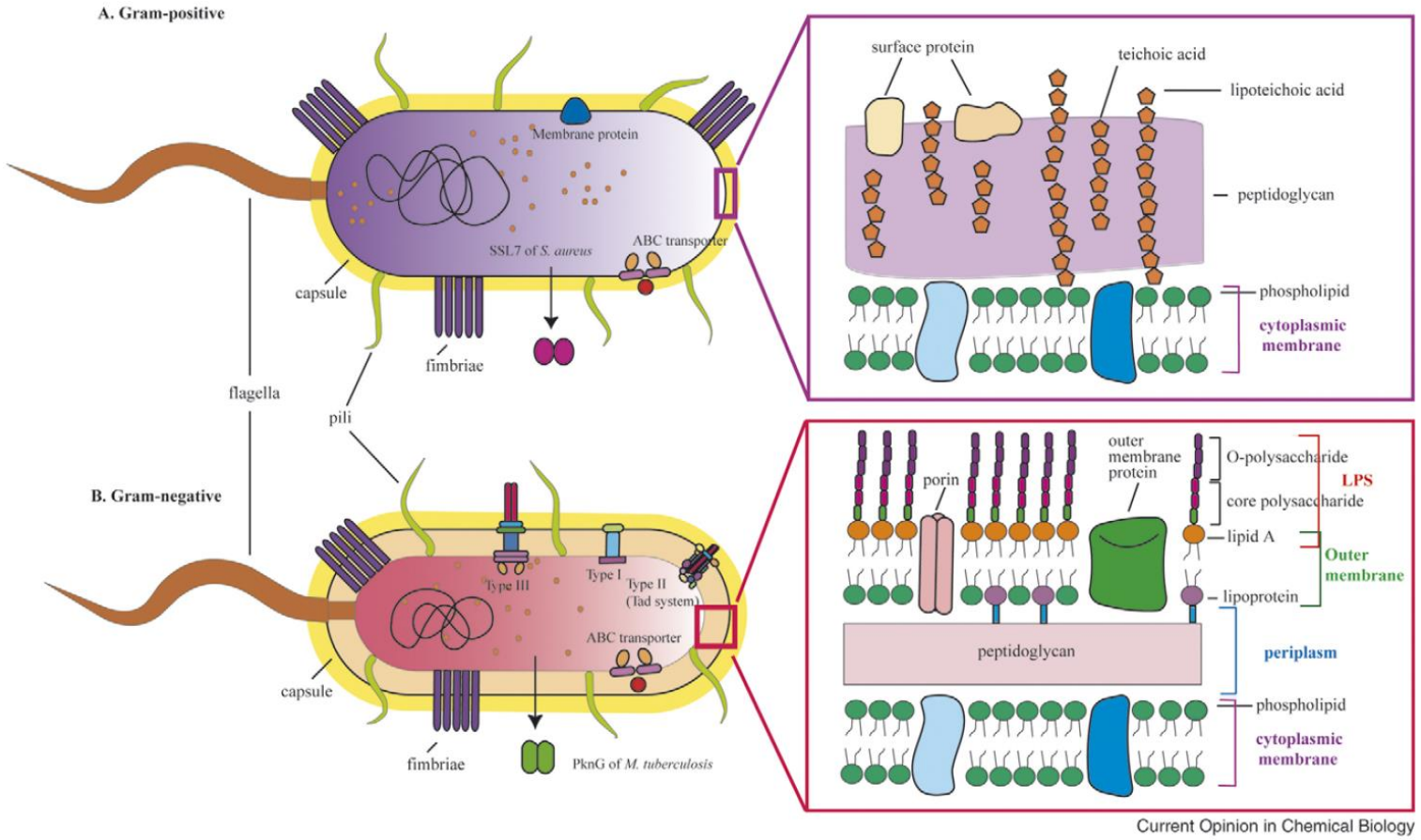
Genomics and outbreak investigation: from sequence to consequence

Esther R Robinson[1], Timothy M Walker[2] and Mark J Pallen[3]*

Discovery of virulence factors of pathogenic bacteria

Hsing-Ju Wu[1], Andrew H-J Wang[1] and Michael P Jennings[2]

Discovering virulence factors of pathogenic bacteria is a key in understanding pathogenesis and for identification of targets for novel drugs and design of new vaccines. Comparative genomics, transcriptomics, and proteomics have become the popular tools in discovering the virulence factors in bacterial pathogens, such as *Neisseria meningitidis*, *Yersinia pestis*, *Mycobacterium tuberculosis*, and *Staphylococcus aureus*. In addition, proteomics has been employed successfully in the study of the mechanism of post-translationally modified proteins of bacterial pathogens. Once the putative virulence factors are identified by genomics and/or proteomics, their functions and mechanisms can be further investigated by phenotypic analyses including mutagenesis and biochemical methods and/or structural biology. Combination of these techniques will accelerate the developments of therapeutic drugs and vaccines in combating bacterial diseases.



The schematic diagram showing the major virulence factors of pathogenic bacteria

Table 1

The classification of the virulence factors of pathogenic bacteria including newly identified virulence factors

Classification	Subclassification	Examples	Reference
1. Membrane proteins	Adhesion	Pilus-associated proteins: microbial surface cell recognition adhesion matrix molecules (MSCRAMMs), for example, Cpa, PrtF1, and PrtF2 of *S. pyogenes*, FnBPA of *S. aureus*	[13,21•]
		Pla and pH 6 fimbriae antigen (PsaA) of *Y. pestis*	[3•]
		Fimbrial adhesins (type I, P and S/F1C) of uropathogenic *E. coli*	[16•]
		Lral family of proteins of *S. pyogenes* and *S. pneumoniae*	[28,30•]
		PsaA of *S. pneumoniae*, ScaA of *S. gordonii*, SsaB from *S. sanguis* and FimA of *S. parasanguis*	[33–36]
	Invasion	Hyaluronidase, lecithinase, and phospholipase of *Clostridium* and Gram-positive cocci	[3•]
	Colonization	Type IV pilus of *N. gonorrhoeae*, *N. menigitidis*, *V. cholerae*, *P. aeruginosa* and entero-pathogenic strains of *E. coli*	[3•,47•]
		Urease of *H. pylori*	[27]
	Surface components	Spa (surface protein A) of *S. aureus*	[17•]
		Surface protein A (SpsA), pneumococcal surface protein A (PspA), choline-binding protein A (CbpA), LytA amidase and pneumococcal surface antigen A (PsaA) of *S. pneumoniae*	[3•]
		LipL32, LipL21 and LipL41 of *Leptospira* spp.	[22]
		Spy0416 of Group A *Streptococcus*	[29]
		VI antigen of *Salmonella typhi*	[3•]
	Outer membrane proteins	YaeT of *E. coli*	[50•]
		FhaC of *B. pertussis*	[51•]
2. Capsule		poly-γ-D-glutamic acid of *B. anthracis*	[3•]
		F1 capsule antigen of *Y. pestis*	[3•]
3. Secretory proteins	Immune response inhibitors	TlpA of *S. enterica* serovar Enteritidis	[10•]
		AvrA of *S. enterica* serovar Typhimurium	[26]
		YopJ of *Yersinia*	[3•]
		Protein kinase G (PknG) and phosphatase (MptpB) of *M. tuberculosis*	[52]
		SSL7 of *S. aureus*	[53•]
	Toxins	Exotoxins: for example, (1) Ymt of *Y. pestis*; (2) Lethal toxin (zinc metalloprotease, Npr599 and InhA) of *B. anthracis*; (3) Protective antigen (PA) and edema toxin of *B. anthracis*; (4) α-Toxin of *S. aureus*; (5) α-Hemolysin (Hly) of uropathogenic *E. coli*; (6) Exotoxin A of *P. aeruginosa*; (7) Diphtheria exotoxin (DT) of *Corynebacterium diptheriae*; (8) Vacoulating toxin of *H. pylori*; (9) Superantigens of *S. pyogenes* and *S. aureus*	[3•,16•, 21•,25,29]
	Transport of toxins	Type I: for example, haermolysin of *E. coli*	[5,7,15•]
		Type II: for example, (1) Pseudopilin XcpT of *Pseudomonas aeruginosa*; (2) The Tad system	
		Type III: for example, (1) Yop of *Y. pestis*; (2) SptP, SgD/SopB and PrgI of *S. typhimurium*; (3) BsaL of *B. peudomallei*; (4) MxiH and Ipa of *S. flexneri*	
		Type V: Autotransporter, for example, (1) AusI of *N. meningitidis*; (2) YapA, C, E-H and K-N of *Y. pestis*	
4. Cell wall and outer membrane components		Peptidoglycan, LPS or endotoxin or teichoic acid	[3•]
5. Others	Biofilm	α-Acetolactate decarboxylase (AlsD) of *S. aureus*	[17•]
		acetolactate synthase of *S. aureus*	[17•]
	Iron acquisition	Siderophore receptor, for example, FrpB, LbpA/B of *N. meningitidis*	[3•]
		Siderophore, for example, (1) Ybt system in *Y. pestis*; (2) Aerobactin, enterobactin, IroN and yersiniabactin of urogenic *E. coli*; (3) Enterochelin of *Salmonella*	[16•,26]
		ABC transport system, for example, YfeABCDE of *Y. pestis*	[3•]
	PhoP/PhoQ two-component system	*Y. pestis*	[3•]

Bacterial virulence factors

In spite of advances in treatment and prevention, bacterial pathogens still pose a major threat on public health worldwide. To understand how pathogenic bacteria interact with their hosts to produce clinical disease is a fundamental issue. A key first step in this process is the identification of novel virulence determinants that may serve as targets for vaccine and drug development.

We expect that combining different genomic, proteomic, and structural results will substantially increase our understanding of complex biological processes associated with virulence factors and assist the development of antibacterial drugs and vaccines.

Gut microbiome-host interactions in health and disease

James M Kinross[1], Ara W Darzi[1] and Jeremy K Nicholson*[2]

The gut microbiome is the term given to describe the vast collection of symbiotic microorganisms in the human gastrointestinal system and their collective interacting genomes. Recent studies have suggested that the gut microbiome performs numerous important biochemical functions for the host, and disorders of the microbiome are associated with many and diverse human disease processes. Systems biology approaches based on next generation 'omics' technologies are now able to describe the gut microbiome at a detailed genetic and functional (transcriptomic, proteomic and metabolic) level, providing new insights into the importance of the gut microbiome in human health, and they are able to map microbiome variability between species, individuals and populations. This has established the importance of the gut microbiome in the disease pathogenesis for numerous systemic disease states, such as obesity and cardiovascular disease, and in intestinal conditions, such as inflammatory bowel disease. Thus, understanding microbiome activity is essential to the development of future personalized strategies of healthcare, as well as potentially providing new targets for drug development. Here, we review recent metagenomic and metabonomic approaches that have enabled advances in understanding gut microbiome activity in relation to human health, and gut microbial modulation for the treatment of disease. We also describe possible avenues of research in this rapidly growing field with respect to future personalized healthcare strategies.

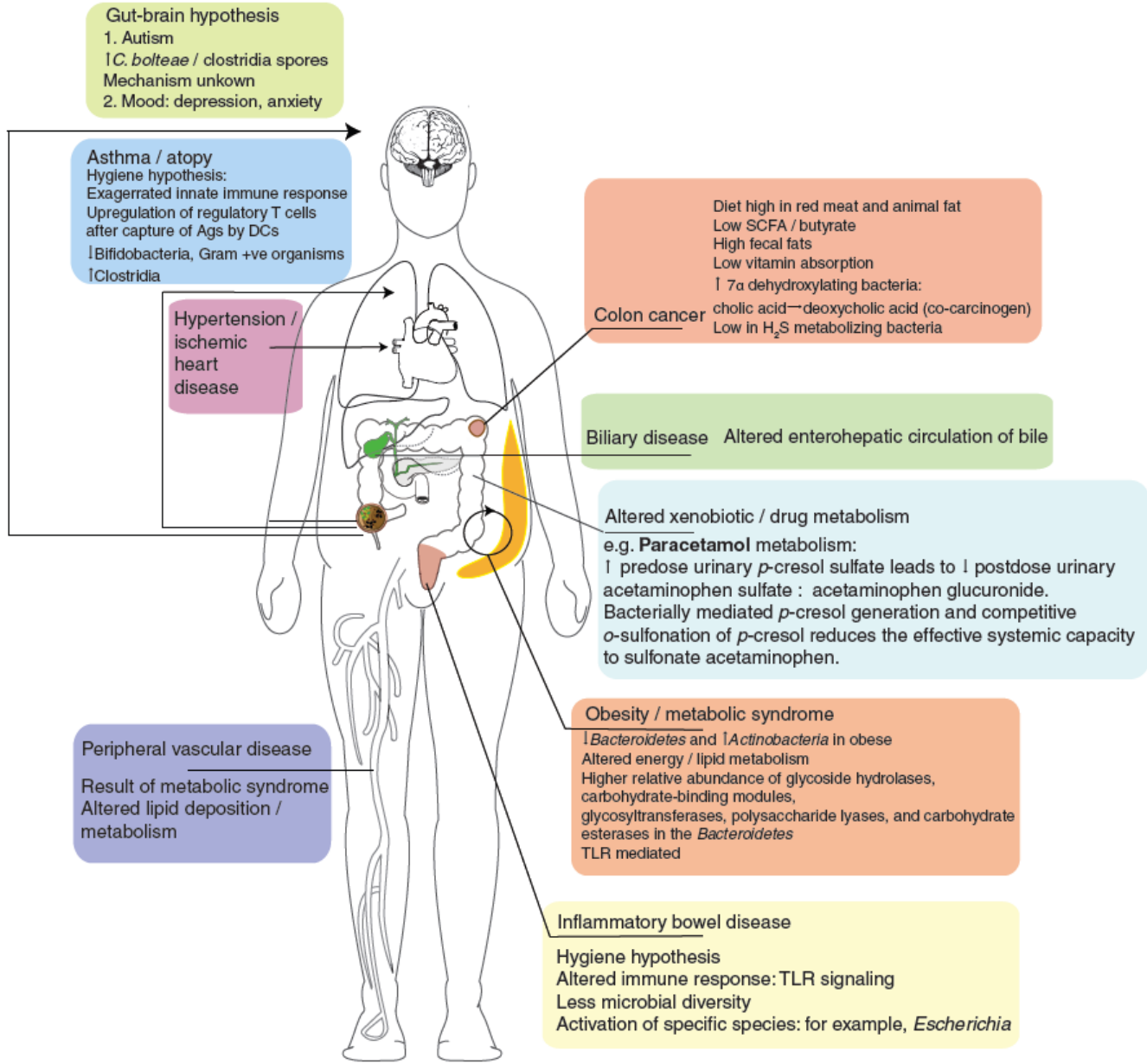
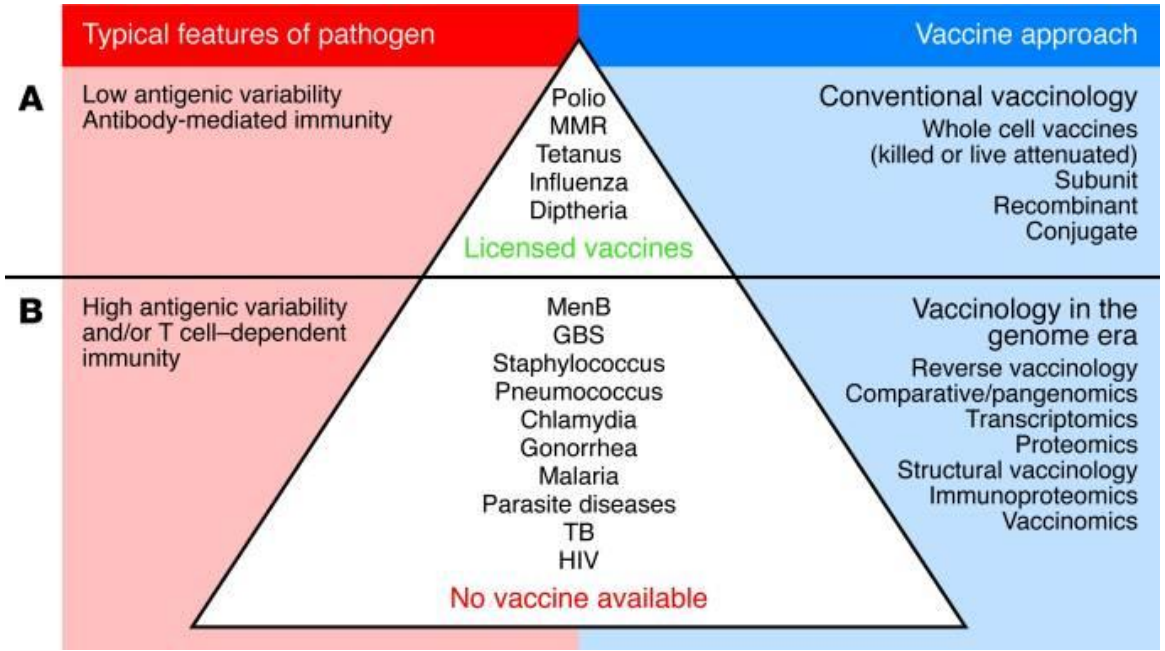


Figure 1. Diseases influenced by gut microbial metabolism. The variety of systemic diseases that are directly influenced by gut microbial metabolism and its influence on other mammalian pathways, such as the innate immune system, are shown. Specifically highlighted are the metabolic pathways involved in drug metabolism and obesity that are directly influenced by the gut microbial content. Ags, antigens; *C. bolteae, Clostridium bolteae*; DCs; dendritic cells; SCFA, short-chain fatty acid; TLR, Toll-like receptor.

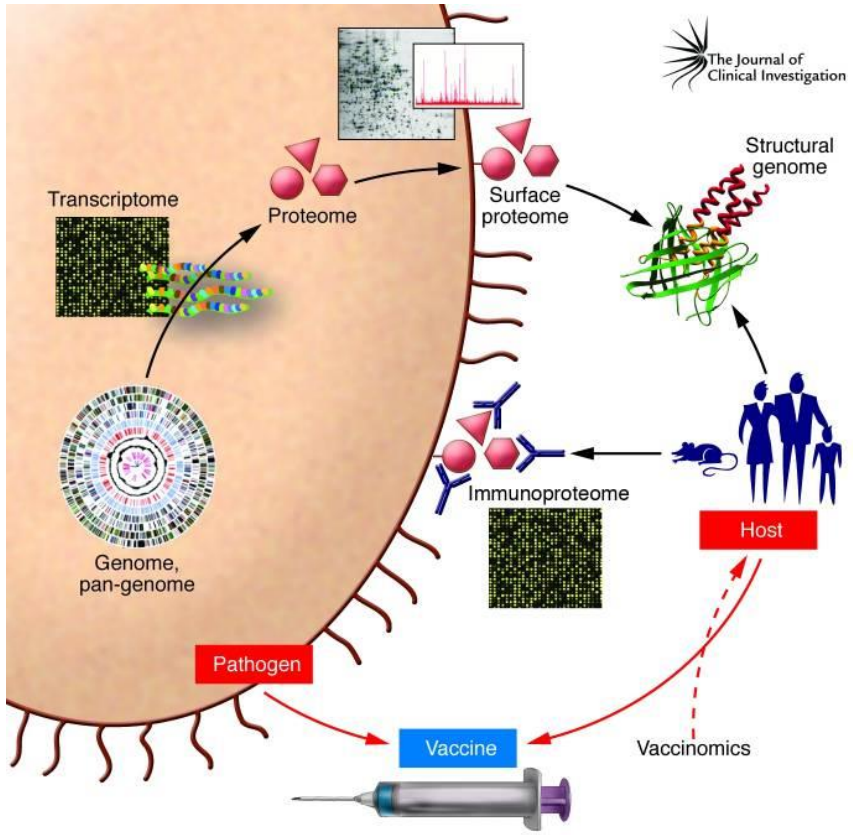
Pomen genomike za razvoj novih cepiv

Vaccination has played a significant role in controlling and eliminating life-threatening infectious diseases throughout the world, and yet currently licensed vaccines represent only the tip of the iceberg in terms of controlling human pathogens. However, as we discuss in this Review, the arrival of the genome era has revolutionized vaccine development and catalyzed a shift from conventional culture-based approaches to genome-based vaccinology. The availability of complete bacterial genomes has led to the development and application of high-throughput analyses that enable rapid targeted identification of novel vaccine antigens. Furthermore, structural vaccinology is emerging as a powerful tool for the rational design or modification of vaccine antigens to improve their immunogenicity and safety.



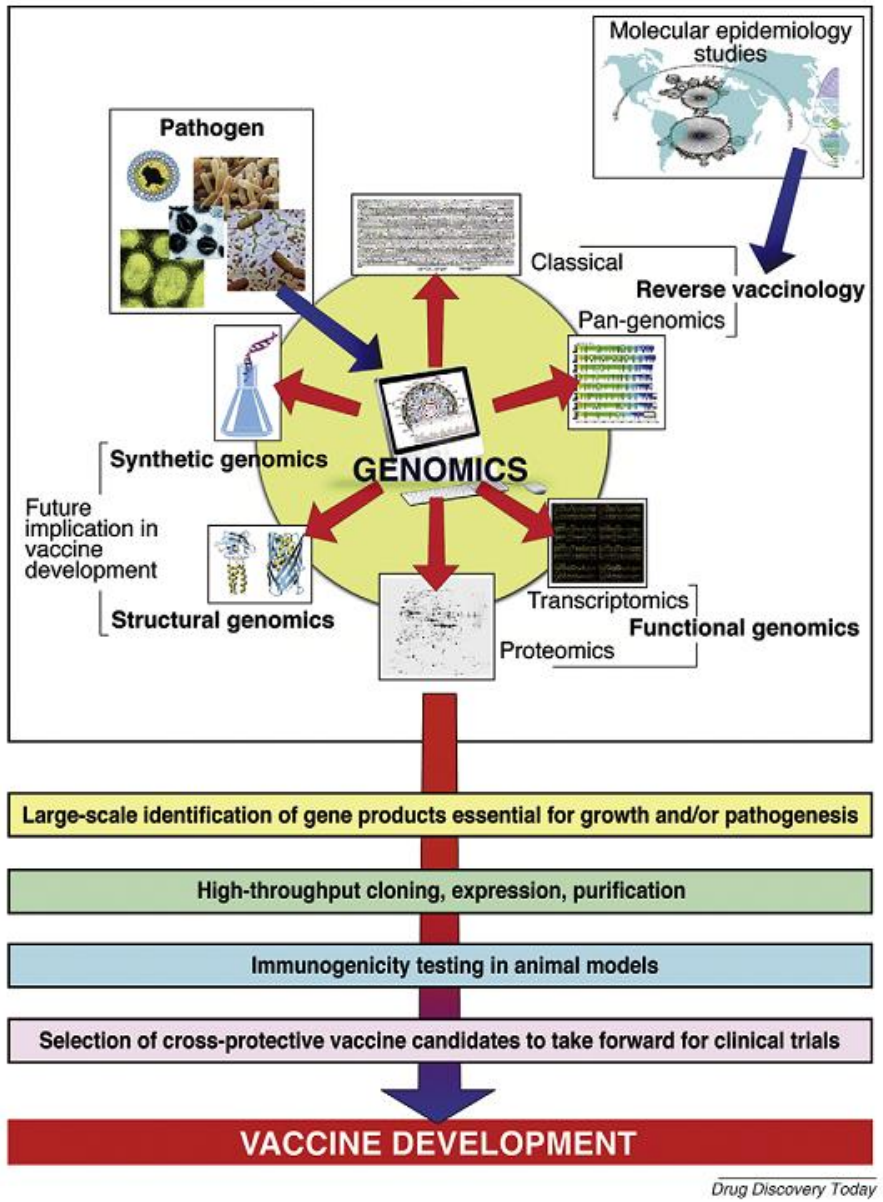
Schematic overview of conventional vaccinology versus vaccinology in the genome era.

Vaccinology in the genome era



Schematic overview of the way in which high-throughput analyses applied to various aspects of a pathogen and its interactions with the host immune system are used to identify vaccine candidates in the genome era.

The use of genomics in microbial vaccine development



Vaccination is one of the most effective tools for the prevention of infectious diseases. The availability of complete genome sequences, together with the progression of high-throughput technologies such as functional and structural genomics, has led to a new paradigm in vaccine development. Pan-genomic reverse vaccinology, with the comparison of sequence data from multiple isolates of the same species of a pathogen, increases the opportunity of the identification of novel vaccine candidates. Overall, the conventional empiric approach to vaccine development is being replaced by vaccine design. The recent development of synthetic genomics may provide a further opportunity to design vaccines.

TABLE 1

Examples of different postgenomics approaches in the development of vaccines against some bacterial pathogens, and the status of the corresponding vaccine development

Pathogen	Disease	Approach	Refs	Status of vaccine development
Neisseria meningitidis serogroup B	Bacterial meningitis and septicemia	Reverse vaccinology Microarray Proteomics	[12,13,15] [37,38] [49]	Phase II clinical trials
Streptococcus pneumoniae	Bacterial pneumonia, sepsis, sinusitis, otitis media and bacterial meningitis	Classical or comparative reverse vaccinology Proteomics	[16,50] [51]	Discovery/preclinical
Bacillus anthracis	Anthrax	Reverse vaccinology CGH microarray Microarray Proteomics and immunoproteomics	[18] [52] [53] [54]	Discovery/preclinical
Staphylococcus aureus	Variety of infections, including 'pelvic syndrome', rapidly progressive pneumonia, ocular infections, septic thrombophlebitis	CGH microarray Immunoproteomics	[55] [41]	Discovery/preclinical
Porphyromonas gingivalis	Periodontitis	Reverse vaccinology	[19]	Discovery/preclinical
Mycobacterium tuberculosis	Tuberculosis	Reverse vaccinology	[21]	Discovery/preclinical
Helicobacter pylori	Ulcer, atrophic gastritis, adenocarcinoma, lymphoma	Reverse vaccinology Immunoproteomics	[20] [56]	Discovery/preclinical
Streptococcus agalactiae (GBS)	Bacterial sepsis, pneumonia, meningitis	Classical or comparative reverse vaccinology	[36]	Discovery/preclinical
Streptococcus pyogenes (GAS)	Many systemic invasive infections including necrotizing fasciitis, myositis, pneumonia, sepsis, arthritis	Genome-wide analysis Proteomics (surface proteome)	[57] [40]	Discovery/preclinical
Chlamydia pneumoniae	Pneumonia, meningitis, middle era infections	Reverse vaccinology and proteomics	[17]	Discovery/preclinical

Next generation deep sequencing and vaccine design: today and tomorrow

Fabio Luciani, Rowena A. Bull and Andrew R. Lloyd

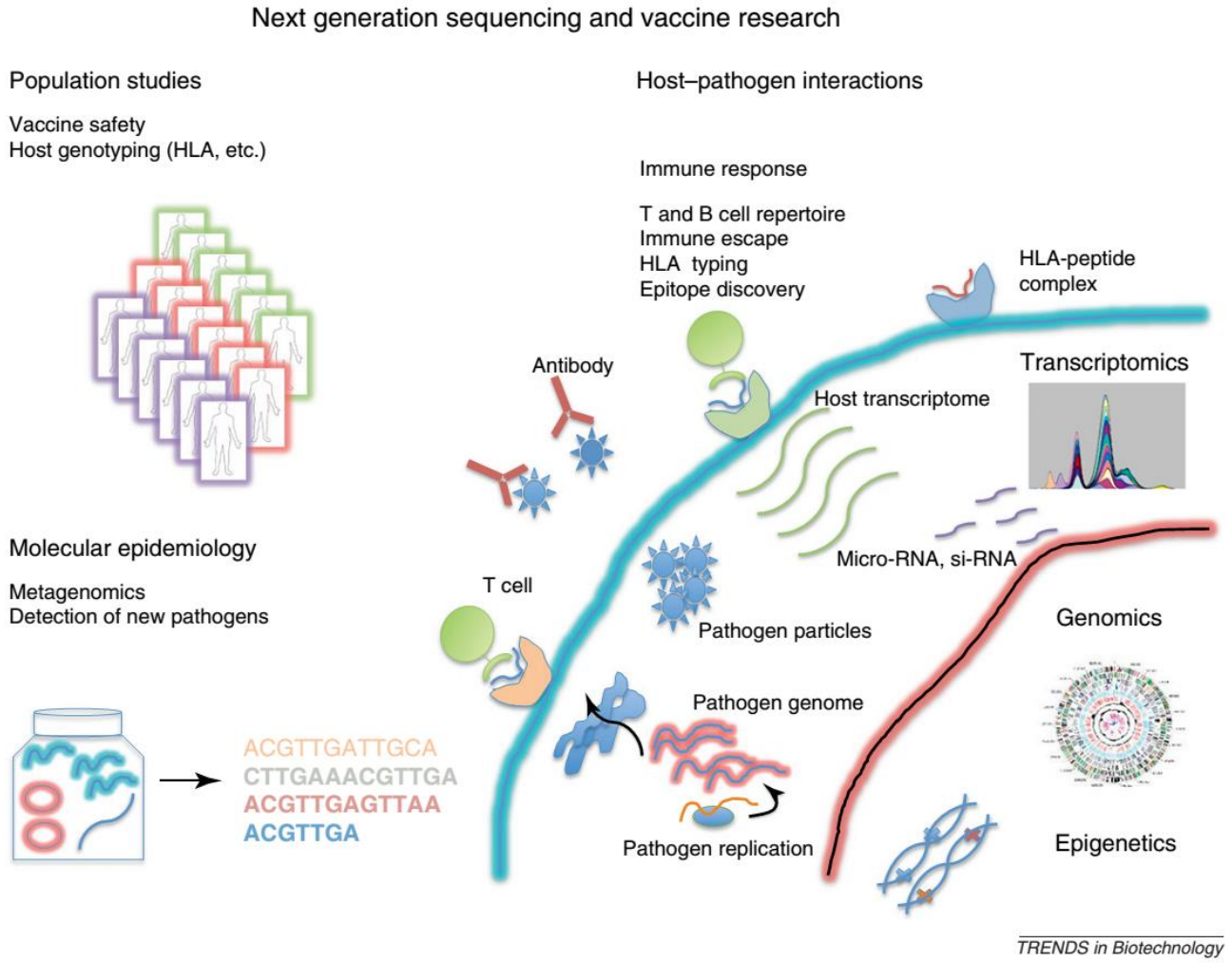
Inflammation and Infection Research Centre, School of Medical Sciences, University of New South Wales, Sydney, Australia

Next generation sequencing (NGS) technologies have redefined the *modus operandi* in both human and microbial genetics research, allowing the unprecedented generation of very large sequencing datasets on a short time scale and at affordable costs. Vaccine development research is rapidly taking full advantage of the advent of NGS. This review provides a concise summary of the current applications of NGS in relation to research seeking to develop vaccines for human infectious diseases, incorporating studies of both the pathogen and the host. We focus on rapidly mutating viral pathogens, which are major targets in current vaccine research. NGS is unraveling the complex dynamics of viral evolution and host responses against these viruses, thus contributing substantially to the likelihood of successful vaccine development.

Table 1. Current applications of NGS to the study of rapidly mutating viruses

Area of research	Pathogen[a]	Refs
Detection of low frequency variants	HCV HIV SARS Influenza Norovirus Rhinovirus	[16,47] [69,70] [71] [49,59] [46] [72]
Drug resistance	Influenza HBV HIV HCV	[49] [73] [70] [74]
Host–pathogen interactions	General HIV HCV	[41] [39,62,65,69] [40]
Mechanisms of viral evolution within-host	HCV HIV Rhinovirus Influenza HBV Poxvirus	[16,47] [75] [72] [49] [7] [76]
Molecular epidemiology of pathogens	Influenza HIV	[16,77] [16]
Detection of contaminants for vaccine safety	Poliovirus	[52–54]
Detection of adaptive host responses	HIV	[11]
Detection of escape variants	HIV HCV	[10,68–70,78] [16,47]
Haplotype reconstruction	HIV HCV Norovirus	[79] [47] [46]
Detection of new strains/ pathogens/genotypes [metagenomics]	Influenza virus Arenavirus Norovirus/ influenza	[14] [80] [76]

[a]HBV, hepatitis B virus; SARS, severe acute respiratory syndrome.



Next generation sequencing (NGS) is applicable to a wide spectrum of settings with a direct impact on vaccine research. Applications of NGS for vaccine studies range from systematic analyses of many samples collected from human populations, to detailed longitudinal studies of host–pathogen interactions within fewer subjects. NGS allows rapid assessment of both human and pathogen genomes, their transcriptomes, as well as examination of host immune responses, such as T and B cell diversity. NGS can be used to assess the quality of vaccine stocks, the diversity of HLA polymorphisms in large populations, and also for detection of new pathogenic strains in mixed samples.

Genomika in odpornost na antibiotike

Antibiotics and the post-genome revolution

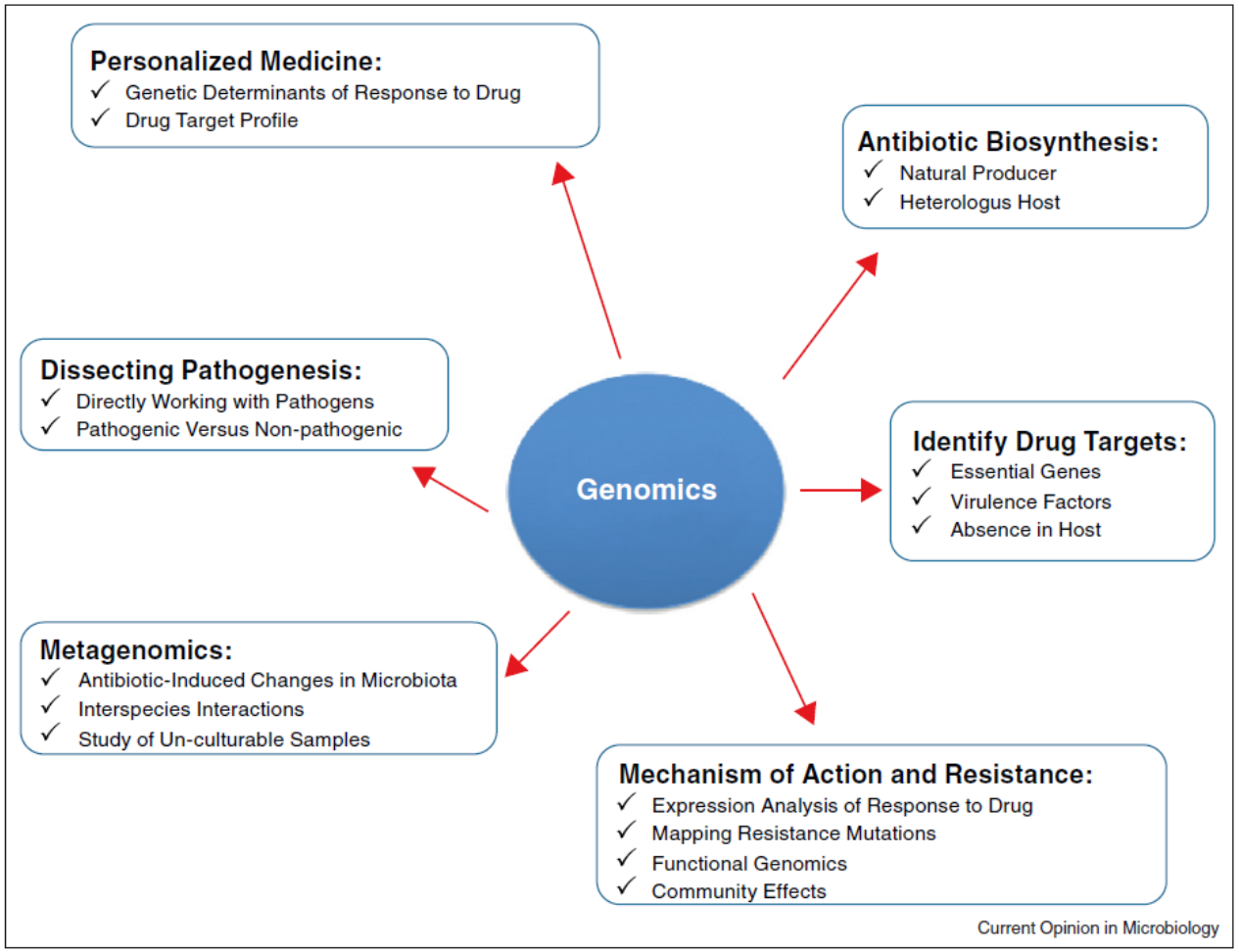
Sasan Amini and Saeed Tavazoie

The emergence of pathogenic bacteria resistant to multiple antimicrobial agents is turning into a major crisis in human and veterinary medicine. This necessitates a serious re-evaluation of our approaches toward antibacterial drug discovery and use. Concurrent advances in genomics including whole-genome sequencing, genotyping, and gene expression profiling have the potential to transform our basic understanding of antimicrobial pathways and lead to the discovery of novel targets and therapeutics.

Conclusions

The contribution of genomics to our knowledge of antibiotics will expand owing to the dropping cost of next-generation sequencing technologies. This will have a multi-dimensional impact on various aspects of the antibiotic development field, including but not limited to new drug target identification, understanding the mechanism of antibiotic action, drug safety and efficacy assessment, bacterial resistance development, understanding the pathogenesis process, optimizing antibiotic biosynthetic process, and devising personalized treatments for specific instances of infectious disease (Figure 1). This combination will be a valuable asset in our endless battle against drug resistant bacterial infections.

Figure 1



Contribution of Genomics to Antibiotics: Genomics has transformed our overall knowledge of antibiotics and our approach toward antibacterial drug discovery. This includes but is not limited to new drug target identification, understanding the mechanism of antibiotic action and emergence of resistance, better understanding of pathogenesis, optimizing antibiotic biosynthetic processes, and personalized medicine.

Acceleration of Emergence of Bacterial Antibiotic Resistance in Connected Microenvironments

Qiucen Zhang,[1] Guillaume Lambert,[1] David Liao,[2] Hyunsung Kim,[3] Kristelle Robin,[4] Chih-kuan Tung,[5] Nader Pourmand,[3] Robert H. Austin[1,4]*

The emergence of bacterial antibiotic resistance is a growing problem, yet the variables that influence the rate of emergence of resistance are not well understood. In a microfluidic device designed to mimic naturally occurring bacterial niches, resistance of *Escherichia coli* to the antibiotic ciprofloxacin developed within 10 hours. Resistance emerged with as few as 100 bacteria in the initial inoculation. Whole-genome sequencing of the resistant organisms revealed that four functional single-nucleotide polymorphisms attained fixation. Knowledge about the rapid emergence of antibiotic resistance in the heterogeneous conditions within the mammalian body may be helpful in understanding the emergence of drug resistance during cancer chemotherapy.

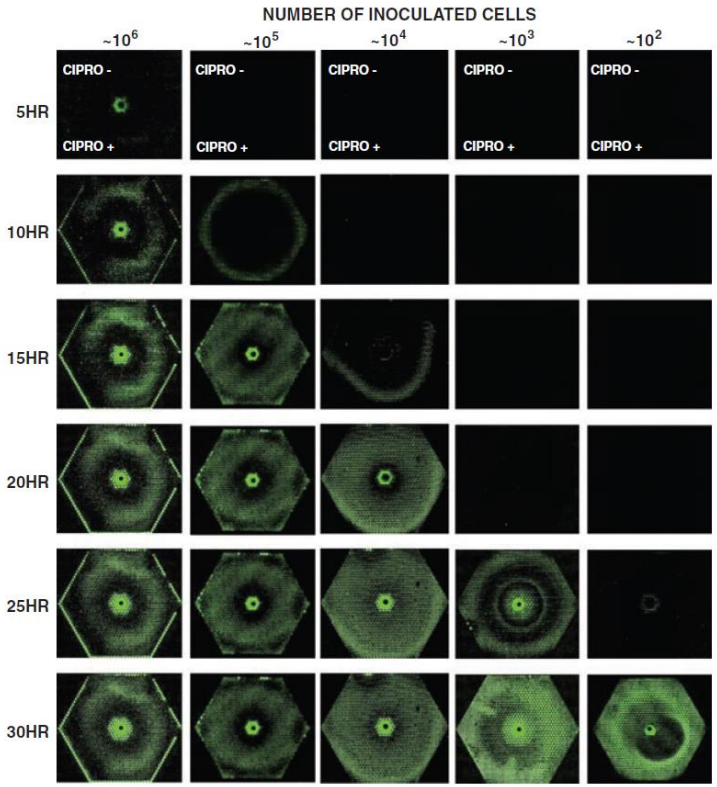


Fig. 2. Bacterial density versus time, shown for initial inoculations of 10^6, 10^5, 10^4, 10^3, and 10^2 wild-type *E. coli* with ciprofloxacin (1 μg/ml) in the bottom channel.

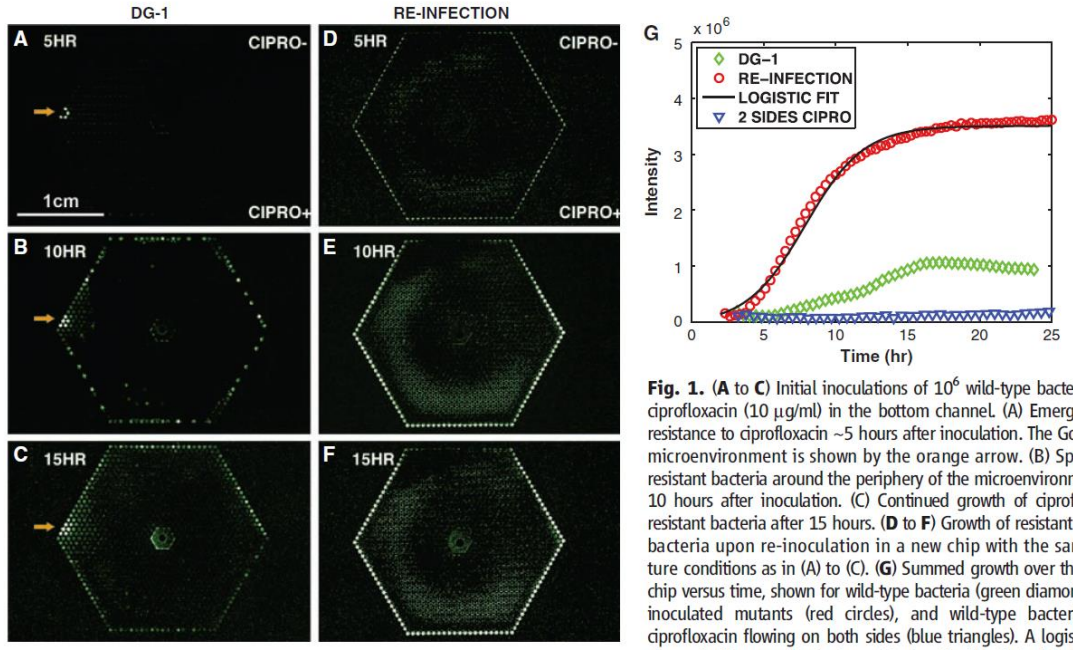
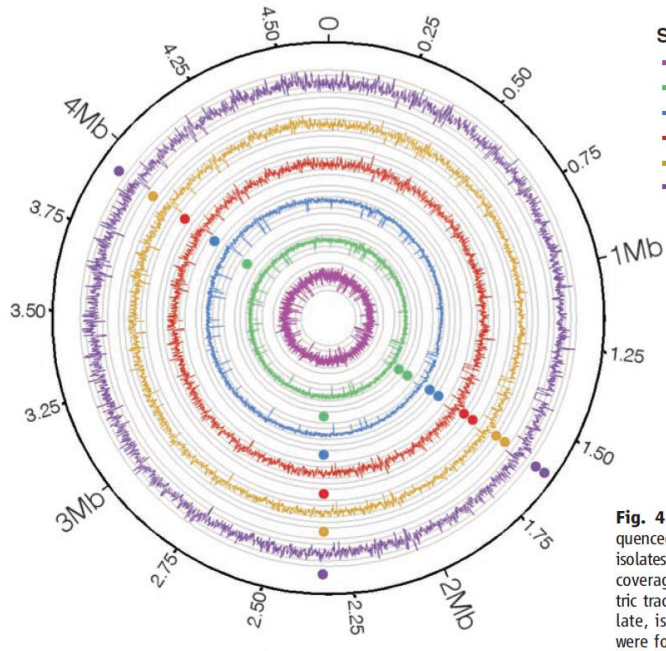


Fig. 1. (**A** to **C**) Initial inoculations of 10^6 wild-type bacteria with ciprofloxacin (10 μg/ml) in the bottom channel. (A) Emergence of resistance to ciprofloxacin ~5 hours after inoculation. The Goldilocks microenvironment is shown by the orange arrow. (B) Spread of resistant bacteria around the periphery of the microenvironments at 10 hours after inoculation. (C) Continued growth of ciprofloxacin-resistant bacteria after 15 hours. (**D** to **F**) Growth of resistant mutant bacteria upon re-inoculation in a new chip with the same culture conditions as in (A) to (C). (**G**) Summed growth over the entire chip versus time, shown for wild-type bacteria (green diamonds), re-inoculated mutants (red circles), and wild-type bacteria with ciprofloxacin flowing on both sides (blue triangles). A logistic fit is shown for the growth of reinoculated mutant bacteria (black line).



Strains

- Wild Type
- DG1-a
- DG1-b
- DG2-a
- DG2-b
- DG3

SNPs

Feature	Position	Mutation
gyrA	2,337,183	D87G
rbsA	3,933,247	T483I
marR	1,617,460	A105S
marR	1,617,461	A105S

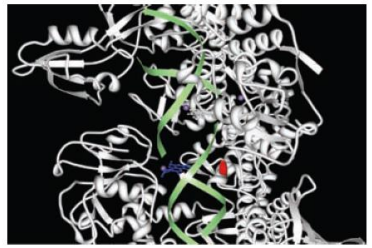
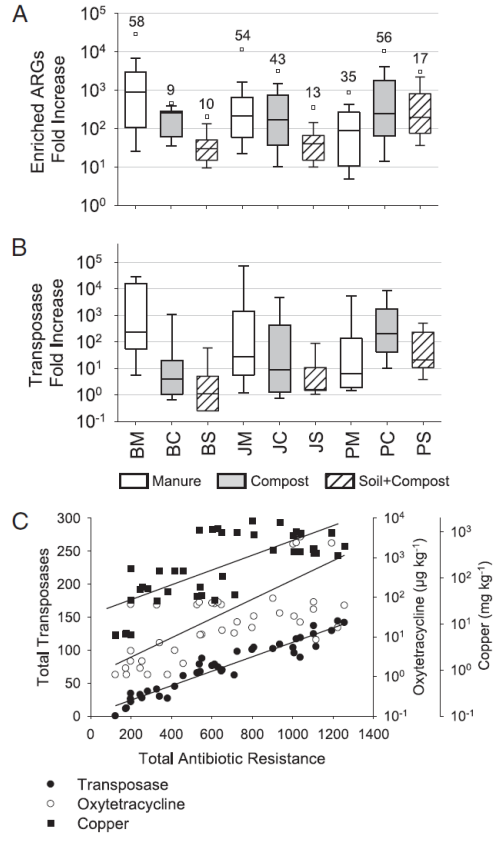


Fig. 4. Summary of sequencing results. Six genomes were sequenced: the wild type, two resistant isolates from DG-1, two resistant isolates from DG-2, and one resistant isolate from DG-3. Sequencing coverage for each isolate is plotted, according to color, on concentric tracks. Each SNP, represented by circles colored according to isolate, is marked at the appropriate genomic position. Four SNPs were found in five strains sampled from three independent experiments with initial inoculation of 10^6 *E. coli*. Lower right: X-ray structure of *gyrA* + double-stranded DNA (green helix), ciprofloxacin (blue), and the local SNP site (red) in *gyrA*.

Diverse and abundant antibiotic resistance genes in Chinese swine farms

Yong-Guan Zhu[a,b,1,2], Timothy A. Johnson[c,d,1], Jian-Qiang Su[a], Min Qiao[b], Guang-Xia Guo[b], Robert D. Stedtfeld[c,e], Syed A. Hashsham[c,e], and James M. Tiedje[c,d,2]

Antibiotic resistance genes (ARGs) are emerging contaminants posing a potential worldwide human health risk. Intensive animal husbandry is believed to be a major contributor to the increased environmental burden of ARGs. Despite the volume of antibiotics used in China, little information is available regarding the corresponding ARGs associated with animal farms. We assessed type and concentrations of ARGs at three stages of manure processing to land disposal at three large-scale (10,000 animals per year) commercial swine farms in China. In-feed or therapeutic antibiotics used on these farms include all major classes of antibiotics except vancomycins. High-capacity quantitative PCR arrays detected 149 unique resistance genes among all of the farm samples, the top 63 ARGs being enriched 192-fold (median) up to 28,000-fold (maximum) compared with their respective antibiotic-free manure or soil controls. Antibiotics and heavy metals used as feed supplements were elevated in the manures, suggesting the potential for coselection of resistance traits. The potential for horizontal transfer of ARGs because of transposon-specific ARGs is implicated by the enrichment of transposases—the top six alleles being enriched 189-fold (median) up to 90,000-fold in manure—as well as the high correlation (r^2 = 0.96) between ARG and transposase abundance. In addition, abundance of ARGs correlated directly with antibiotic and metal concentrations, indicating their importance in selection of resistance genes. Diverse, abundant, and potentially mobile ARGs in farm samples suggest that unmonitored use of antibiotics and metals is causing the emergence and release of ARGs to the environment.



The diversity and abundance of ARGs reported in this study is alarming and clearly indicates that unmonitored use of antibiotics and metals on swine farms has expanded the diversity and abundance of the antibiotic resistance reservoir in the farm environment. The coenrichment of ARGs and transposases further exacerbates the risks of transfer of ARGs from livestock animals to human-associated bacteria, and then spread among human populations (4, 6). Policies and management tools to facilitate prudent use of antibiotics and heavy metals, including their combined use, in animal industries and animal waste management are needed. Decreased resistance levels have been observed in Europe after the disuse of agricultural antibiotics (51). Pig manure, with its abundant and diverse ARGs and sheer volume, is a major source of resistance genes and as such a public health hazard. Microbes from manure, compost, or soil containing the ARGs are subject to dispersal via runoff into rivers (22), leaching to subsurface waters, air dispersal via dust, human travel, and distribution of agricultural products, including compost for gardening, which could expand a local contamination to regional and even global scales (6, 11).

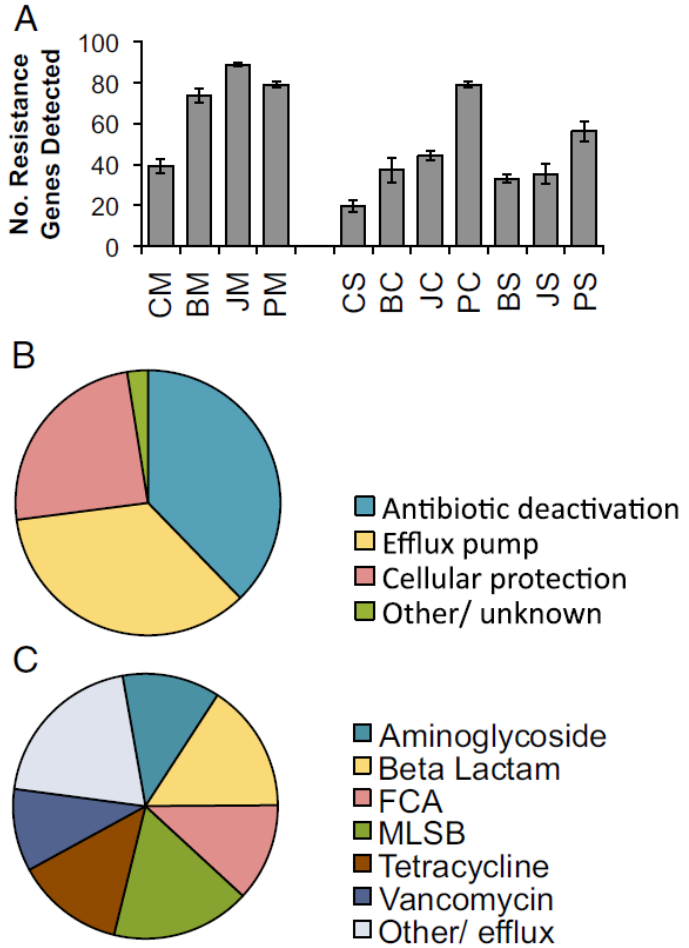


Fig. 1. Antibiotic resistance gene detection statistics. Sample names are abbreviated with two letters representing location and sample type: first C, B, J, and P (control, Beijing, Jiaxing, and Putian, respectively) and second M, C, and S [manure, compost, and soil (with compost amendment), respectively]. Because many resistance genes were targeted with multiple primers, if multiple primer sets detected the same gene, this was only counted as detection of a single unique resistance gene. (*A*) Average number of unique resistance genes detected in each sample. Error bars represent SEM of four field replicates. The resistance genes detected in all samples were classified based on (*B*) the mechanism of resistance, and (*C*) the antibiotic to which they confer resistance. FCA, fluoroquinolone, quinolone, florfenicol, chloramphenicol, and amphenicol resistance genes; MLSB, Macrolide-Lincosamide-Streptogramin B resistance.

Coping with antibiotic resistance: contributions from genomics

Gian Maria Rossolini[1]* and Maria Cristina Thaller[2]

Antibiotic resistance is a public health issue of global dimensions with a significant impact on morbidity, mortality and healthcare-associated costs. The problem has recently been worsened by the steady increase in multiresistant strains and by the restriction of antibiotic discovery and development programs. Recent advances in the field of bacterial genomics will further current knowledge on antibiotic resistance and help to tackle the problem. Bacterial genomics and transcriptomics can inform our understanding of resistance mechanisms, and comparative genomic analysis can provide relevant information on the evolution of resistant strains and on resistance genes and cognate genetic elements. Moreover, bacterial genomics, including functional and structural genomics, is also proving to be instrumental in the identification of new targets, which is a crucial step in new antibiotic discovery programs.

The role of genomics in surveillance and control of resistance

Tracing the epidemiology of resistant strains and resistance genes is of paramount importance for surveillance and control of antibiotic resistance, which can no longer rely on the simple phenotypic characterization of bacterial isolates. Molecular epidemiology is the discipline that studies the epidemiology of resistant strains and resistance genes by characterizing them at the molecular level, and it has provided major breakthroughs in the understanding of this phenomenon, with practical implications for resistance control strategies. The information provided by molecular analysis has a variable degree of resolution depending on the analytical tools, with genetic ones being the most versatile and of highest resolution. From this perspective, full genomic analysis of major resistant clones can be very informative for understanding their lifestyle and evolution and would be the golden standard for their comparison, as it was clearly shown with MRSA [12].

Genomic knowledge can also be instrumental to the development of sets of molecular probes for easy and specific identification of resistant clones of high spreading propensity and clinical impact, such as those described above, which is crucial in infection control practices. These probes can be used by reference laboratories or even by the largest diagnostic laboratories, in

Genomics can inform our understanding of resistance mechanisms

Understanding resistance mechanisms to novel drugs is crucial in the process of discovery and development of antimicrobial drugs. Knowledge on newly emerging resistance mechanisms to antibiotics already available for clinical use is also of paramount importance for the prediction of resistance evolution, antibiotic policies and resistance surveillance and control strategies. Genome-scale investigations may provide relevant insights into unknown mechanisms of antibiotic resistance.

For instance, a comparative genomic analysis between *S. aureus* strains showing heterogeneous or homogeneous intermediate resistance to vancomycin (hVISA or VISA strains) and the susceptible parent strain has recently revealed the role of mutations of the genes encoding the *vraSR* and *graSR* two-component regulatory system in conferring this resistance phenotype, which is associated with clinical failures of glycopeptides [15]. A similar

Genomic knowledge and antibacterial drug discovery

Knowledge on bacterial genomes has progressed at a fast pace during recent years, with almost 1,500 completed bacterial genomes and more than 600 additional genome projects in progress at the beginning of 2010 [13]. Indeed, since its very beginning, the advent of bacterial genomics was not only regarded as a fascinating scientific tool, but also raised great hopes for renewing the golden era of antimicrobial discovery at a time when this is sorely needed because of the growing impact of bacterial resistance. The rationale behind this expectation was that comparative genomic analysis could reveal valuable information on bacterial genes that presumably encode proteins that are essential to survival or fitness of bacterial pathogens and do not have close eukaryotic counterparts; these proteins could then be potential targets for new antimicrobial agents.

This approach has been extensively pursued and has returned some potential targets for new antimicrobial agents that fulfilled the above criteria, including enzymes involved in biosynthetic pathways (for example coenzyme A, chorismate, lipid A and fatty acids), protein synthesis (such as aminoacyl tRNA synthetases and peptide deformylase), protein secretion (such as signal peptidase 1) and DNA replication (such as FtsZ/FtsA) [20,21]. Some two-component signal transduction systems have also been considered [20].

However, it was soon evident that, although identification of potential targets is important, a comprehensive understanding of bacterial biochemistry, physiology and pathogenicity are essential for exploiting this information for antimicrobial drug discovery. After the identification of the potential targets, there are several bottlenecks to be overcome in the process of developing new drugs; in particular, the need to set up high-throughput screening of banks of small molecules to obtain potential hits, which has prompted great efforts in the fields of functional and structural genomics [22]. Moreover, many screenings have turned out to be unproductive, or it was discovered that effective drugs could not be developed from the potential inhibitors [20].

This accounts for the overall dearth of new drugs discovered using genomic approaches that have made it into advanced clinical phases of the antibiotic pipeline after almost 15 years of efforts in this area. However, the

The Lancet Infectious Diseases Commission

Antibiotic resistance—the need for global solutions

Ramanan Laxminarayan, Adriano Duse, Chand Wattal, Anita K M Zaidi, Heiman F L Wertheim, Nithima Sumpradit, Erika Vlieghe, Gabriel Levy Hara, Ian M Gould, Herman Goossens, Christina Greko, Anthony D So, Maryam Bigdeli, Göran Tomson, Will Woodhouse, Eva Ombaka, Arturo Quizhpe Peralta, Farah Naz Qamar, Fatima Mir, Sam Kariuki, Zulfiqar A Bhutta, Anthony Coates, Richard Bergstrom, Gerard D Wright, Eric D Brown, Otto Cars

The causes of antibiotic resistance are complex and include human behaviour at many levels of society; the consequences affect everybody in the world. Similarities with climate change are evident. Many efforts have been made to describe the many different facets of antibiotic resistance and the interventions needed to meet the challenge. However, coordinated action is largely absent, especially at the political level, both nationally and internationally. Antibiotics paved the way for unprecedented medical and societal developments, and are today indispensible in all health systems. Achievements in modern medicine, such as major surgery, organ transplantation, treatment of preterm babies, and cancer chemotherapy, which we today take for granted, would not be possible without access to effective treatment for bacterial infections. Within just a few years, we might be faced with dire setbacks, medically, socially, and economically, unless real and unprecedented global coordinated actions are immediately taken. Here, we describe the global situation of antibiotic resistance, its major causes and consequences, and identify key areas in which action is urgently needed.

Peter Chadwick/Science Photo Library

Figure 3: **Waste-water treatment facilities can be hotspots for horizontal transfer of resistance**

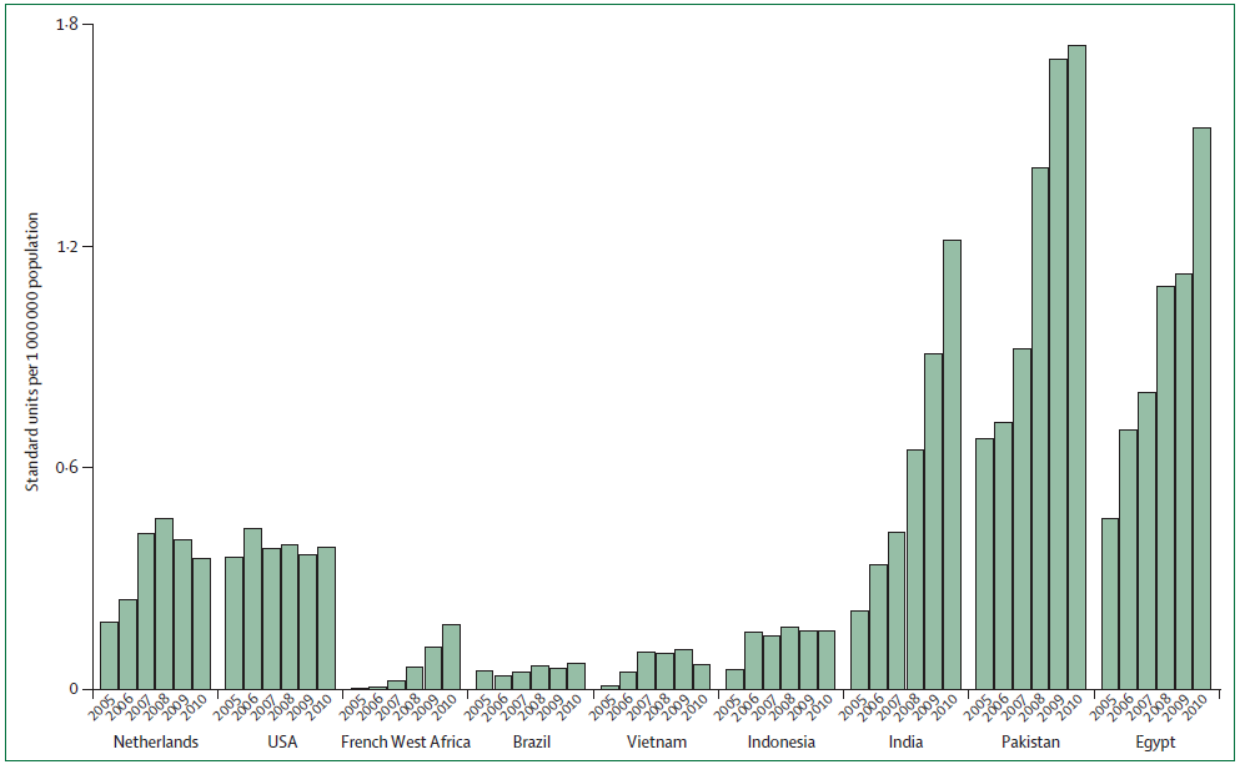


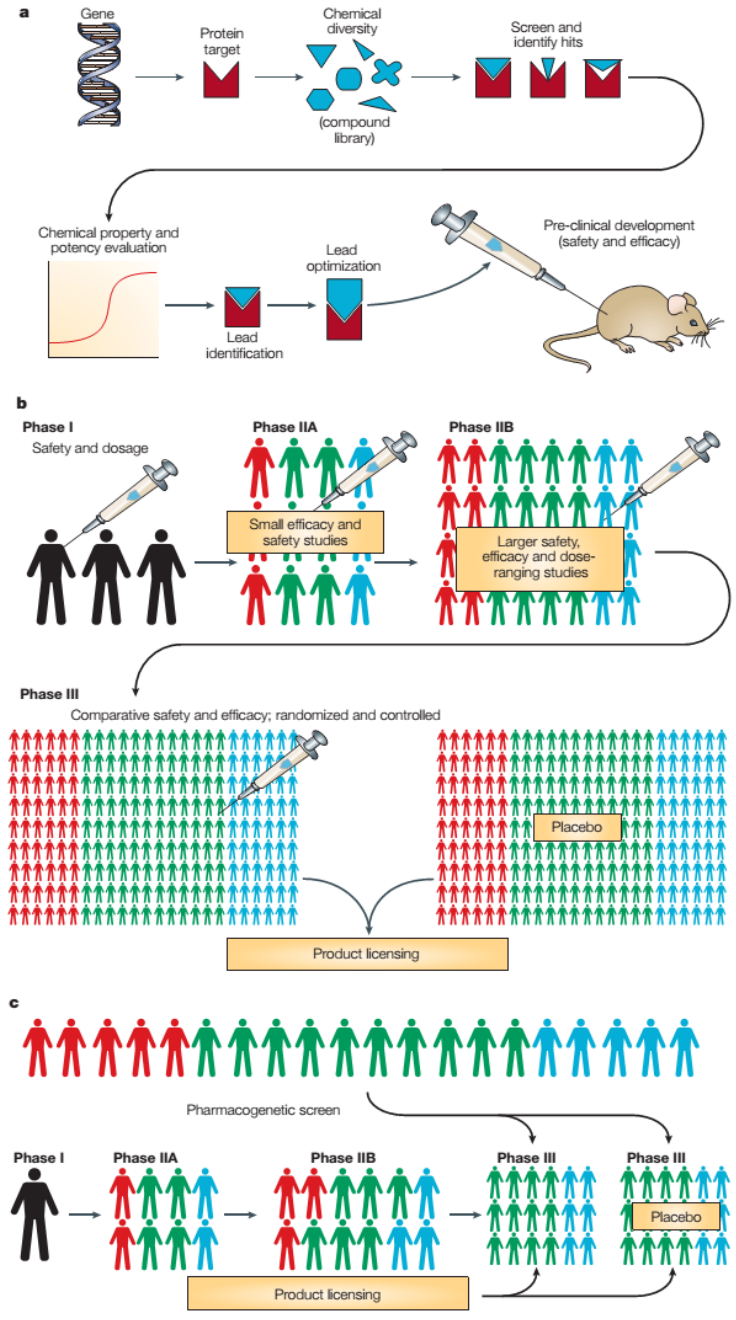
Figure 1: **Trends in retail sales of carbapenem antibiotics for Gram-negative bacteria**
Based on data obtained from IMS Health's MIDAS™ database. *An IMS grouping of Benin, Burkina Faso, Cameroon, Congo (Brazzaville), Gabon, Guinea, Ivory Coast, Mali, Senegal, and Togo.

Personalizirana medicina in farmakogenomika

Box 1 | **Pharmacogenetics and the pharmaceutical pipeline**

From having the sequence of a gene that encodes a potential target for drugs to actually having a medicine that interacts with that target involves a lengthy, expensive and complicated pharmaceutical pipeline that usually requires several years of basic science for target validation before chemical screening (see figure part **a**). Once there is a positive decision to progress a target, an effective screening assay that can allow the high throughput of many thousands of known chemical entities must be designed and implemented. Molecules that affect the target (hits) must then be evaluated for chemical properties and potency before those that are worth pursuing (leads) can be identified, synthesized, evaluated and modified for drug qualities (lead optimization). When a lead is identified, considerable pre-clinical development must also occur, particularly in the fields of toxicology, drug kinetics and drug metabolism. All of these processes must occur before the first dose of any new molecule can be tested in humans (part **b**).

Following initial clinical testing for safety in humans (PHASE I), the molecule enters the most crucial phase, PHASE IIA, during which the desired clinical effect (that is, efficacy) is addressed in a relatively small, but still expensive, clinical trial. For example, if 100 patients participate in a Phase-IIA study at a minimum cost of US $10,000 per patient, the study would cost US $1 million. Even after a molecule demonstrates efficacy, more extensive dose-ranging PHASE-IIB studies, as well as PHASE-III studies, cost several hundreds of millions of dollars. With costs as high as this, only molecules for which there is good evidence of efficacy and a reasonable biological rationale for its mechanism of action are selected for full development. Prospective efficacy pharmacogenetics early in clinical development can be used to select patients who would be predicted to be potential responders (green figures in parts **b** and **c**) and hyper-responders (blue figures). By selecting out patients in Phase-II and -III clinical trials who would be predicted to be non-responders (red figures in parts **b** and **c**), subsequent clinical trials can be made smaller, faster and less expensive (part **c**).



PHARMACOGENETICS AND DRUG DEVELOPMENT: THE PATH TO SAFER AND MORE EFFECTIVE DRUGS

Allen D. Roses

Abstract | Pharmacogenetics provides opportunities for informed decision-making along the pharmaceutical pipeline. There is a growing literature of retrospective studies of marketed medicines that describe efficacy or safety on the basis of patient genotypes. These studies emphasize the potential prospective use of genome information to enhance success in finding new medicines. An example of a prospective efficacy pharmacogenetic Phase-IIA proof-of-concept study is described. Inserting a rapidly performed efficacy pharmacogenetic step after initial clinical data are obtained can provide confidence for a commitment to full drug development. The rapid identification of adverse events during and after drug development using genomic mapping tools is also reviewed.

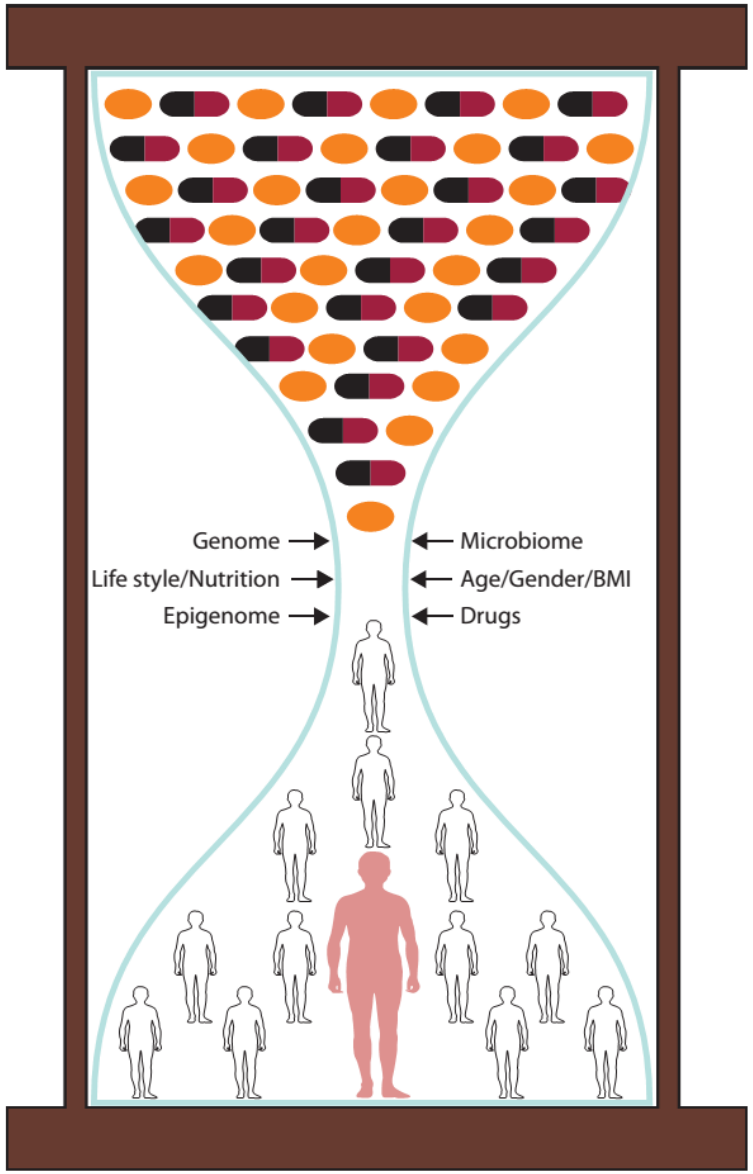
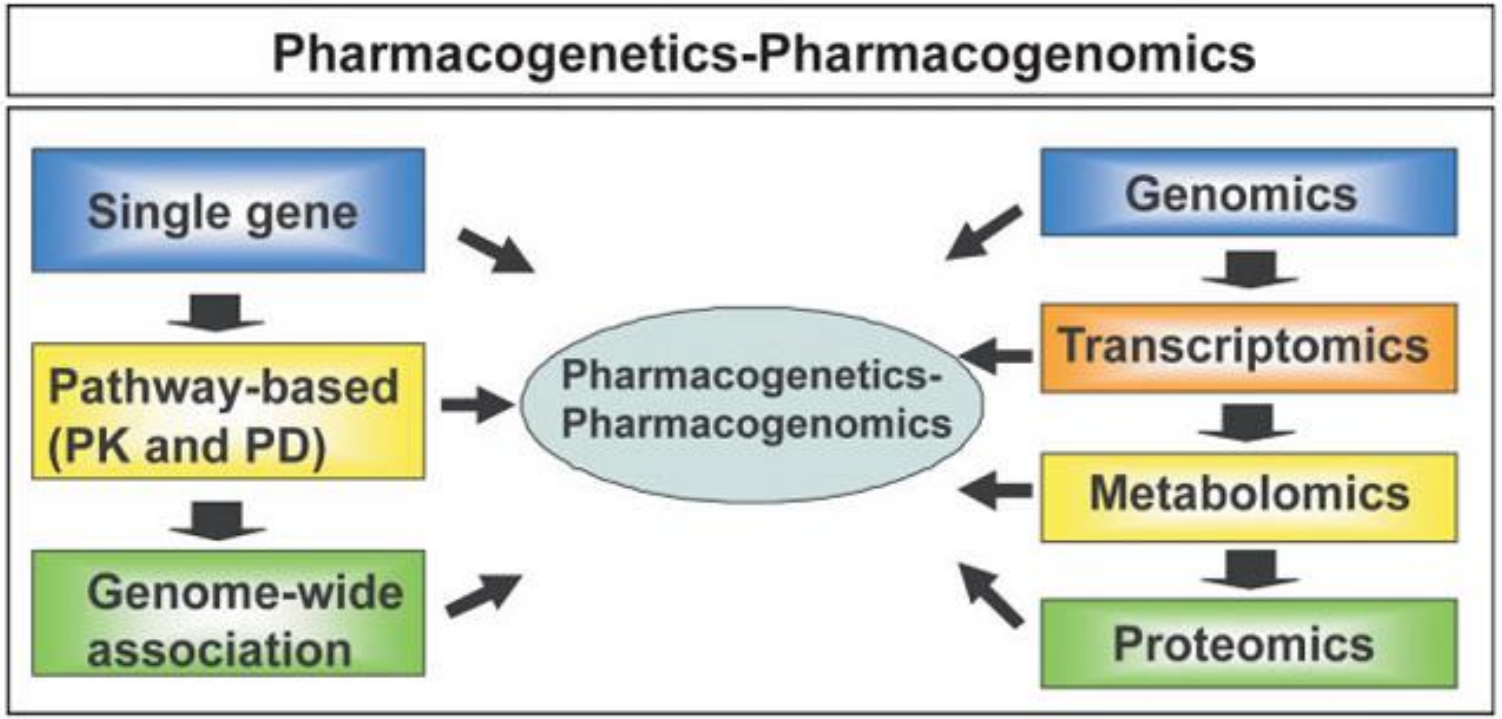


Figure 1. Pharmacogenomics. Interindividual variation in drug response is the consequence of a combination of genetic and environmental factors as well as patient characteristics, which affect the pharmacokinetics and/or pharmacodynamics of drugs. Pharmacogenomics affects not only therapeutic efficacy but also disease susceptibility and drug development. BMI, body mass index.

TABLE 1
Major factors affecting individual drug response

Factors	Effects
Genetic Factors	Major variables; stable and inherited
Therapeutic targets	Drug efficacy (pharmacodynamics)
Drug-metabolizing enzymes	Drug metabolism (pharmacokinetics)
Drug transporters	Drug disposition (pharmacokinetics)
Targets of adverse drug reactions	Drug toxicity (pharmacodynamics and pharmacokinetics)
Factors with indirect effects	Drug efficacy, pharmacokinetics, and toxicity
Other Factors	Mostly transient
Environmental factors	Drug efficacy, pharmacokinetics, and toxicity
Environmental chemicals, coadministered drugs, tobacco smoking, alcohol drinking, and dietary constituents	
Physiological factors	Drug efficacy, pharmacokinetics, and toxicity
Age, sex, disease state, pregnancy, exercise, circadian rhythm, and starvation	



The evolution of pharmacogenetics and pharmacogenomics. Pharmacogenomics has evolved from a single gene approach to incorporate pathway-based and genome-wide approaches (left side of the diagram). In parallel, it has increasingly incorporated a variety of high-throughput technologies including genomics, transcriptomics, metabolomics, and proteomics to significantly enhance the ability to generate and test pharmacogenomic hypotheses and to translate those hypotheses into clinical practice. PK, pharmacokinetics; PD, pharmacodynamics.

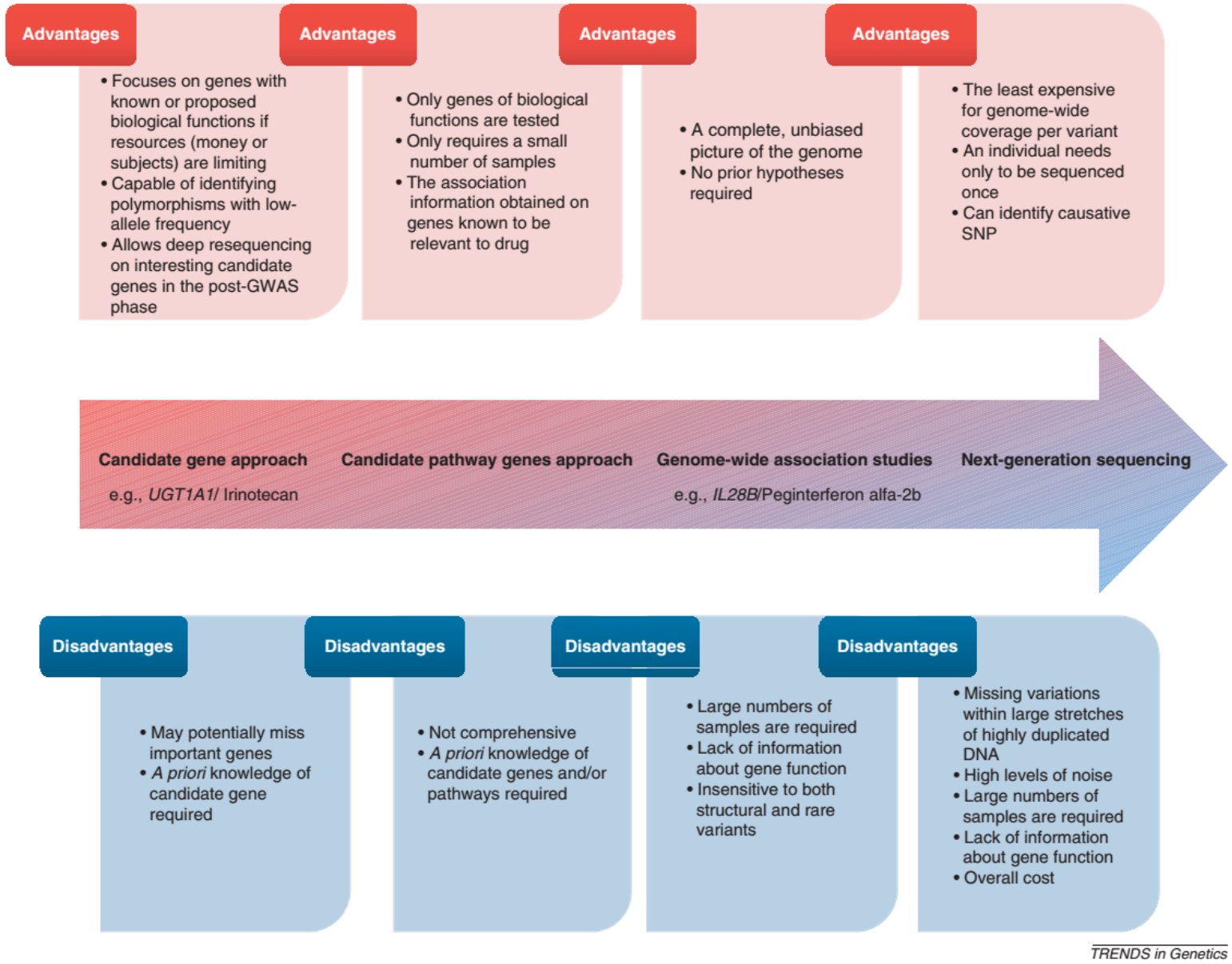
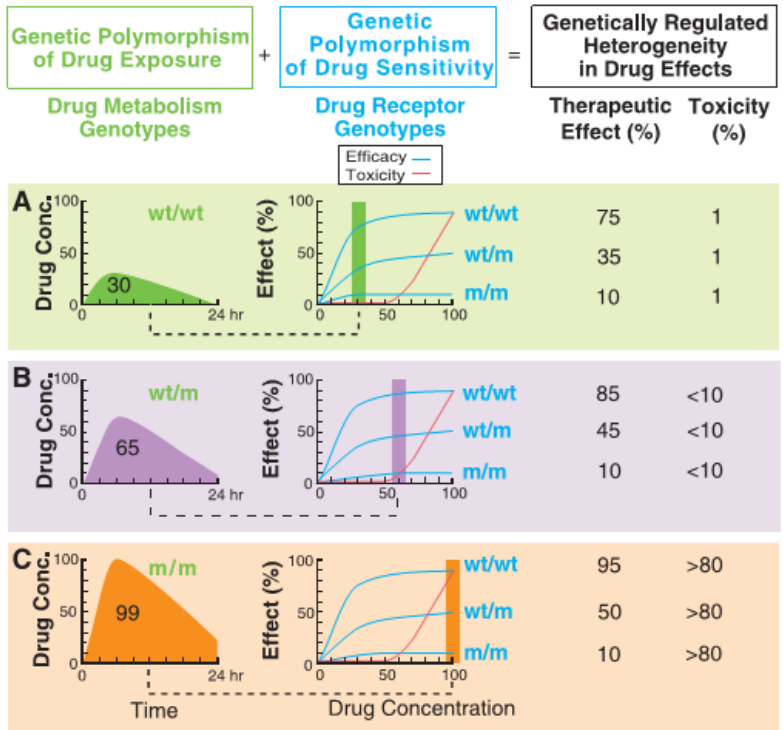


Figure 1. The evolution of pharmacogenomics. The diagram depicts the different advantages and disadvantages for candidate gene approaches, candidate pathway genes approaches, genome-wide association studies (GWAS), and next-generation sequencing. New techniques (e.g., GWAS) do not necessarily replace old strategies (e.g., candidate gene approach). Examples of the candidate gene approach include the gene–drug pair UDP glucuronosyltransferase 1 family, polypeptide A1 (*UGT1A1*)–irinotecan. Examples of GWAS include the gene–drug pair interleukin 28B (*IL28B*)–peginterferon alfa-2b.

Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics

William E. Evans* and Mary V. Relling

Genetic polymorphisms in drug-metabolizing enzymes, transporters, receptors, and other drug targets have been linked to interindividual differences in the efficacy and toxicity of many medications. Pharmacogenomic studies are rapidly elucidating the inherited nature of these differences in drug disposition and effects, thereby enhancing drug discovery and providing a stronger scientific basis for optimizing drug therapy on the basis of each patient's genetic constitution.



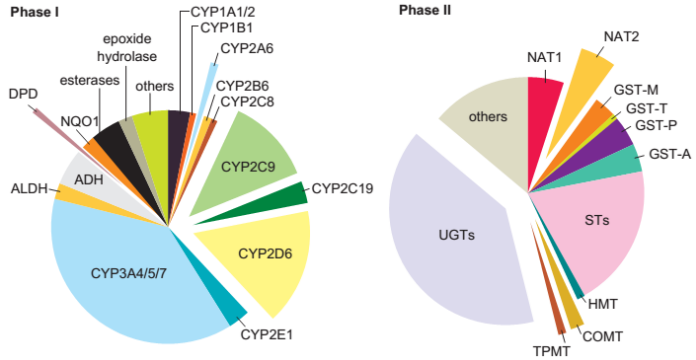


Fig. 2. Most drug-metabolizing enzymes exhibit clinically relevant genetic polymorphisms. Essentially all of the major human enzymes responsible for modification of functional groups [classified as phase I reactions (**left**)] or conjugation with endogenous substituents [classified as phase II reactions (**right**)] exhibit common polymorphisms at the genomic level; those enzyme polymorphisms that have already been associated with changes in drug effects are separated from the corresponding pie charts. The percentage of phase I and phase II metabolism of drugs that each enzyme contributes is estimated by the relative size of each section of the corresponding chart. ADH, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; CYP, cytochrome P450; DPD, dihydropyrimidine dehydrogenase; NQO1, NADPH:quinone oxidoreductase or DT diaphorase; COMT, catechol *O*-methyltransferase; GST, glutathione *S*-transferase; HMT, histamine methyltransferase; NAT, *N*-acetyltransferase; STs, sulfotransferases; TPMT, thiopurine methyltransferase; UGTs, uridine 5′-triphosphate glucuronosyltransferases.

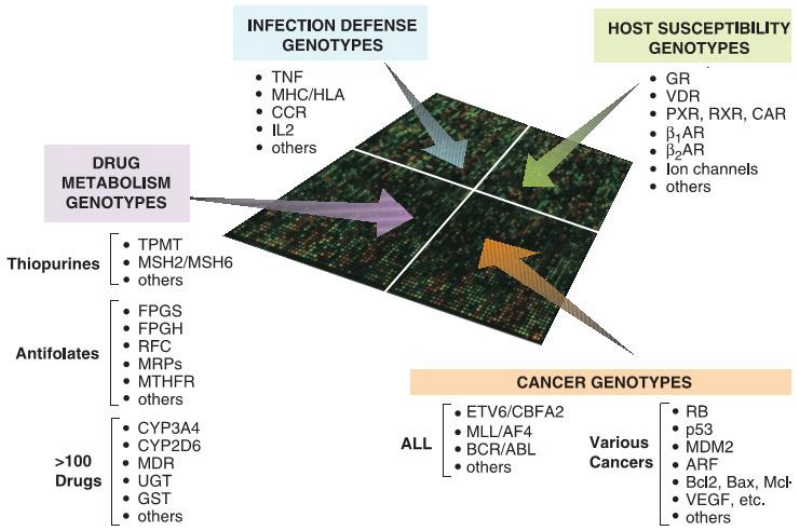


Fig. 3. Molecular diagnostics of pharmacogenomic traits. DNA arrays are being made for automated, high-throughput detection of functionally important mutations in genes that are important determinants of drug effects, such as drug-metabolizing enzymes, drug targets (receptors), disease pathogenesis, and other polymorphic genes that influence an individual's susceptibility to drug toxicities or environmental exposures (such as pathogens, carcinogens, and others). This figure exemplifies components of a potential diagnostic DNA array for genes that could influence a patient's response to chemotherapy for acute lymphoblastic leukemia, including genes that determine drug metabolism, disease sensitivity, and the risk of adverse effects of treatment (cardiovascular or endocrine toxicities, infections, and so forth).

Genomics and Drug Response

PHARMACOGENOMICS IS THE STUDY OF THE ROLE OF INHERITED AND ACquired genetic variation in drug response.[1] Clinically relevant pharmacogenetic examples, mainly involving drug metabolism, have been known for decades, but recently, the field of pharmacogenetics has evolved into "pharmacogenomics," involving a shift from a focus on individual candidate genes to genome-wide association studies. Such studies are based on a rapid scan of markers across

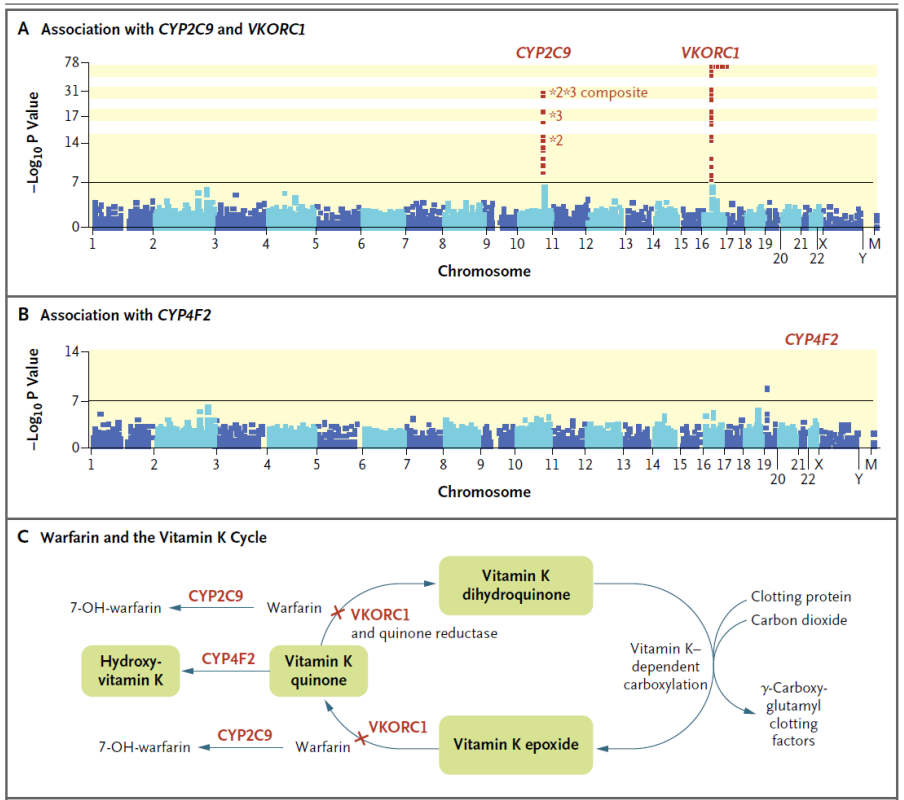


Figure 1. Warfarin Pharmacogenomics.

Panels A and B show Manhattan plots of P values (negative $\log_{10}$) for the association between single-nucleotide-polymorphisms (SNPs) across the genome and the final warfarin dose. The horizontal line indicates a P value of 1.5×10^{-7}, which is the level of genomewide statistical significance. In Panel A, the results of univariate regression analysis highlight SNP signals in or near *CYP2C9* and *VKORC1*. In Panel B, the results of multivariate regression analysis with adjustment for the contributions of *CYP2C9* and *VKORC1* show the *CYP4F2* signal on chromosome 19. (Data are from Takeuchi et al.[18]) The label *2 indicates the nonsynonymous SNP rs1799853, *3 indicates the nonsynonymous SNP rs1057910, and the *2*3 composite indicates the SNP rs4917639. M denotes mitochondrial SNPs. Panel C shows the sites of action of warfarin in the vitamin K cycle, as well as the roles of CYP2C9, CYP4F2, and VKORC1 in this process.

Table 1. Anticancer Drugs Approved by the Food and Drug Administration (FDA) with Labeling Regarding Pharmacogenomic Biomarkers.*

Type of Biomarker and Associated Drug
Biomarker with pharmacokinetic effect
TPMT
Mercaptopurine
Thioguanine
UGT1A1
Irinotecan
Nilotinib
Biomarker with pharmacodynamic effect
EGFR
Cetuximab
Erlotinib
Gefitinib
Panitumumab
KRAS
Cetuximab
Panitumumab
ABL
Imatinib
Dasatinib
Nilotinib
C-Kit (*KIT*)
Imatinib
HER2/neu (*ERBB2*)
Lapatinib
Trastuzumab
Estrogen receptor
Tamoxifen

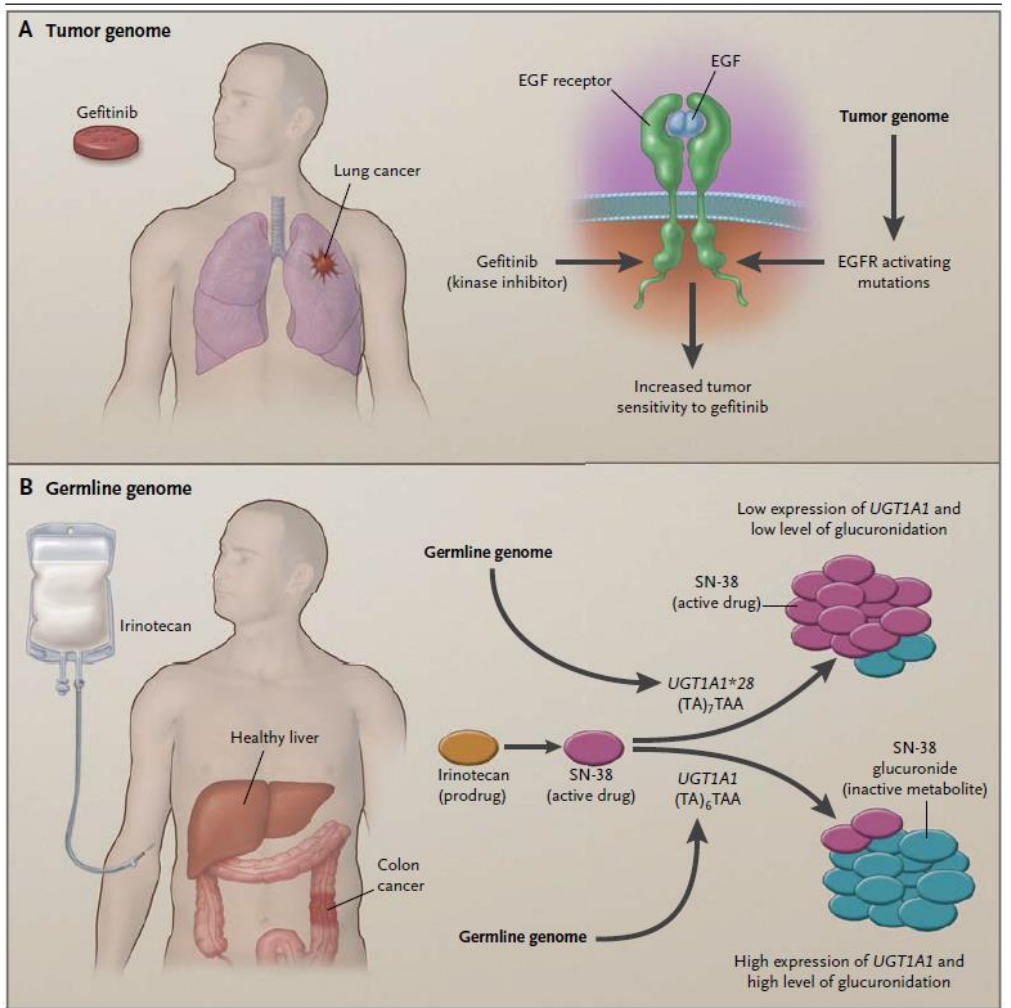


Figure 3. Cancer Pharmacogenomics and Tumor and Germline Genomes.

Both the tumor genome (e.g., in the case of gefitinib therapy) and the patient's germline genome (e.g., in the case of irinotecan therapy) can contribute to pharmacogenomic variation in response to antineoplastic drugs. The tumor genome plays a critical role in the response to gefitinib (Panel A), since the sensitivity of non–small-cell lung cancer to this drug is enhanced by activating mutations in the kinase domain of the gene encoding epidermal growth factor receptor (*EGFR*).[58,59] Tumor *EGFR* encoding activating mutations within the kinase domain results in enhanced tumor sensitivity to gefitinib. The rate of toxic effects associated with irinotecan (diarrhea and myelosuppression) is increased in patients with seven TA dinucleotide repeats rather than the more common six repeats in the promoter region of *UGT1A1* encoding a UDP-glucuronosyltransferase in germline DNA, resulting in lower enzyme activity and a decreased rate of drug metabolism (Panel B).[1,62]

Cancer pharmacogenomics: strategies and challenges

Heather E. Wheeler[1,2], Michael L. Maitland[1,2,3], M. Eileen Dolan[1,2,3], Nancy J. Cox[1,3,4] and Mark J. Ratain[1,2,3]

Abstract | Genetic variation influences the response of an individual to drug treatments. Understanding this variation has the potential to make therapy safer and more effective by determining selection and dosing of drugs for an individual patient. In the context of cancer, tumours may have specific disease-defining mutations, but a patient's germline genetic variation will also affect drug response (both efficacy and toxicity), and here we focus on how to study this variation. Advances in sequencing technologies, statistical genetics analysis methods and clinical trial designs have shown promise for the discovery of variants associated with drug response. We discuss the application of germline genetics analysis methods to cancer pharmacogenomics with a focus on the special considerations for study design.

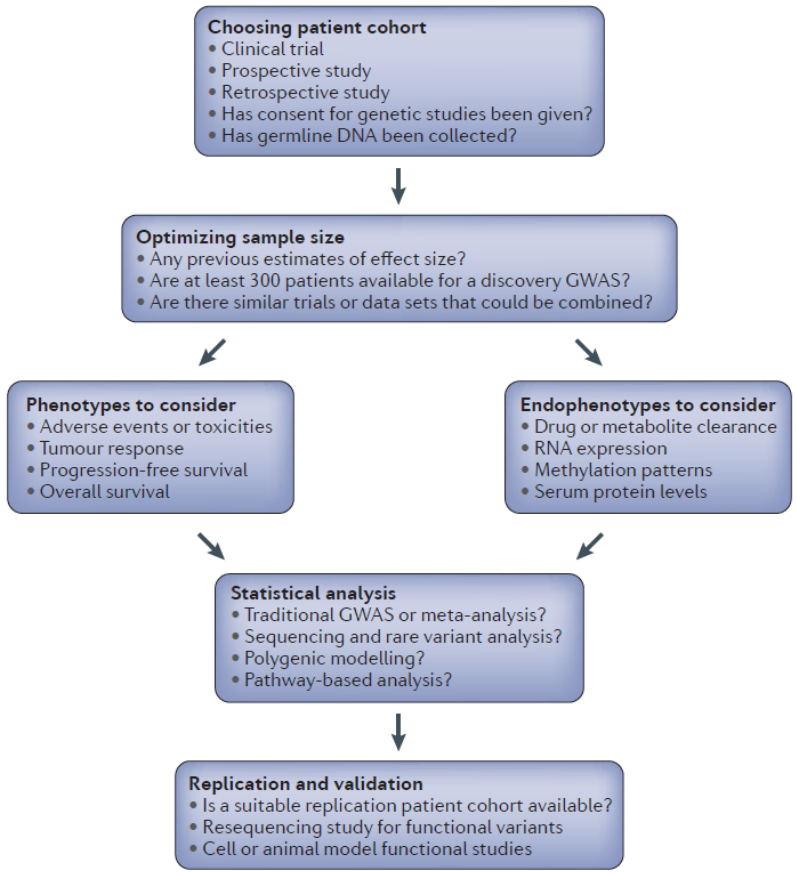


Figure 1 | **Steps in cancer pharmacogenomic study design.** This flow diagram outlines the main steps in a cancer pharmacogenomic study design. In addition to making these key decisions, potential covariate data should be collected, as discussed in BOX 2. GWAS, genome-wide association study.

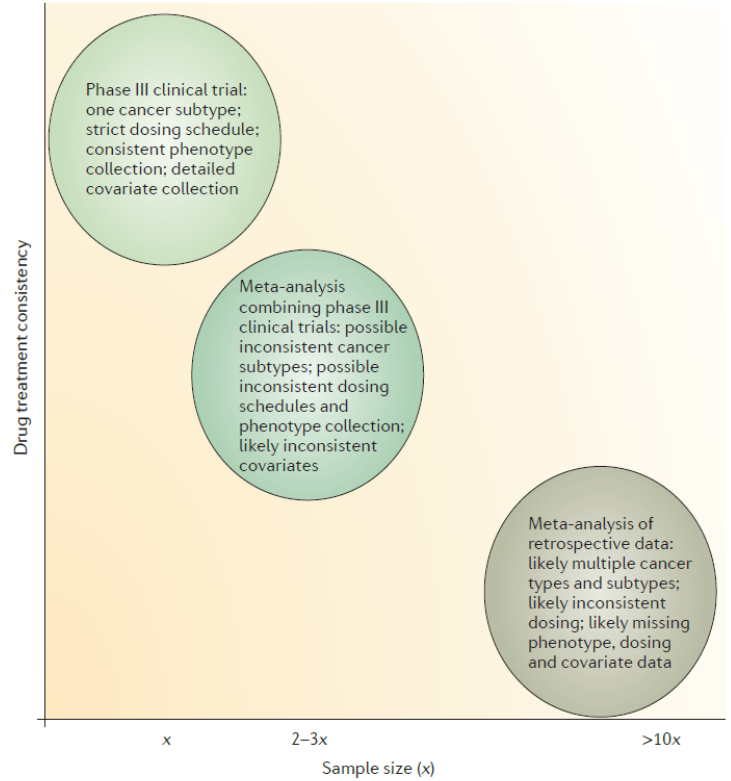
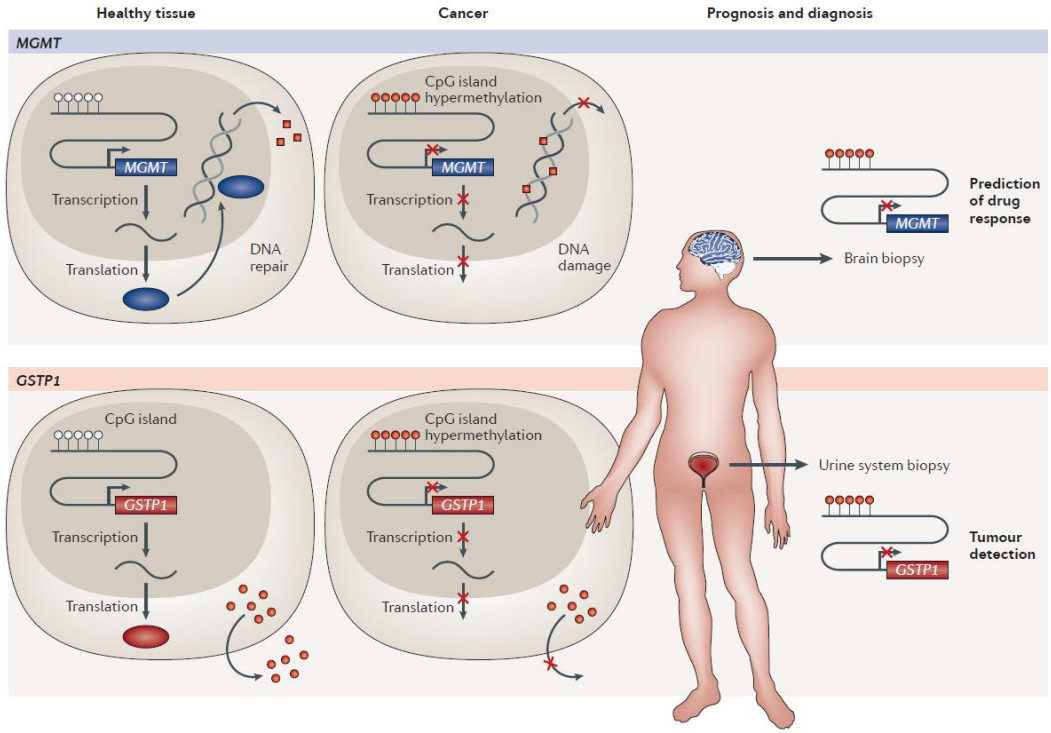


Figure 2 | **Negative relationship between sample size and drug treatment consistency in cancer pharmacogenomics.** To test for replication of findings from preliminary genome-wide association studies (GWASs), it is necessary to combine data sets from multiple trials and retrospective patient collections. Therefore, the phenotype and covariate data become less consistent, increasing the potential for confounding variables. The sample size (x) will vary depending on the prevalence of the type of cancer under study and the prevalence of the drug's use when collecting retrospective data, but often *x* is ~1,000.

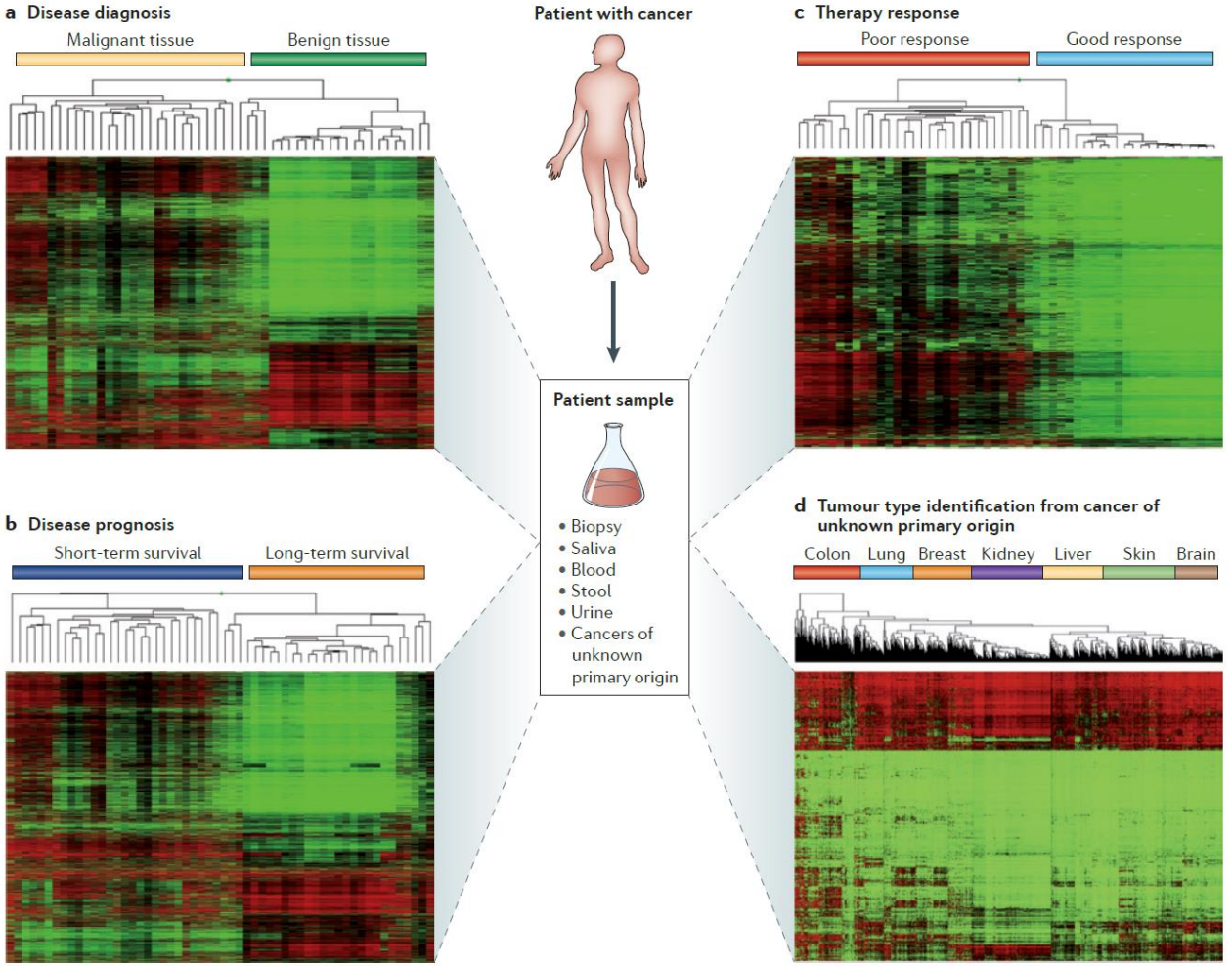
DNA methylation profiling in the clinic: applications and challenges

Holger Heyn[1] and Manel Esteller[1,2,3]

Abstract | Knowledge of epigenetic alterations in disease is rapidly increasing owing to the development of genome-wide techniques for their identification. The ever-growing number of genes that show epigenetic alterations in disease emphasizes the crucial role of these epigenetic alterations — particularly DNA methylation — for future diagnosis, prognosis and prediction of response to therapies. This Review focuses on epigenetic profiling, which has started to be of clinical value in cancer and may in the future be extended to other diseases, such as neurological and autoimmune disorders.



GSTP1 and MGMT: case examples for epigenetic profiling in diagnosis and prognosis. O6-methylguanine-DNA methyltransferase (MGMT) protects normal cells against transition mutations by removing alkyl groups (red squares), which have been introduced by carcinogens such as nitrosamides, from guanine bases. Glutathione S-transferase pi 1 (GSTP1) has been established as biomarker for prostate cancer diagnosis and prognosis.



High-resolution screening of patient samples in disease diagnosis, prognosis, prediction of drug response and tumour-type identification. Future cancer diagnosis, prognosis and therapy will benefit from epigenetic high-resolution screening technologies. Profiling and subsequent classification can be carried out on primary tissue or biological fluids, such as saliva, blood, stool and urine.

Clan Genomics and the Complex Architecture of Human Disease

James R. Lupski,[1,2,3,*] John W. Belmont,[1,2] Eric Boerwinkle,[4,5] and Richard A. Gibbs[1,5,*]
[1]Department of Molecular and Human Genetics
[2]Department of Pediatrics
Baylor College of Medicine, Houston, TX 77030, USA
[3]Texas Children's Hospital
[4]Human Genetics Center
University of Texas Health Science Center at Houston, Houston, TX 77030-1501, USA
[5]The Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA
*Correspondence: jlupski@bcm.edu (J.R.L.), agibbs@bcm.edu (R.A.G.)
DOI 10.1016/j.cell.2011.09.008

Human diseases are caused by alleles that encompass the full range of variant types, from single-nucleotide changes to copy-number variants, and these variations span a broad frequency spectrum, from the very rare to the common. The picture emerging from analysis of whole-genome sequences, the 1000 Genomes Project pilot studies, and targeted genomic sequencing derived from very large sample sizes reveals an abundance of rare and private variants. One implication of this realization is that recent mutation may have a greater influence on disease susceptibility or protection than is conferred by variations that arose in distant ancestors.

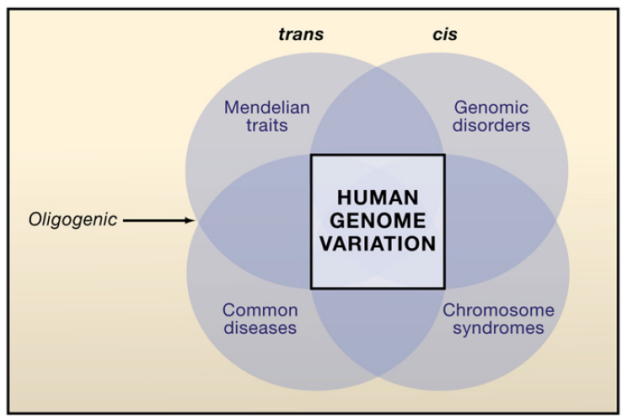


Figure 6. A Continuum for the Genetics of Human Disease
The square (center) represents genomic variation that can influence the different categories of genetic disease. The circles represent the overlapping categories of human disease with darker regions depicting intersection with greater overlap in the underlying genetic influences on these given disease categories. A unified model for human genetic disease proposes that all major categories of disease with genetic influence—Mendelian disease, common disease or complex traits, genomic disorders, and chromosomal syndromes—can be explained by variation in DNA sequence (SNV) or copy number (CNV) from a "wild-type" diploid state. Whereas *trans*-genetic interactions at a single locus (alleles) or between loci may contribute to Mendelian disease and complex traits, *cis*-genetic interactions can be important to phenotypic manifestations in genomic disorders (CNV) and chromosomal syndromes (segmental aneuploidy). Digenic and triallelic inheritance bridge Mendelian traits and complex disease; each represents an oligogenic inheritance model.

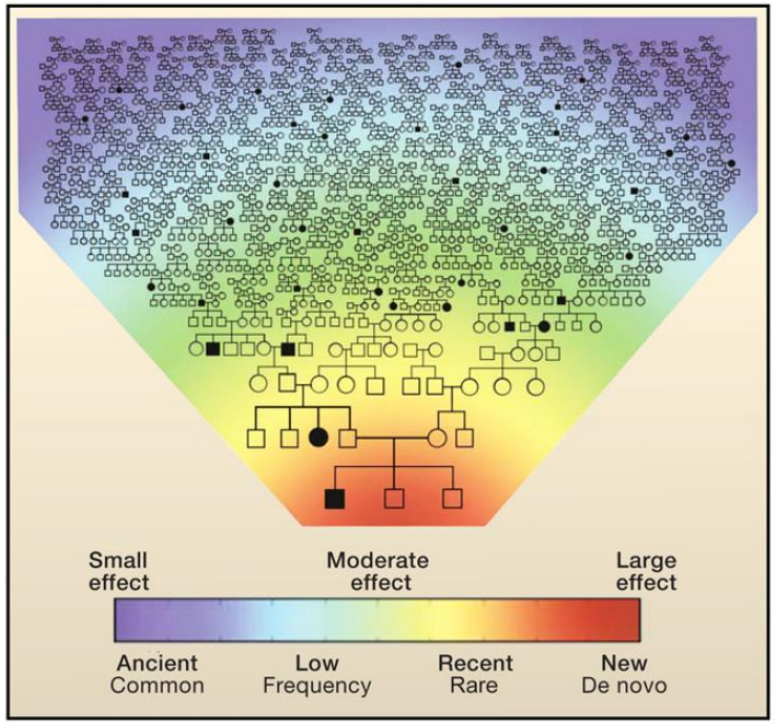


Figure 1. Clan Genomics
Heat map and extended pedigree showing the conceptual relationship among de novo mutations leading to disease (red), recent mutations with moderate effects arising within a clan (yellow and green), and older common variants with small effects segregating in the population (blue). An individual's genetic disease risk emerges from the collection of variants he or she has inherited from both parental lineages of distant ancestors (typically common and of individually small effect), more recent ancestors (rare, but potentially larger effect), and de novo mutations.

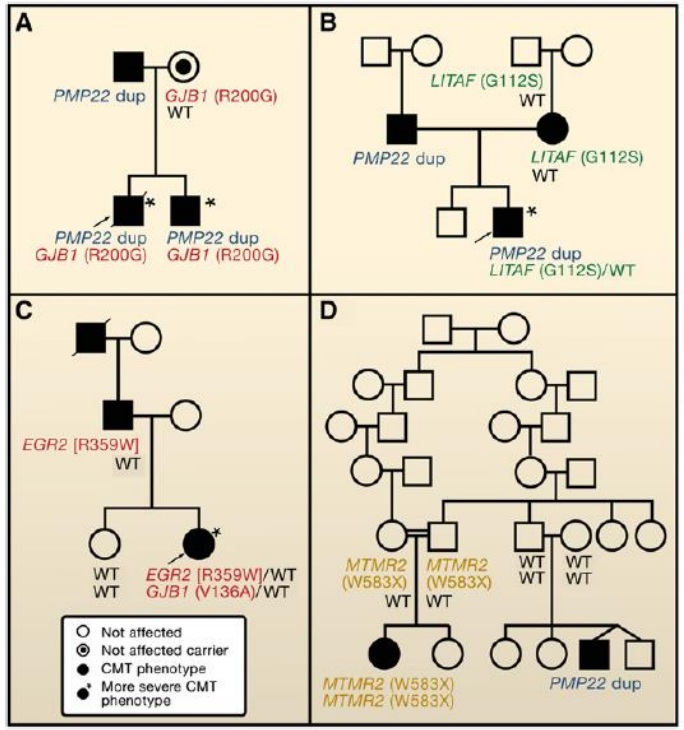
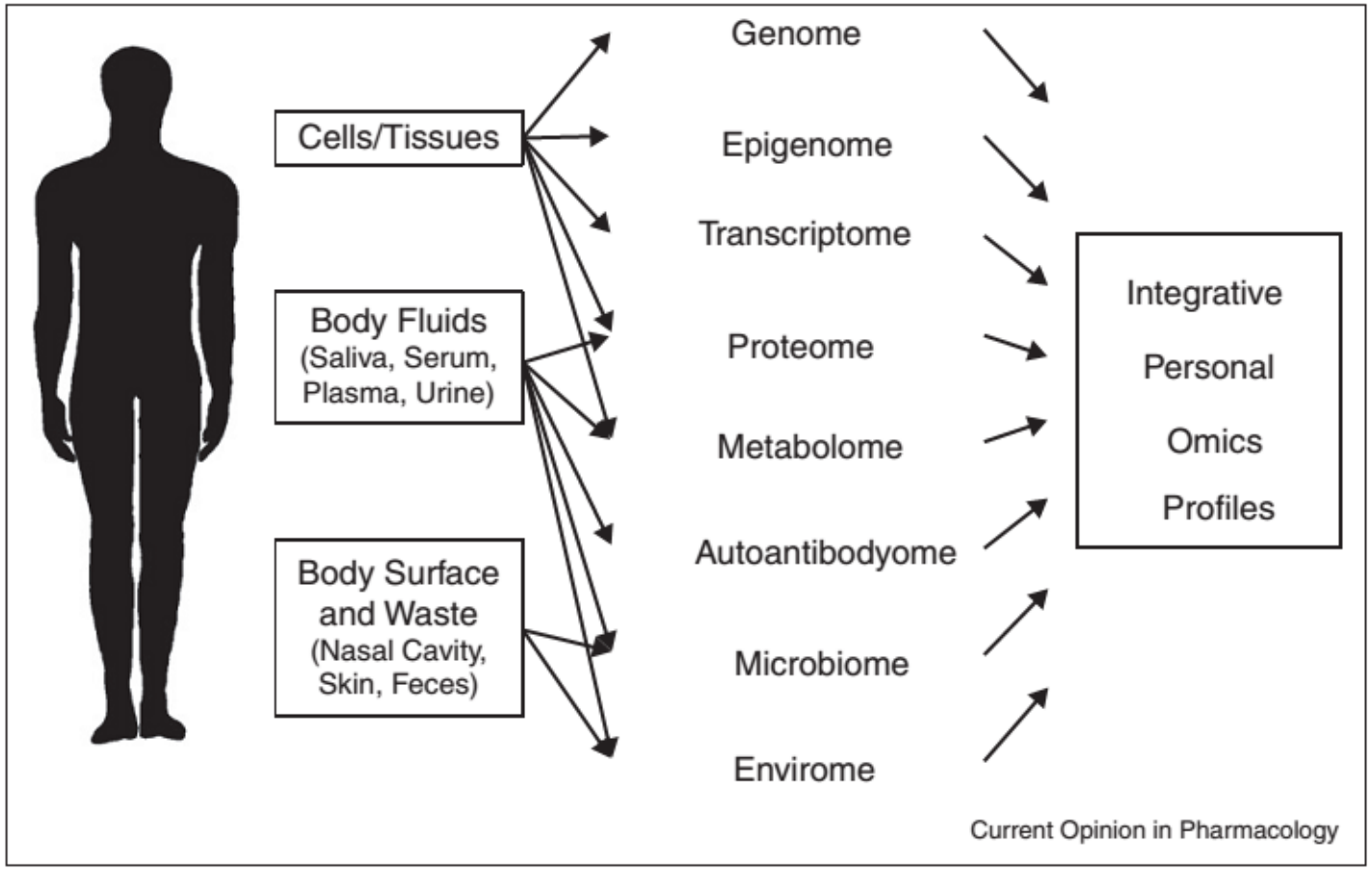


Figure 4. Totality of Pathogenic Variants, Disease Severity, and Clan Genomics
Pedigrees of families segregating Charcot-Marie-Tooth (CMT) neuropathy, illustrating that disease severity is directly related to pathogenic mutational burden.
(A–C) Mutations at two different CMT loci result in a more severe phenotype. These double heterozygotes may be due to either a single-nucleotide variant (SNV) + copy-number variant (CNV) (A and B) or two SNV (C) (Chung et al., 2005; Hodapp et al., 2006; Meggouh et al., 2005).
(D) In a single family, disease results from homozygous *MTMR2* mutation (likely related to consanguinity) or de novo CNV—the CMT1A duplication (*PMP22*) (Verny et al., 2004); an example of clan genomics.

Systems biology: personalized medicine for the future?

Rui Chen and Michael Snyder

Systems biology is actively transforming the field of modern health care from symptom-based disease diagnosis and treatment to precision medicine in which patients are treated based on their individual characteristics. Development of high-throughput technologies such as high-throughout sequencing and mass spectrometry has enabled scientists and clinicians to examine genomes, transcriptomes, proteomes, metabolomes, and other omics information in unprecedented detail. The combined 'omics' information leads to a global profiling of health and disease, and provides new approaches for personalized health monitoring and preventative medicine. In this article, we review the efforts of systems biology in personalized medicine in the past 2 years, and discuss in detail achievements and concerns, as well as highlights and hurdles for future personalized health care.



Conclusion

Personalized medicine is the future direction of health care and systems biology serves as the enabling force. Despite various clinical and technological concerns, we still believe that personalized health monitoring and preventative medicine will greatly improve the health of the general public. We are envisioning that in the near

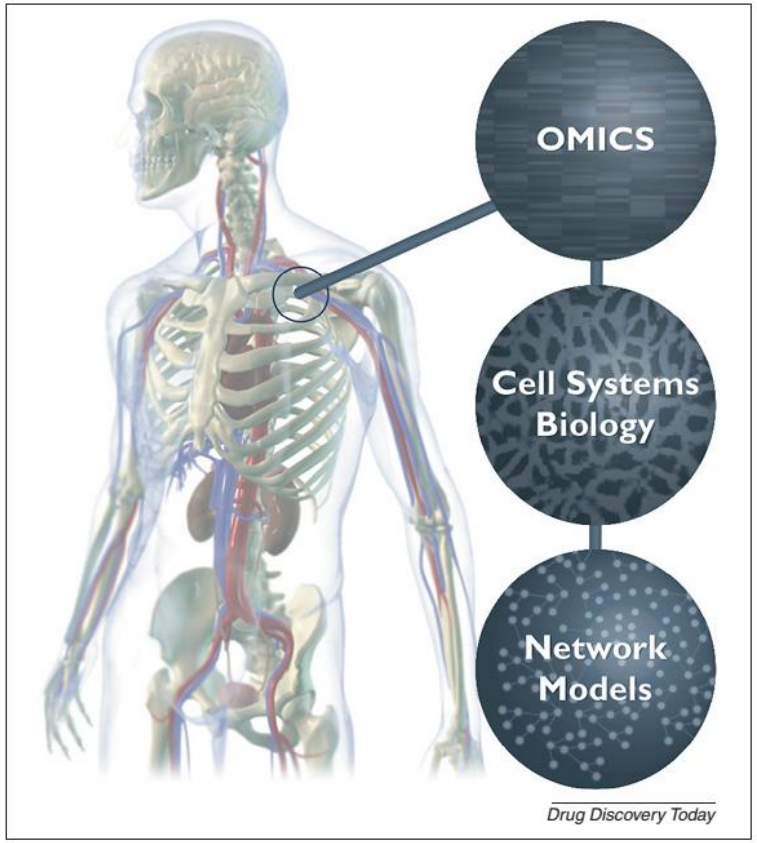


FIGURE 2

Systems biology in drug development involves integration of diverse data types and connections between multiple levels of organization.

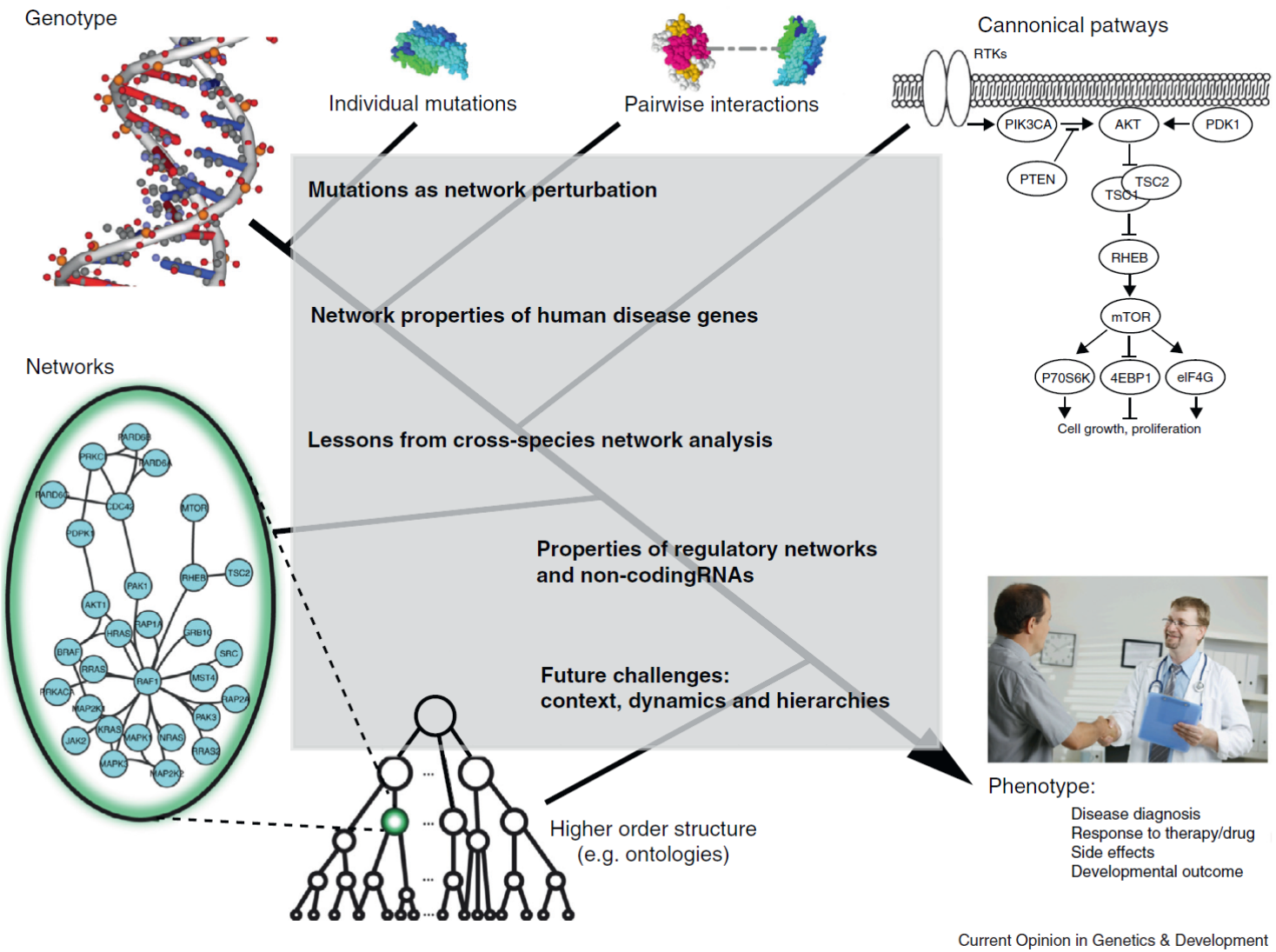
Novi pristopi za razvoj bioloških zdravil

Genotype to phenotype via network analysis☆

Hannah Carter[1], Matan Hofree[1,2] and Trey Ideker[1,2]

A prime objective of genomic medicine is the identification of disease-causing mutations and the mechanisms by which such events result in disease. As most disease phenotypes arise not from single genes and proteins but from a complex network of molecular interactions, *a priori* knowledge about the molecular network serves as a framework for biological inference and data mining. Here we review recent developments at the interface of biological networks and mutation analysis. We examine how mutations may be treated as a perturbation of the molecular interaction network and what insights may be gained from taking this perspective. We review work that aims to transform static networks into rich context-dependent networks and recent attempts to integrate non-coding RNAs into such analysis. Finally, we conclude with an overview of the many challenges and opportunities that lie ahead.

Biological networks are increasingly being applied to study the mechanisms by which genetic alterations cause phenotypic changes at the cellular level. Network organization and structure can help explain many disease phenomena such as locus heterogeneity, variable penetrance, pleiotropy, inheritance models and comorbidity. We believe these efforts are in their infancy. Limited knowledge of the dynamic and context-specific interplay of molecules within cell and our incomplete understanding of the makeup of the human genome has prevented effective modeling of the heritable contributions to human disease. Advances in experimental measurement technologies will soon enable large-scale screens to fill in much of our missing knowledge.



A hierarchical perspective of biological interactions mediating genotype–phenotype relationships. Protein activity is determined by protein amino acid sequence and structure. Proteins contribute to biological processes through interactions with other molecules in the cell. Biological processes arise from coordinated groups of molecular interactions, and in turn can interact to mediate higher order cellular behaviors and responses to environmental cues. Advances in several areas of network research are improving our understanding of how the organization of biological systems mediates genotype–phenotype relationships. This knowledge will be essential for identifying mutations underlying disease associations and their mechanisms of pathogenesis.

Network Systems Biology for Drug Discovery

Systems biology provides a platform for integrating multiple components and interactions underlying cell, organ, and organism processes in health and disease. Beyond traditional approaches focused on individual molecules or pathways, bioinformatic network analysis of high-throughput data sets offers an opportunity for integration of biological complexity and multilevel connectivity. Emerging applications in rational drug discovery range from targeting and modeling disease-corrupted networks to screening chemical or ligand libraries to identification/validation of drug–target interactions for improved efficacy and safety.

Network medicine is an emerging tool that is increasingly being applied to generate novel biological insights pertinent to human health and disease.[17,18,31,34] Network analysis enables informed systems interpretations by revealing particular nodes as effective targets for drug intervention or as unsuspected sites of off-target effects for existing drugs—or subsequent modeling efforts may prove that they contribute little to the mechanistic change observed for a particular drug perturbation despite exhibiting extreme modification within the system.[37] A network approach also enhances the understanding of disease pathobiology and systems pharmacology by serving as a template for downstream integration of multiple levels of complexity, for drug–protein connectivity, and for visualization of drug–ligand interactions.[38] It is critical to refer back constantly to the role of the target in the physiologic control system where it resides. This is particularly important in drug development, while trying to predict all the potential adverse effects of an intervention. It involves consideration of both on- and off-target effects of the main and subsidiary properties of the molecule. Although broad network applications are yet to be realized, the emergence of systems biology offers an innovation that is potentially capable of transforming the design and implementation of future drug discovery and drug development.

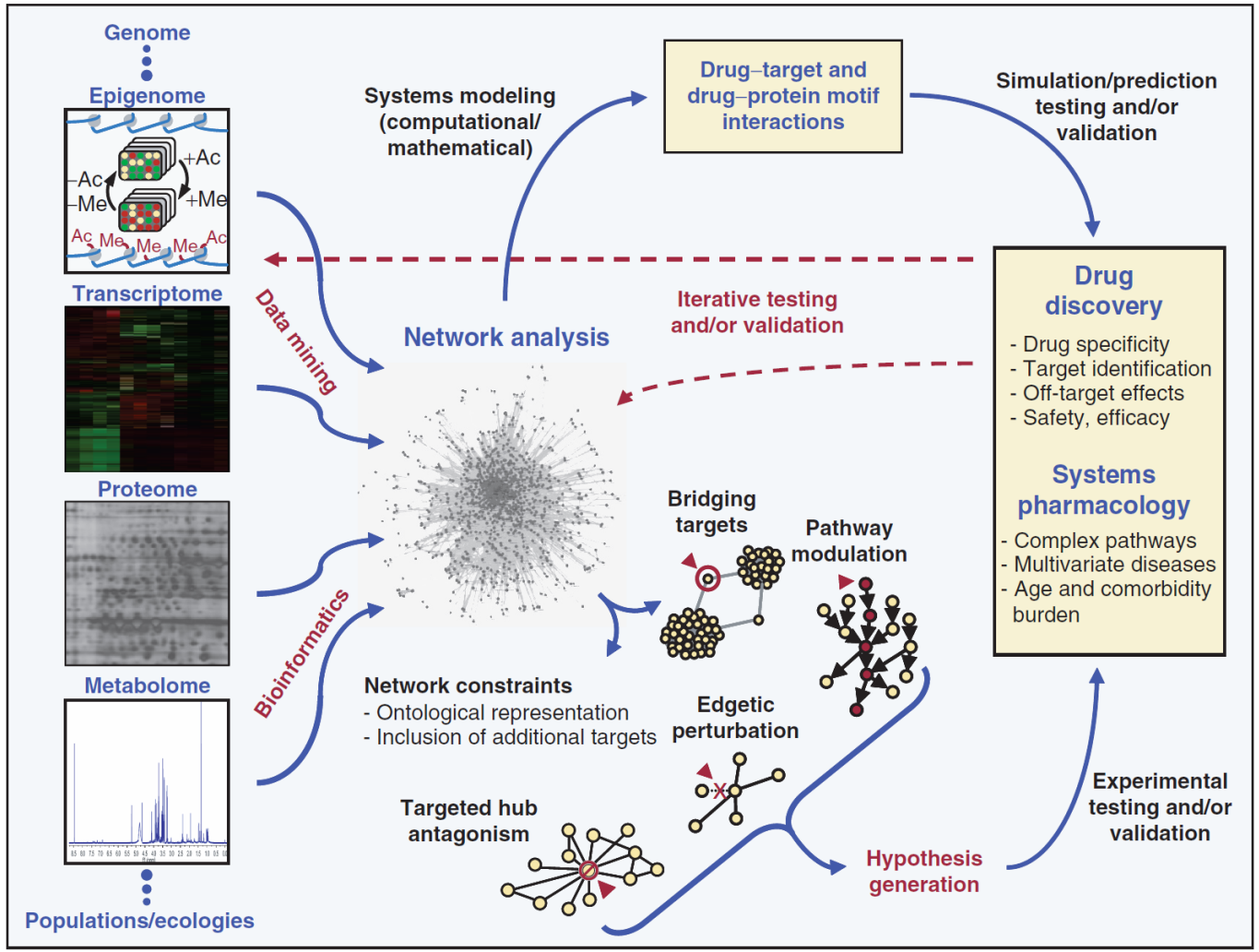
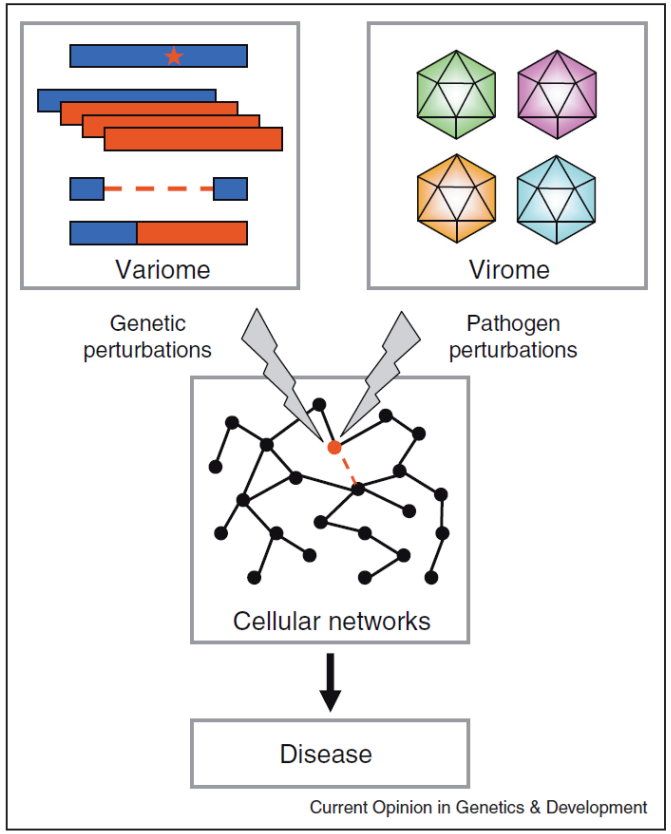


Figure 1 Network analysis serves an integral role in systems approaches to drug discovery. High-throughput data acquisition from multiple levels of biological complexity can be integrated by network analysis. Interaction networks provide a global template for computational and mathematical systems modeling, simulation, and prediction. Network topological parameters also provide foci and targets for hypothesis generation and experimental testing. Together, network-based approaches facilitate efforts in drug discovery and systems pharmacology, a process that can cycle iteratively for further data acquisition and/or network analysis. Specific applications of these various network-based systems approaches are outlined in the text. Ac, histone acetylation; Me, histone and/or DNA methylation.

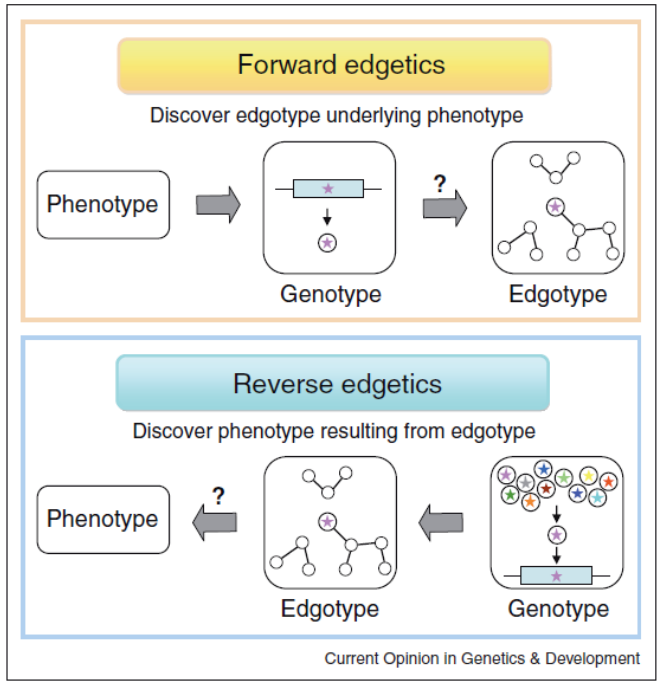
Edgotype: a fundamental link between genotype and phenotype

Nidhi Sahni[1,2], Song Yi[1,2], Quan Zhong[1,2], Noor Jailkhani[1,2], Benoit Charloteaux[1,2], Michael E Cusick[1,2] and Marc Vidal[1,2]

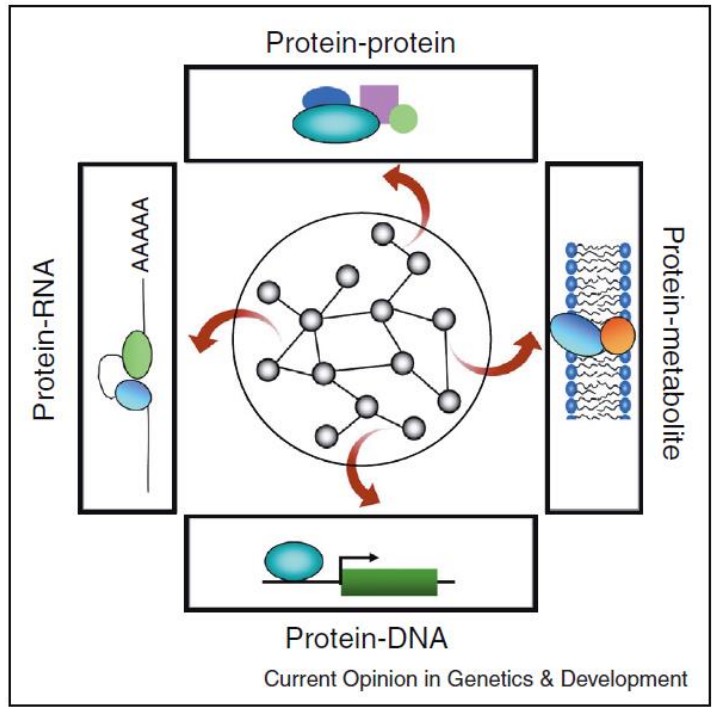
Classical 'one-gene/one-disease' models cannot fully reconcile with the increasingly appreciated prevalence of complicated genotype-to-phenotype associations in human disease. Genes and gene products function not in isolation but as components of intricate networks of macromolecules (DNA, RNA, or proteins) and metabolites linked through biochemical or physical interactions, represented in 'interactome' network models as 'nodes' and 'edges', respectively. Accordingly, mechanistic understanding of human disease will require understanding of how disease-causing mutations affect systems or interactome properties. The study of 'edgetics' uncovers specific loss or gain of interactions (edges) to interpret genotype-to-phenotype relationships. We review how distinct genetic variants, the genotype, lead to distinct phenotypic outcomes, the phenotype, through edgetic perturbations in interactome networks altogether representing the 'edgotype'.



Human genetic variations (the variome) and pathogenic viral proteins (the virome) similarly influence local and global properties of networks to induce disease states.



Forward and reverse edgetics to functionally characterize genomic variants. Forward edgetics studies the underlying edgotype for a given phenotype (disease), introducing known disease-causing mutations to study mutation-mediated loss or maintenance of known protein interactions and to relate the corresponding edgotype to a disease phenotype. Reverse edgetics introduces novel mutations into proteins of interest, finding those mutations that cause loss or maintenance of interactions against known interactors. The obtained mutations can then be introduced *in vivo* to characterize the resulting phenotype.

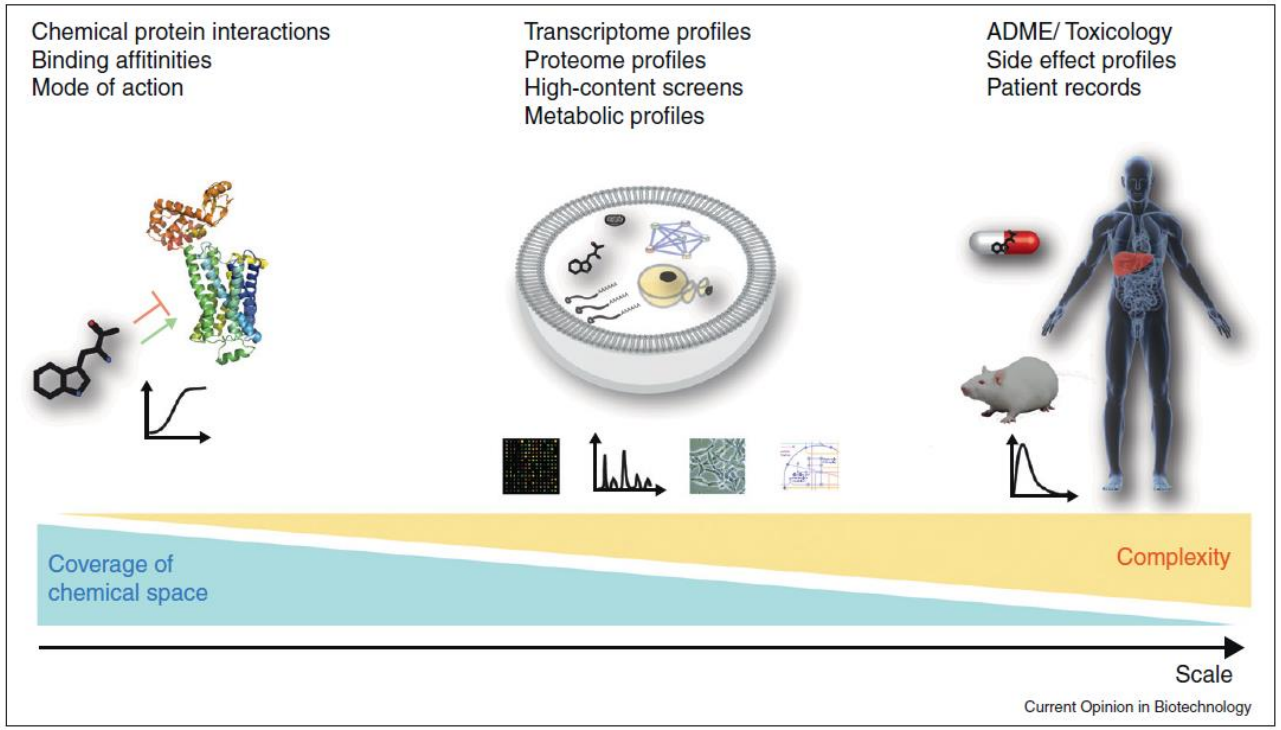


Edgetics can apply to all types of biomolecular interaction networks, including but not limited to protein–protein, protein–DNA, protein–RNA, and protein–metabolite interactions. Perturbations of these distinct types of interactions have been shown to be a factor in human disease.

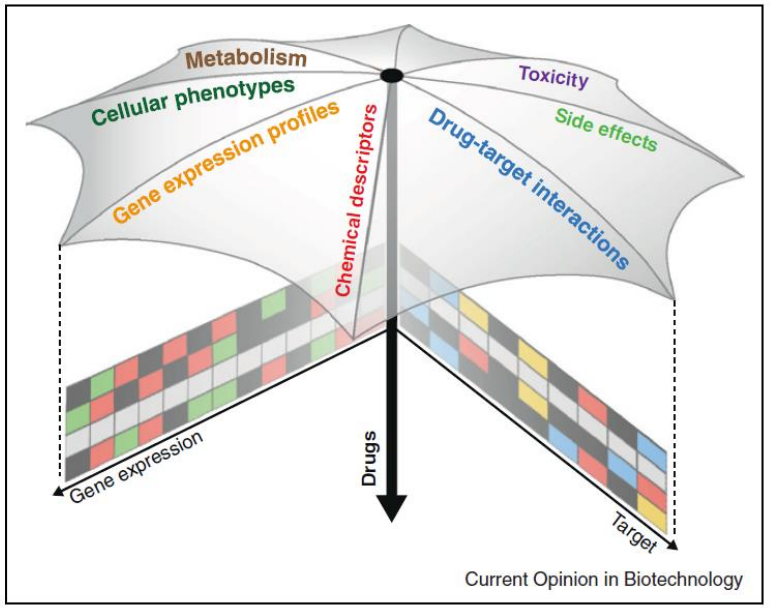
Drug discovery in the age of systems biology: the rise of computational approaches for data integration

Murat Iskar[1,a], Georg Zeller[1,a], Xing-Ming Zhao[1], Vera van Noort[1] and Peer Bork[1,2]

The increased availability of large-scale open-access resources on bioactivities of small molecules has a significant impact on pharmacology facilitated mainly by computational approaches that digest the vast amounts of data. We discuss here how computational data integration enables systemic views on a drug's action and allows to tackle complex problems such as the large-scale prediction of drug targets, drug repurposing, the molecular mechanisms, cellular responses or side effects. We particularly focus on computational methods that leverage various cell-based transcriptional, proteomic and phenotypic profiles of drug response in order to gain a systemic view of drug action at the molecular, cellular and whole-organism scale.



The different scales at which information on bioactivity of chemicals can be profiled. From *in vitro* binding assays to cellular profiling and whole-organism readouts, complexity of the data typically increases. Similarly, assay cost generally increases, whereas experimental throughput and hence availability of the data (coverage of chemical space) decreases along the same axis.



The promise of integrative methods for gaining a systems level understanding of complex biological responses to drug treatment. The growing diversity of publicly accessible bioactivity data has created many possibilities for integrative approaches. For example, gene expression profiles have been combined with drug-target information to infer mechanism of action and feedback regulation of target expression [13•,32,33••]. Integrating heterogeneous data from many bioactivity resources under one umbrella (i.e. on the basis of a common set of drugs) will be an important strategy to advance a systems level understanding of drug action.

For most pharmacologically relevant protein targets, experimental binding data are incomplete, even though high-throughput screens are more and more employed. This makes computational drug-target prediction attractive for exploring the interaction space between potential targets and bioactive chemicals [20]. Here we will focus on computational approaches that can make inferences for the majority of drug-target interactions. In this

Computational biology approaches for selecting host–pathogen drug targets

James R. Brown[1], Michal Magid-Slav[2], Philippe Sanseau[3] and Deepak K. Rajpal[4]

[1] Computational Biology, GlaxoSmithKline, 1250 South Collegeville Road, UP1345, PO Box 5089, Collegeville, PA 19426-0989, USA
[2] 709 Swedeland Road, UW2230, King of Prussia, PA 19406, USA
[3] Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK
[4] 5 Moore Drive, 3.2084, Research Triangle Park, NC 27709, USA

The proliferation of genomic platform data, ranging from silencing RNAs through mRNA and microRNA expression to proteomics, is providing new insights into the interplay between human and pathogen genes during infection: the so-called 'host–pathogen interactome'. Exploiting the interactome for novel human drug targets could provide new therapeutic avenues towards the treatment of infectious disease, which could ameliorate the growing clinical challenge of drug-resistant infections. Using the hepatitis C virus interactome as an example, here we suggest a computational biology framework for identifying and prioritizing potential human host targets against infectious diseases.

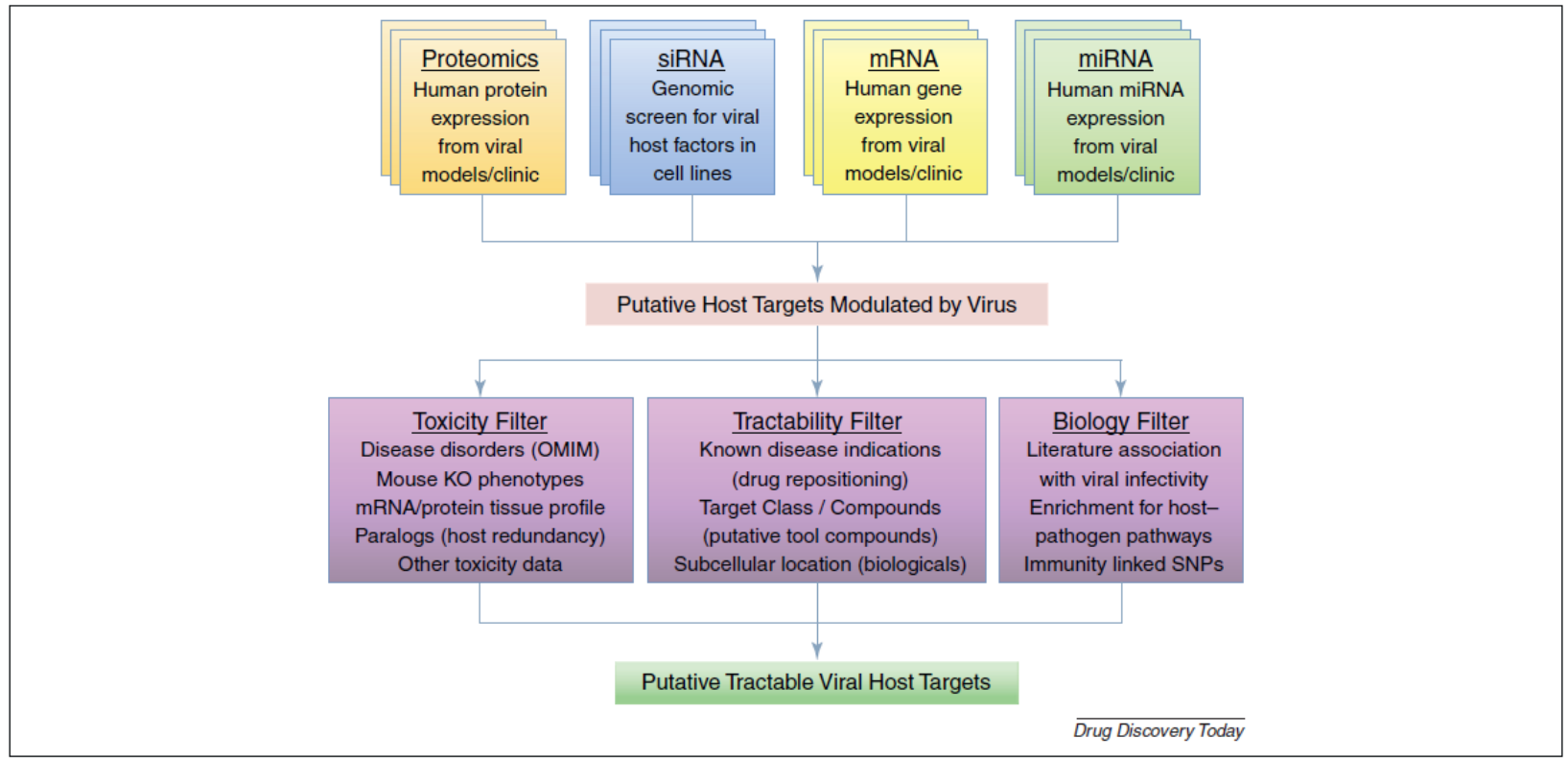


FIGURE 1

A pipeline for selecting putative tractable viral host targets. Conceptually, there are two components. The first is the integrative statistical analysis of multiple platform genomic data sets of the human or animal model host response to a common virus, such as HCV. The second phase involves further refining of putative host targets using quantitative and qualitative filters focused on toxicity, tractability and biological criteria. Specific steps are discussed in the main text.

The druggable genome

An assessment of the number of molecular targets that represent an opportunity for therapeutic intervention is crucial to the development of post-genomic research strategies within the pharmaceutical industry. Now that we know the size of the human genome, it is interesting to consider just how many molecular targets this opportunity represents. We start from the position that we understand the properties that are required for a good drug, and therefore must be able to understand what makes a good drug target.

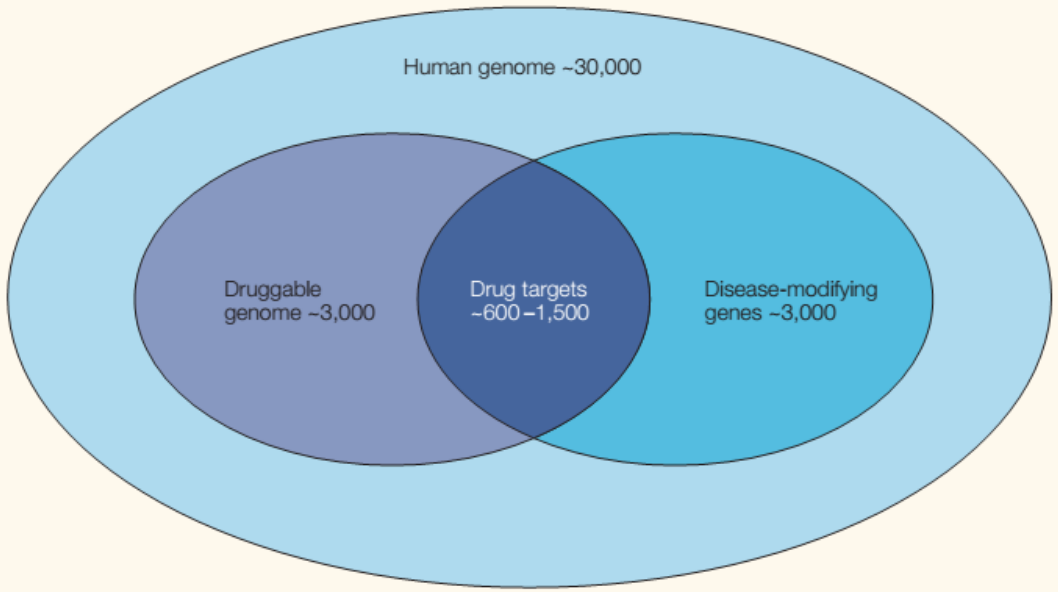


Figure 2 | **Number of drug targets.** The effective number of exploitable drug targets can be determined by the intersection of the number of genes linked to disease and the 'druggable' subset of the human genome.

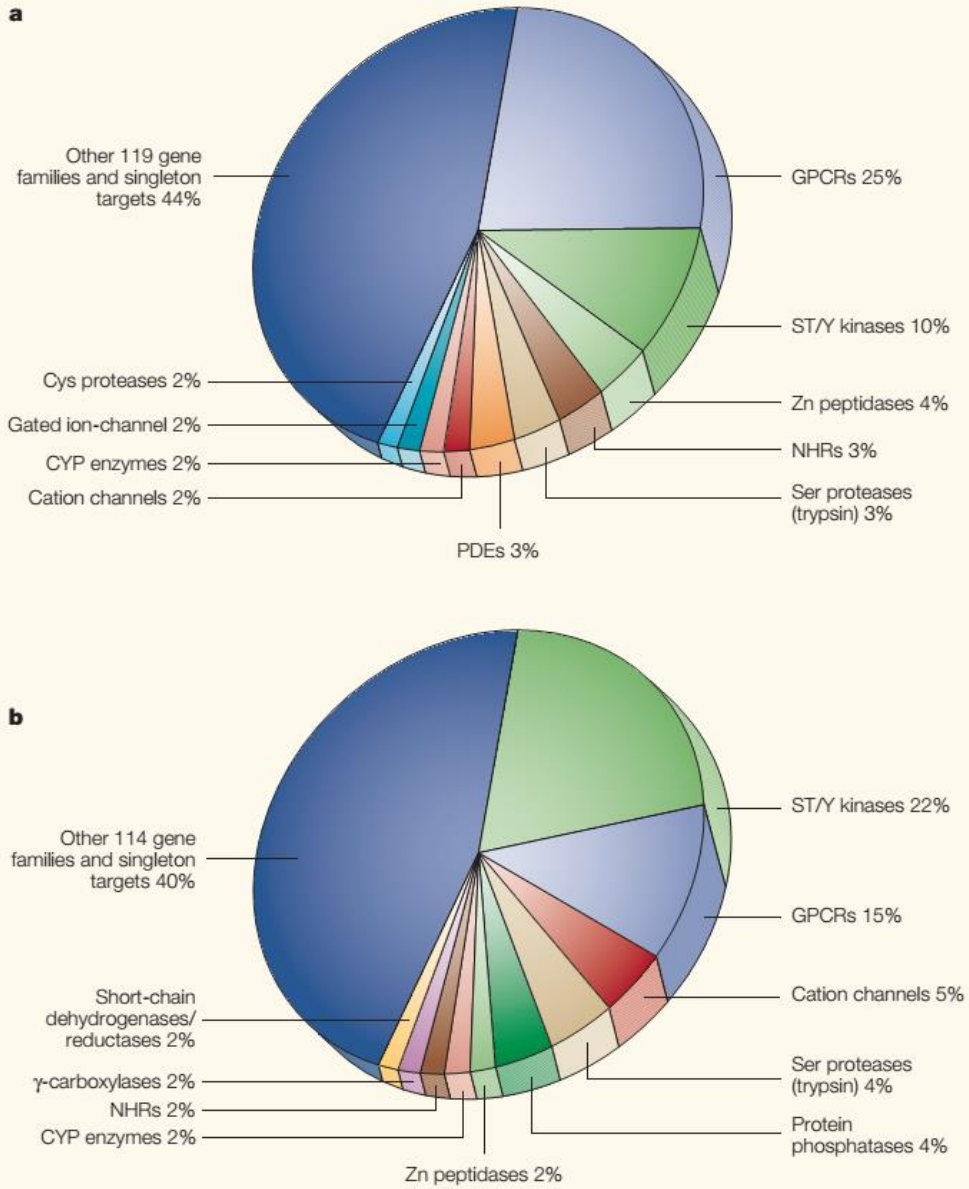
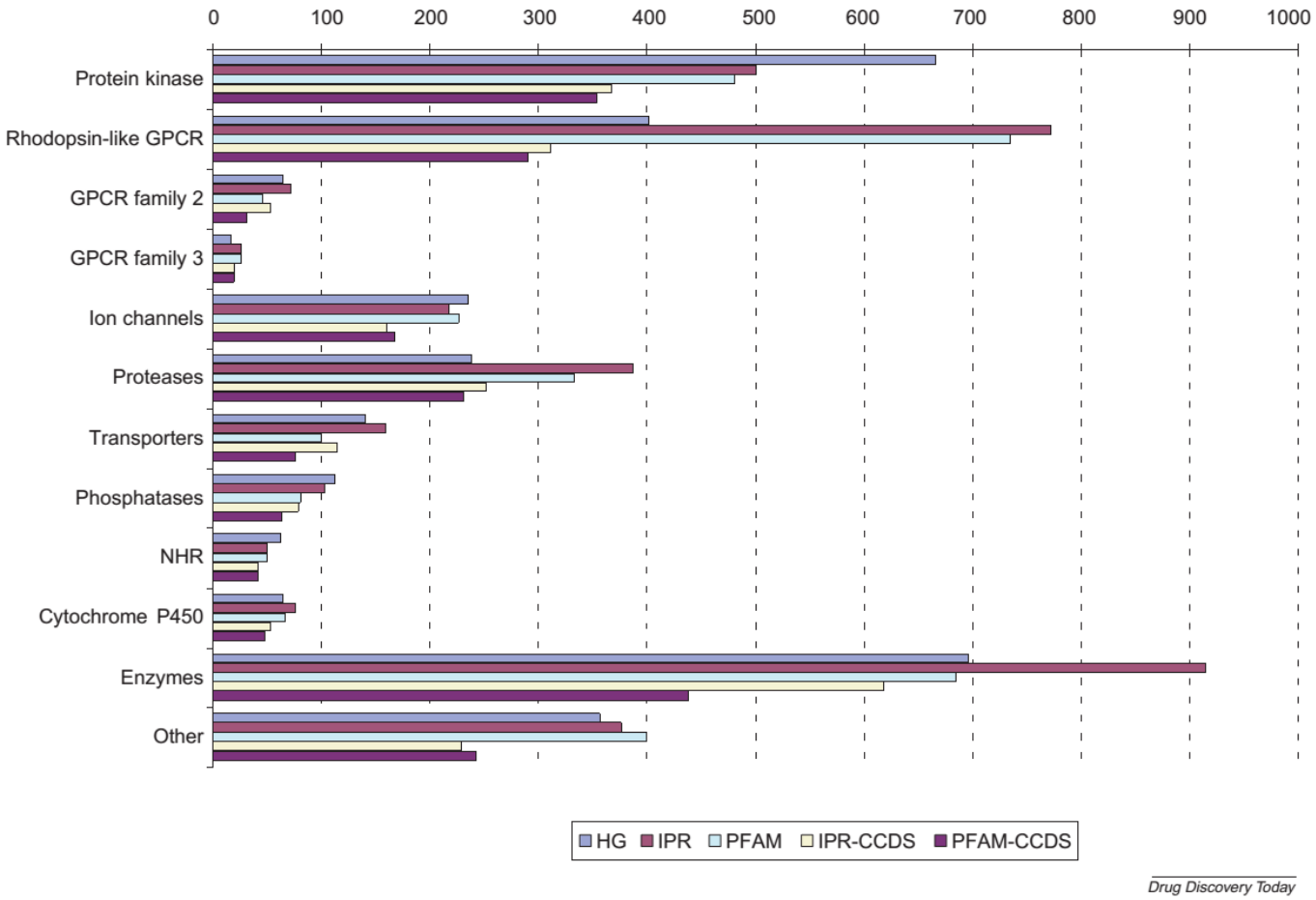
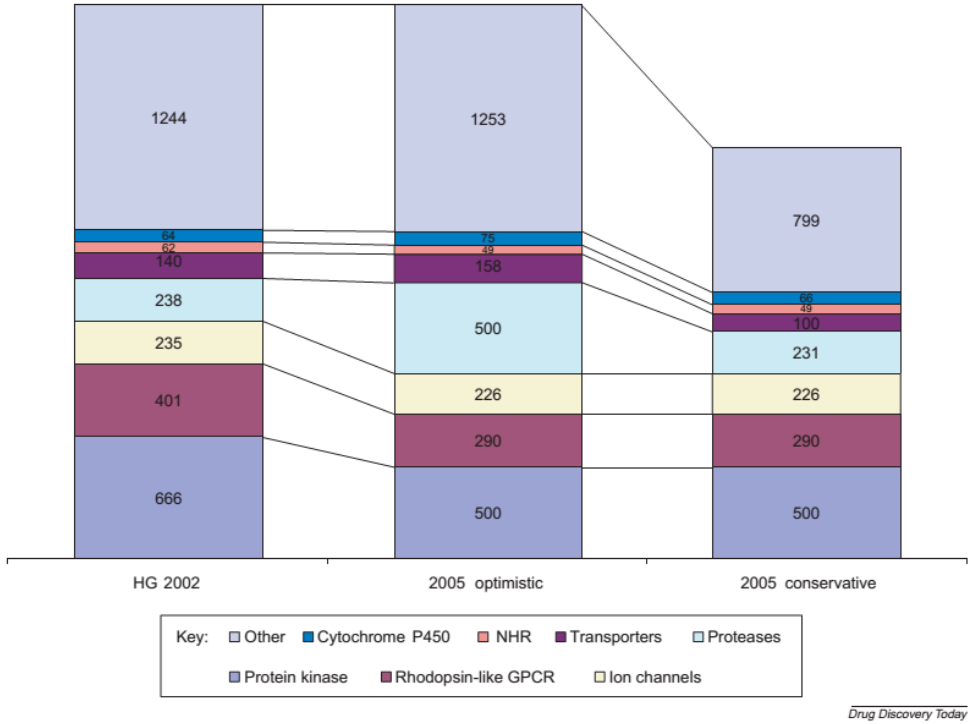


Figure 1 | **Drug-target families.** Gene-family distribution of **a** | the molecular targets of current rule-of-five-compliant experimental and marketed drugs, and **b** | the druggable genome. Serine (Ser)/threonine and tyrosine protein kinases are grouped as one gene family (ST/Y kinases), as are class 1 and class 2 G-protein-coupled receptors (GPCRs). CYP, cytochrome P450; Cys, cysteine; NHR, nuclear hormone receptor; PDE, phosphodiesterase; Zn, zinc.

The druggable genome: an update





Optimistic and conservative estimates of the druggable genome compared with previous predictions. NHR, nuclear hormone receptors.

The count of druggable protein classes based on Ensembl and CCDS annotations of the human genome sequence (build 35) compared with the data of Hopkins and Groom (HG) [2]. IPR represents the Interpro annotation of the full Ensembl gene set. PFAM represents the PFAM annotation of full Ensembl gene set. IPR-CCDS represents the Interpro annotation of full CCDS gene set. PFAM-CCDS represents the PFAM annotation of full CCDS gene set. The sum of druggable domains exceeds the count of druggable genes because some proteins contain more than one druggable domain. IPR and PFAM predictions of rhodopsin-like GPCR include ~400 sensory receptors. Abbreviation: NHR, nuclear hormone receptors.

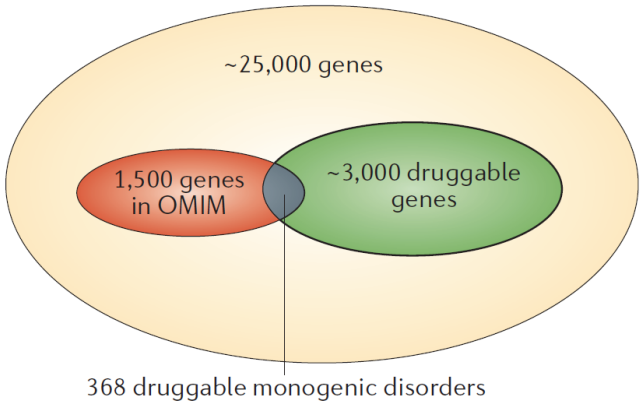
Box 1 | **The druggable human genome**

Recent reviews have suggested that there are approximately 3,000 classically 'druggable human genes' in the entire annotated genome[104,105]. These druggable genes were defined on the basis of classes of proteins that are responsive to the small-molecule drugs currently in use, and so have been validated for pharmaceutical development using established biochemical methods in the industry. The list of the druggable human genes includes well-characterized classes of targets, such as G-protein coupled receptors, voltage or ligand gated-ion channels, proteases, kinases and phosphatases. Not all of these genes are necessarily legitimate potential drug targets, as issues of specificity, side-effects and bioavailability could present insuperable obstacles in particular cases. Moreover there are already drugs on the market with targets that do not conform to these 'druggable' rules, such as humanized antibodies, protein therapeutics and peptidomimetics.

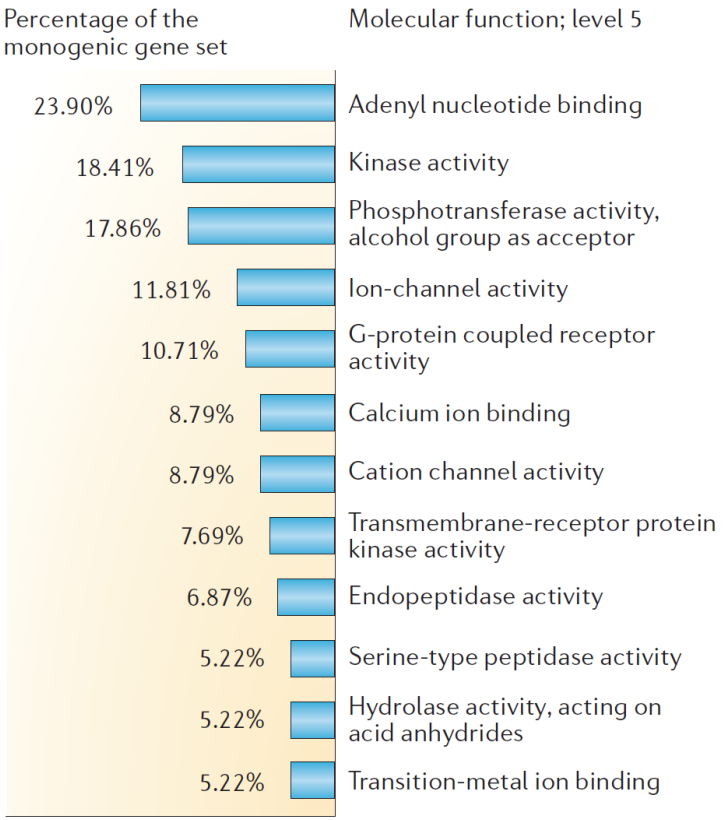
The intersection of the two gene sets — those with allelic variants that lead to monogenic phenotypes (FIG. 1) together with those that contain classically druggable functional domains — provides a group of 368 genes that we define as the druggable monogenic genome (panel **a**; see also Supplementary information S1–S3 (figure and tables)). To define and characterize these genes, the non-overlapping complete set of gene entries with curated allelic variants was extracted from the OMIM database and compared with the set used by Orth and co-workers of druggable genes in the human genome[105] to yield a preliminary set of 418 genes[105]. In this study, causal mutations were identified by manual inspection, reducing the set to 368 genes, and functional protein domains were documented using the Gene Ontology (GO) system. The count of these genes will grow as new gene annotations and new monogenic phenotypes continue to be identified.

The 368 genes were queried for GO annotations using Fatigo[106]. The monogenic druggable gene set includes well-understood chemical activities such as adenylate binding (24%), protein kinase (18%), ion channel (12% distributed between anionic and cationic, voltage-gated and ligand-gated ion channels), and G-protein coupled receptors (11%). Overall, 24% of identified genes in monogenic disorders have classically druggable motifs (368/1,500 = 24%); interestingly, this fraction is greater than the fraction of all annotated genes that contain such motifs (2,935/25,000 = 10.5%). Panel **b** shows these results for GO level-5 domains that occur in at least 5% of genes; for a complete listing see Supplementary information S1 (figure). Other GO levels can be queried using the gene list in Supplementary information S3 (table). The percentages for the receptors, kinases and proteases that are quoted in the text are with respect to the 368 genes. These percentages are non-exclusive as the same gene product can be annotated with multiple GO functions.

a



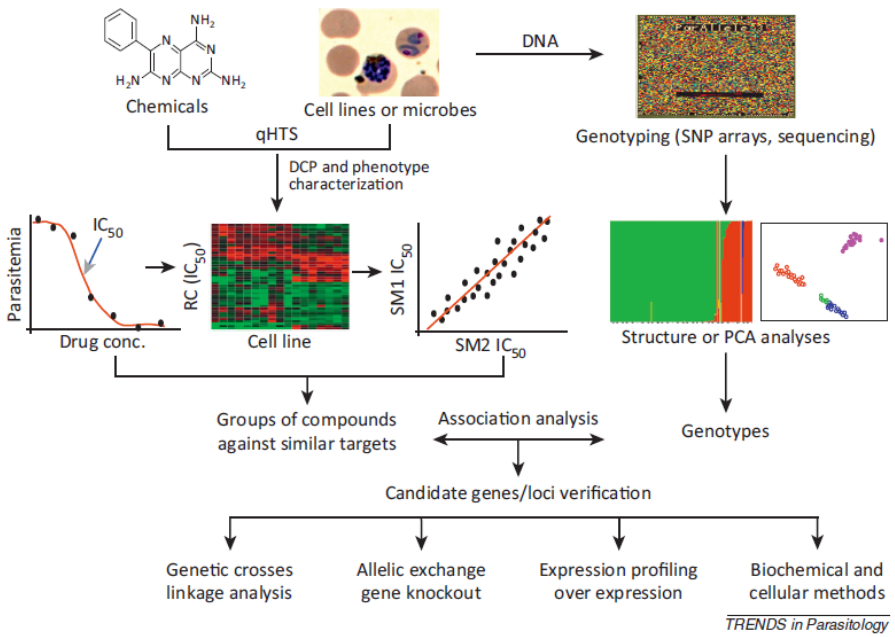
b



Chemical genomics for studying parasite gene function and interaction

Jian Li[1], Jing Yuan[1], Ken Chih-Chien Cheng[2], James Inglese[2,3], and Xin-zhuan Su[4]

With the development of new technologies in genome sequencing, gene expression profiling, genotyping, and high-throughput screening of chemical compound libraries, small molecules are playing increasingly important roles in studying gene expression regulation, gene–gene interaction, and gene function. Here we briefly review and discuss some recent advancements in drug target identification and phenotype characterization using combinations of high-throughput screening of small-molecule libraries and various genome-wide methods such as whole-genome sequencing, genome-wide association studies (GWAS), and genome-wide expression analysis. These approaches can be used to search for new drugs against parasite infections, to identify drug targets or drug resistance genes, and to infer gene function.



A chemical genomics approach to identify targets of small molecules. Cell lines or microorganisms are screened against chemical libraries for differences in their responses to the chemicals (differential chemical phenotypes, DCPs). The compounds can be clustered based on the similarity of response patterns (or positive/negative correlation between pairs of compounds) from a panel of cell lines or isolates. Similar response patterns among compounds may suggest that they act on the same or similar targets. DNA samples from the cell lines or isolates are typed using high-throughput genotyping methods such as microarrays or parallel sequencing. Genotypes obtained are analyzed for potential population structures before association analysis. Candidate loci or genes can be further characterized using linkage analysis if progeny from genetic crosses are available. The candidate genes can also be genetically modified or overexpressed to evaluate phenotypic changes. Gene functions can be deduced after additional biochemical and cellular characterizations. Abbreviations: PCA, principal component analysis; qHTS, quantitative high-throughput screening; RC, response to chemicals; SM1 and SM2, small molecules 1 and 2; SNP, single-nucleotide polymorphism.

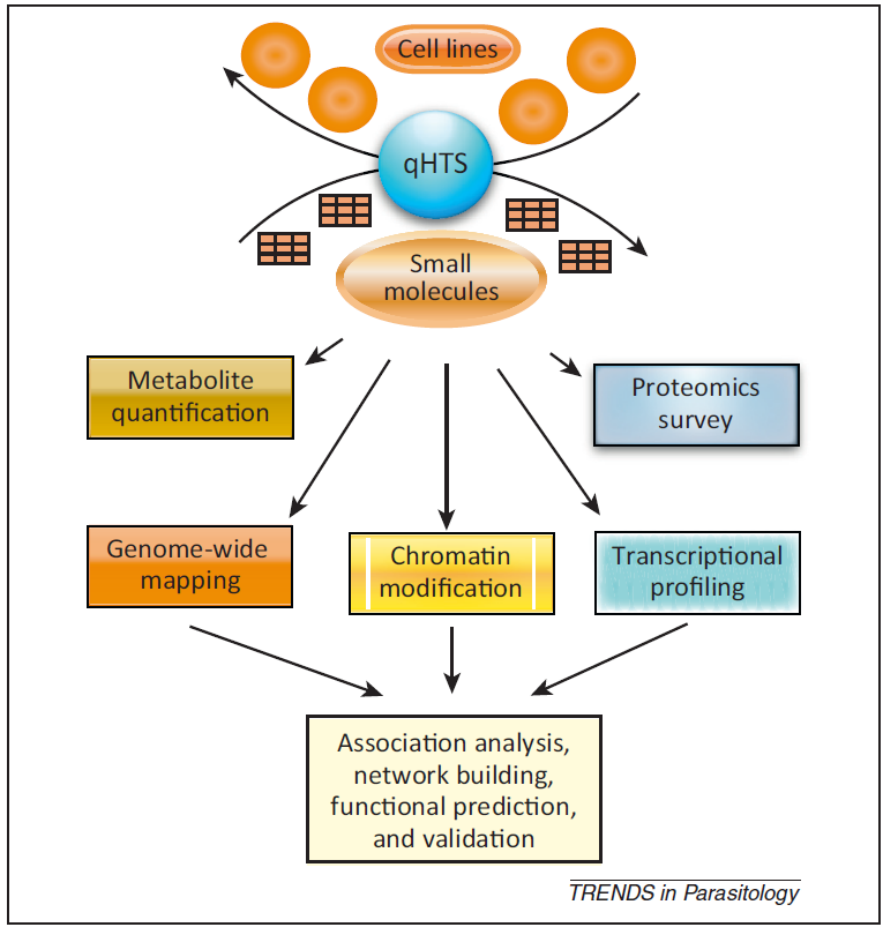


Figure 1. Strategies for identifying chemical targets and developing gene interaction networks. Cell lines or individual isolates of microbes can be screened or treated with small molecules, and variations in their responses can be studied using genome-wide approaches. Association analysis and construction of gene interaction networks will provide important information for predicting gene function and gene interaction. Abbreviation: qHTS, quantitative high-throughput screening.

The application of next-generation sequencing technologies to drug discovery and development

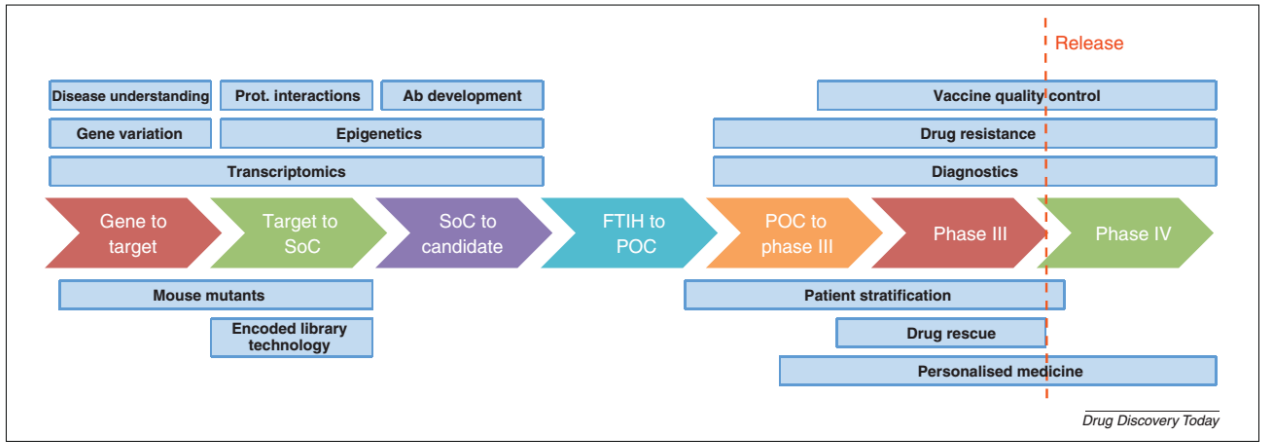


FIGURE 2

Summary of NGS applications in drug discovery.

Next-generation sequencing (NGS) technologies represent a paradigm shift in sequencing capability. The technology has already been extensively applied to biological research, resulting in significant and remarkable insights into the molecular biology of cells. In this review, we focus on current and potential applications of the technology as applied to the drug discovery and development process. Early applications have focused on the oncology and infectious disease therapeutic areas, with emerging use in biopharmaceutical development and vaccine production in evidence. Although this technology has great potential, significant challenges remain, particularly around the storage, transfer and analysis of the substantial data sets generated.

TABLE 2

Broad applications of NGS to drug discovery

Applications	Pros of NGS	Cons of NGS	Alternatives	Refs
Mutation detection: personalised medicine	Can sequence large genome regions to identify efficacy markers	Initial setup and running cost for NGS	Large-scale Sanger sequencing technology	[64]
ChIP-Seq: target identification and/or validation and compound profiling for epigenetics	Enables study of epigenetic targets at the whole-genome level	Many possible algorithms for data analysis and complex data interpretation	ChIP-on-chip assay using microarray-based technology	[44]
CNV: target identification, personalised medicine, for example, cancer	Uncovers all types of CNV; no a priori assumptions about location of CNVs required	Large and complex rearrangements might not be detected	Comparative genomic hybridisation	[35,65, 66]
Exome sequencing: target identification and/or drug resistance studies, biomarker discovery	Identify rare variants, using deep sequence coverage	Sequence variation in non-coding regions and introns not detected	Large-scale Sanger sequencing technology	[16]
RNA-Seq: target identification and/or validation by studying differential gene or miRNA expression between normal and diseased tissue	Detects alternative splicing and low expression transcripts; has large dynamic range	Bias during library preparation can result in over-representation of transcript 3′ ends	Microarray-based technology	[11,12, 67]
HITS-CLIP: study of RNA–protein interactions	Enables study of RNA–protein interactions	A relatively new application; not many studies to date	Microarray-based technology	[68]
Ribosome profiling: target identification by measuring protein translation rates using sequencing to identifying ribosomal footprints	Potential to enable analysis of the whole cell proteome by sequencing	A relatively new application, not many studies to date	Conventional proteomics technologies, for example, mass spectrometry	[69]
Small RNA sequencing (e.g. miRNA): biomarker discovery	Assay and quantify all small RNAs present	Data analysis complex owing to the presence of isomiRs	Fluorescently labelled PCR techniques	[22,40]
Bacterial genome sequencing: target identification, validation and diagnostics to identify new strains and mechanisms of drug resistance	Small genomes mean many strains can be sequenced per run	Short read alignment can result in gaps in coverage owing to repeat sequence	Large-scale Sanger sequencing technology	[45,47, 48,70]

The impact of epigenomics on future drug design and new therapies

Christopher A. Hamm and Fabricio F. Costa

The future of drug design and the development of new therapeutics will rely on our ability to unravel the complexities of the epigenome in normal and disease states. Proper epigenetic regulation is essential for normal differentiation in embryogenesis and development. Conversely, abnormal epigenetic regulation is a feature of complex diseases, including cancer, diabetes, heart disease and other pathologies. Epigenetic therapies hold promise for a wide range of biological applications, from cancer treatment to the establishment of induced pluripotent stem cells. The creation of more specific and effective epigenetic therapies, however, requires a more complete understanding of epigenomic landscapes. Here, we give a historical overview of the epigenomics field and how epigenetic modifications can affect embryo development and disease etiology. We also discuss the impact of current and future epigenetic drugs.

TABLE 1 (*Continued*)

Diseases and disorders	Relationship between epigenetic/ epigenomic alterations and diseases	Current and future areas of therapeutic intervention	Refs
	DNA methyltransferase 3A is upregulated in the eutopic endometrium of women with endometriosis		
Inherited glycosylphosphatidylinositol (GPI) deficiency	Mutation (C–G substitution) in the GPI mannosyltransferase 1 gene (*PIGM*) results in inherited GPI deficiency. The mutation causes hypoacetylation of the PIGM promoter	The HDAC inhibitor, sodium butyrate, increased *PIGM* expression and alleviated disease-associated symptoms of GPI deficiency in a patient	[116]
Schizophrenia and bipolar disorder	Several loci are epigenetically different in the brains of patients with either schizophrenia or bipolar disorder compared with unaffected controls SNPs in the *HDAC3* and *HDAC4* genes are associated with schizophrenia and could be involved in the pathophysiology of the disease	ND	[117,118]
Additional psychiatric disorders	See [119] for a detailed description of epigenetic and/or epigenomic alterations in Fragile X syndrome, Coffin–Lowry syndrome, Rett syndrome, myotonic dystrophy, Prader–Willi syndrome and Angelman syndrome		[119]

[a] Diseases that are classified as the top ten worldwide causes of death according to the World Health Organization [56].
[b] ND: topic not discussed in the report referenced.

TABLE 1
Epigenetic and epigenomic associations with complex disease

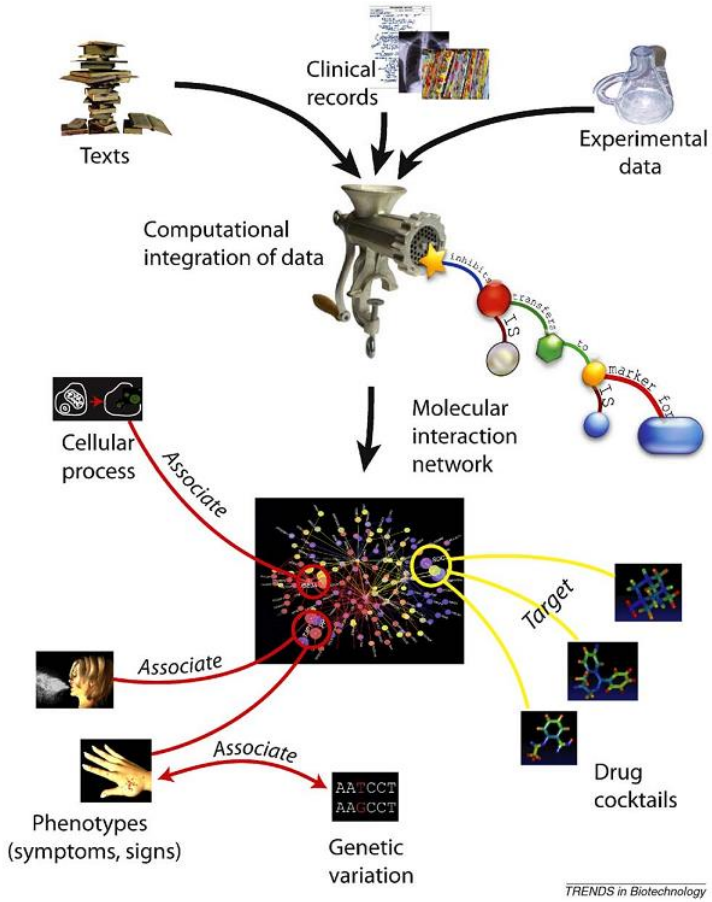
Diseases and disorders	Relationship between epigenetic/ epigenomic alterations and diseases	Current and future areas of therapeutic intervention	Refs
Alzheimer's disease[a]	Changes in DNA methylation and histone modifications were reported for Alzheimer's disease. For a summary of alterations, see [103]	The association between epigenetic abnormalities and Alzheimer's disease suggests that drugs targeting epigenetic pathways have therapeutic value Animal model studies indicate that HDAC inhibitors might be useful for the treatment of Alzheimer's disease	[103]
Asthma and COPD[a]	Chromatin remodeling affects the expression of inflammation-related genes in COPD and asthma HDAC2 mediates the repression of inflammatory genes; HDAC2 levels are reduced in alveolar macrophages of patients with COPD and asthma	Theophylline can restore HDAC2 activity in COPD macrophages HDAC inhibitors could be used to decrease inflammatory gene expression in COPD and asthma	[67,104,105]
Atherosclerosis	Hypermethylation of specific genes was widely reported; however, hypomethylation has also been detected in advanced human atherosclerotic plaques	Drugs targeting epigenetic changes could provide new avenues of treatment for cardiovascular disease	[106]
Cancer (several types)[a]	See references for a general overview of this complex disease	HDAC inhibitors, DNA methyltransferase inhibitors, small molecules targeting chromatin-associated proteins, and others	[49,102,107]
Coronary artery disease[a]	Global DNA methylation levels significantly higher in patients with coronary artery disease compared with normal controls	ND[b]	[108]
Depression	Early-life stress in rodents can lead to DNA hypomethylation and expression of arginine vasopressin (AVP), which is associated with features of depression Levels of DNA methyltransferase-2 are elevated in a rodent model of depression	Behavioral phenotypes induced by the early-life stress could be reversed by treatment with a AVP V1b receptor antagonist (SSR149415) Imipramine (antidepressant) reversed the downregulation of brain-derived neurotrophic factors and increased histone acetylation at the promoters of these genes	[71,109,110]
	Brain-derived neurotrophic factors III and IV are downregulated in a rodent model of depression, and this downregulation is correlated with increased repressive histone methylation at their corresponding promoters		
Diabetes mellitus (Type 1)[a]	Association of type 1 diabetes mellitus and the imprinted DLK1-MEG3 locus on chromosome 14q32 DNA methylation affects the expression of candidate type 1 diabetes causal genes: insulin precursor (INS), interleukin 2 and 10	ND	[111,112]
Diabetes mellitus (Type 2)[a]	The following epigenetic changes have been documented during the development of type 2 diabetes mellitus in rats: Pdx-1 DNA methylation and histone deacetylation; Glut4 histone deacetylation	Therapies targeting epigenetic gene regulation could be used to treat type 2 diabetes mellitus	[74,75,113]
		Glucagon-like peptide 1, glucose-dependent insulinotropic-peptide 1, and pioglitazone (PPAR-γ agonist) are used to treat patients with type 2 diabetes mellitus and they have been shown to reverse epigenetic modifications *in vitro*	
Endometriosis and implantation failure (pregnancy)	The *HOXA10* gene is hypermethylated in the endometrium of women with endometriosis compared with controls. Silencing of *HOXA10*, partially due to hypermethylation, could lead to impaired embryo implantation in endometriosis	The reversion of aberrant methylation through pharmacological means could be a novel form of treatment for endometriosis and implantation failure	[57,114,115]

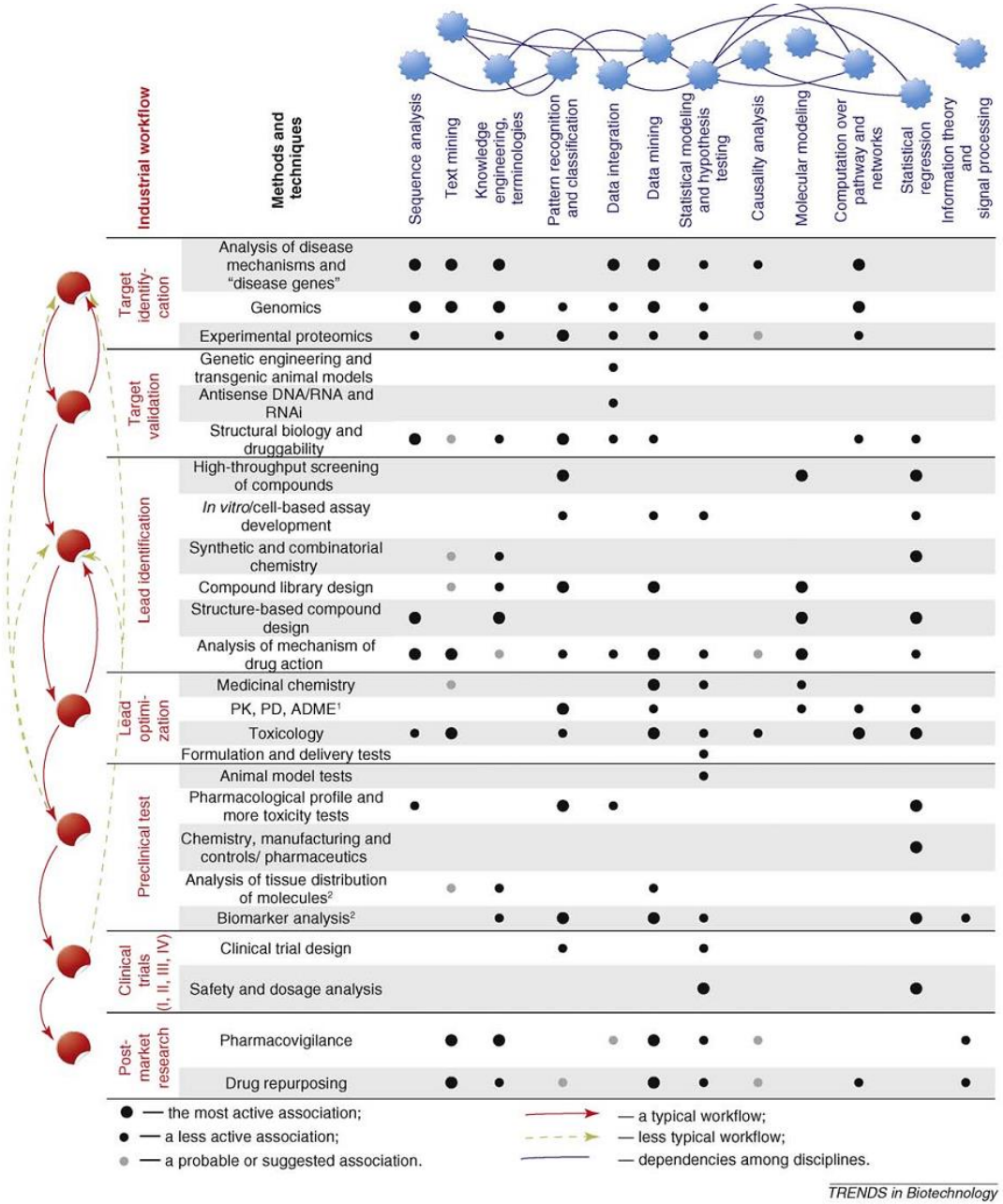
Novel opportunities for computational biology and sociology in drug discovery☆

Corrected paper

Lixia Yao[1], James A. Evans[2,3] and Andrey Rzhetsky[3,4]

Current drug discovery is impossible without sophisticated modeling and computation. In this review we outline previous advances in computational biology and, by tracing the steps involved in pharmaceutical development, explore a range of novel, high-value opportunities for computational innovation in modeling the biological process of disease and the social process of drug discovery. These opportunities include text mining for new drug leads, modeling molecular pathways and predicting the efficacy of drug cocktails, analyzing genetic overlap between diseases and predicting alternative drug use. Computation can also be used to model research teams and innovative regions and to estimate the value of academy–industry links for scientific and human benefit. Attention to these opportunities could promise punctuated advance and will complement the well-established computational work on which drug discovery currently relies.

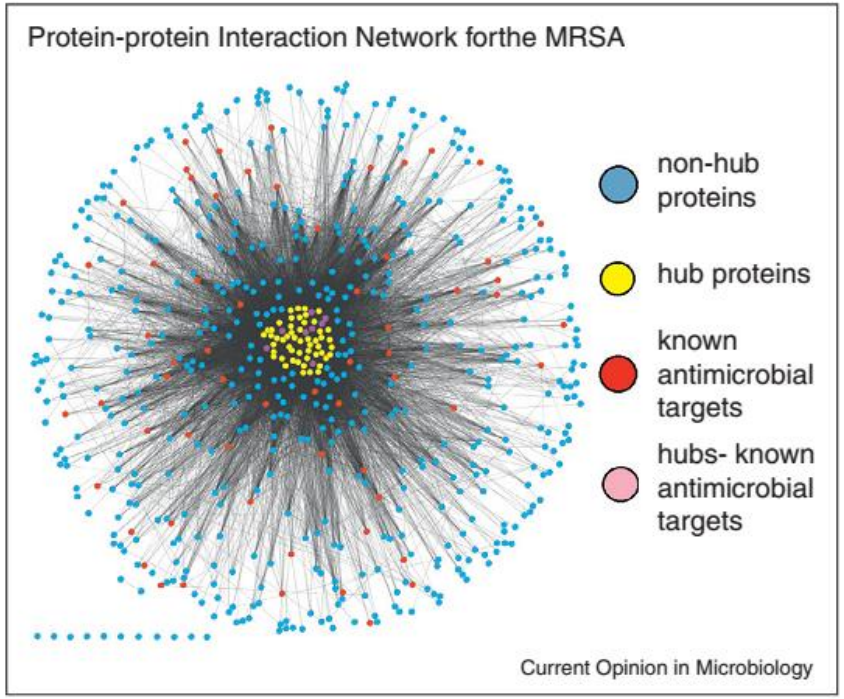




Protein interaction networks as starting points to identify novel antimicrobial drug targets

Roya Zoraghi[1] and Neil E Reiner[1,2]

Novel classes of antimicrobials are needed to address the challenge of multidrug-resistant bacteria. Current bacterial drug targets mainly consist of specific proteins or subsets of proteins without regard for either how these targets are integrated in cellular networks or how they may interact with host proteins. However, proteins rarely act in isolation, and the majority of biological processes are dependent on interactions with other proteins. Consequently, protein–protein interaction (PPI) networks offer a realm of unexplored potential for next-generation drug targets. In this review, we argue that the architecture of bacterial or host–pathogen protein interactomes can provide invaluable insights for the identification of novel antibacterial drug targets.



2D representation of the developed MRSA partial PIN. Hub proteins are shown in yellow and non-hub in blue. Established antimicrobial drug targets are shown in red if they were classified as non-hubs and in purple if they were categorized as hubs
Reprinted from Cherkasov A, *et al.*: **Mapping the protein interaction network in methicillin-resistant *Staphylococcus aureus***. *J Proteome Res* 2011, 10:1139–1135. Used with permission.

Conclusions and future directions

Both bacterial and host–pathogen protein complexes represent a reliable source of potential targets for novel classes of antimicrobials. However, many more highly validated interaction datasets are required to evaluate the biological significance of individual interactions. In particular it is imperative to distinguish between conserved and non-conserved (but biologically relevant) hubs and separate them from false-positives and false-negatives by combining different PPI technologies with complementary experimental approaches and computational analyses. Currently, comparative studies suggest that the development of accurate and complete repertoires of bacterial PPIs is still in its infancy, but given the progress that has been made and the importance of this targeting strategy, it is likely to receive increased attention in the future. Moreover, to reduce the likelihood of resistance development, targeting conserved hubs in bacterial or pathogen–host interactomes is consistent with the recent trends in antibacterial drug discovery favoring potential polypharmacology (i.e. a single drug acting upon multiple targets of a unique pathway or a single drug acting upon multiple targets pertaining to multiple pathways), over single target drugs [12•,82].

Živalski modeli za študij bolezni pri človeku

Life in the Fast Lane: Mammalian Disease Models in the Genomics Era

Lukas E. Dow[1] and Scott W. Lowe[1,2,*]
[1]Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
[2]Howard Hughes Medical Institute, New York, NY 10065, USA
*Correspondence: lowes@mskcc.org
DOI 10.1016/j.cell.2012.02.023

Analyses of the human genome have proven extremely successful in identifying changes that contribute to human disease. Genetically engineered mice provide a powerful tool to analyze these changes, although they are slow and costly and do not always recapitulate human biology. Recent advances in genomic technologies, rodent-modeling approaches, and the production of patient-derived reprogrammed cell lines now provide a plethora of complementary systems to study disease states and test new therapies. Continued evolution and integration of these model systems will be the key to realizing the benefits of the genomic revolution and refining our understanding and treatment of human diseases.

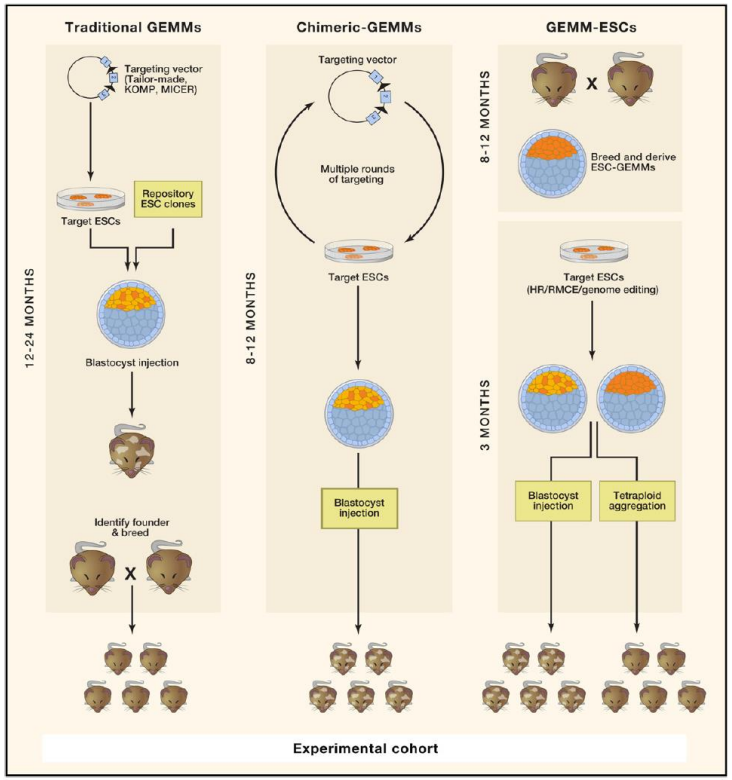


Figure 2. Traditional and "Speedy" Mouse Model Development
The development of traditional knockout and conditional knockout mice requires the generation of unique targeting vectors, low-efficiency homologous recombination, identification of founders that transmit the targeted allele through the germline, and subsequent breeding and/or backcrossing. ESC-based GEMMs make use of multiallelic ESCs developed through multiple rounds of ESC targeting (chimeric-GEMMs) or rederivation of ESCs carrying established disease alleles (speedy or GEMM-ESCs). The initial establishment phase of GEMM-ESC models demands breeding but, once created, requires little manipulation in vitro. Experimental cohorts can be produced directly from the multiallele ESC and do not require any breeding steps. Experimental mice can be generated by blastocyst injection, by producing chimeric mice (depicted by brown and white coat color), or through tetraploid embryo complementation, generating wholly ESC-derived animals (brown coat).

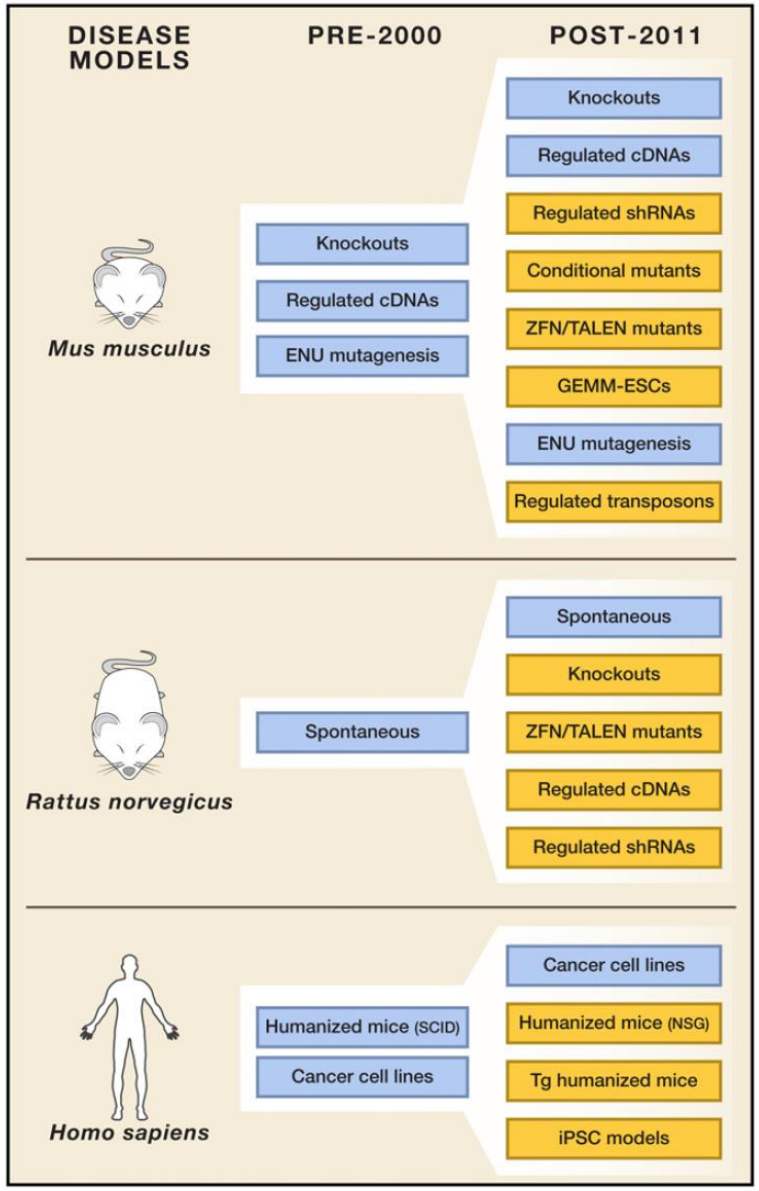


Figure 1. New Tools in Mammalian Disease Models
Platforms for the genetic manipulation of mammalian model systems (primarily mouse, rat, and human) have developed at an incredible rate in the past 10–15 years. Recent sophisticated adaptations in each of these systems (indicated in yellow) now provide an extremely flexible array of tools for investigating and understanding the underlying causes of human disease.

Animal models of human disease: zebrafish swim into view

Graham J. Lieschke and Peter D. Currie‡*

Abstract | Despite the pre-eminence of the mouse in modelling human disease, several aspects of murine biology limit its routine use in large-scale genetic and therapeutic screening. Many researchers who are interested in an embryologically and genetically tractable disease model have now turned to zebrafish. Zebrafish biology allows ready access to all developmental stages, and the optical clarity of embryos and larvae allow real-time imaging of developing pathologies. Sophisticated mutagenesis and screening strategies on a large scale, and with an economy that is not possible in other vertebrate systems, have generated zebrafish models of a wide variety of human diseases. This Review surveys the achievements and potential of zebrafish for modelling human diseases and for drug discovery and development.

Generating zebrafish disease models

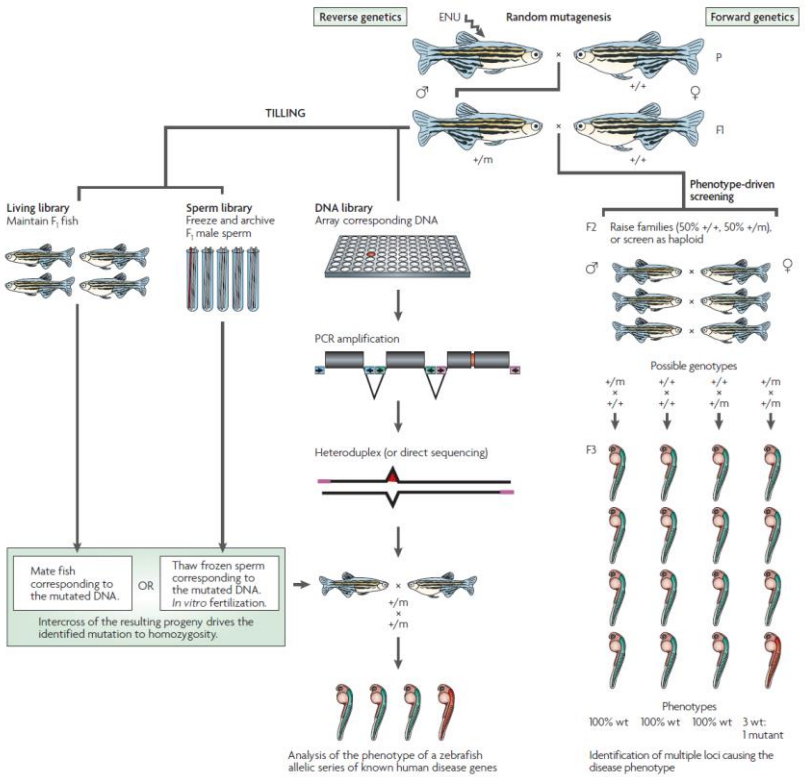


Table 1 | **Attributes of some key animals used to model human disease**

Attribute of disease model	Model organism			
	Fly	*Zebrafish*	*Mouse*	*Rat*
Practical issues				
Husbandry infrastructure	$	$	$$$	$$$
Cost per animal per year	$	$	$$$	$$$
Characterized inbred strains	+	-	++++	+++
Outbred laboratory strains	+	+++	++	++
Anatomical similarity	-	+	++	++
Molecular or genetic similarity	+	++	+++	+++
Pathological similarity	-	++	+++	+++
Storage; for example, freezing sperm	No	Yes	Yes	Yes
Molecular biology tools				
Transgenesis*	++	++	++	++
Targeted gene modification*	+	-	++++	+
Transient *in vivo* assays*	++	++++	+	+
Allelic series from TILLING*	+++	++++	++	+
Feasibility of large-scale screens‡	++++	+++	++	+
Affordability of large-scale screens‡	++++	+++	+	-
Sequencing progress§	+++	++	+++	++
Annotation progress§	++	++	++++	++
Cell-biology tools				
Cell lines and tissue culture	++	+	++++	+
Antibody reagents	++	+	++++	++

*Reverse-genetics approach; ‡forward-genetics approach; §genome sequence; -, not relevant, or not a strength; $, $$, $$$ and +, ++, +++, relative cost ($) and strength (+) of the model in each category; ++++, outstanding strength of the model; TILLING, targeting induced local lesions in genomes.

Od genoma do protivirusnih zdravil

From genome to antivirals: SARS as a test tube

Yossef Kliger, Erez Y. Levanon and Doron Gerber

The severe acute respiratory syndrome (SARS) epidemic brought into the spotlight the need for rapid development of effective anti-viral drugs against newly emerging viruses. Researchers have leveraged the 20-year battle against AIDS into a variety of possible treatments for SARS. Most prominently, based solely on viral genome information, silencers of viral genes, viral-enzyme blockers and viral-entry inhibitors were suggested as potential therapeutic agents for SARS. In particular, inhibitors of viral entry, comprising therapeutic peptides, were based on the recently launched anti-HIV drug enfuvirtide. This could represent one of the most direct routes from genome sequencing to the discovery of antiviral drugs.

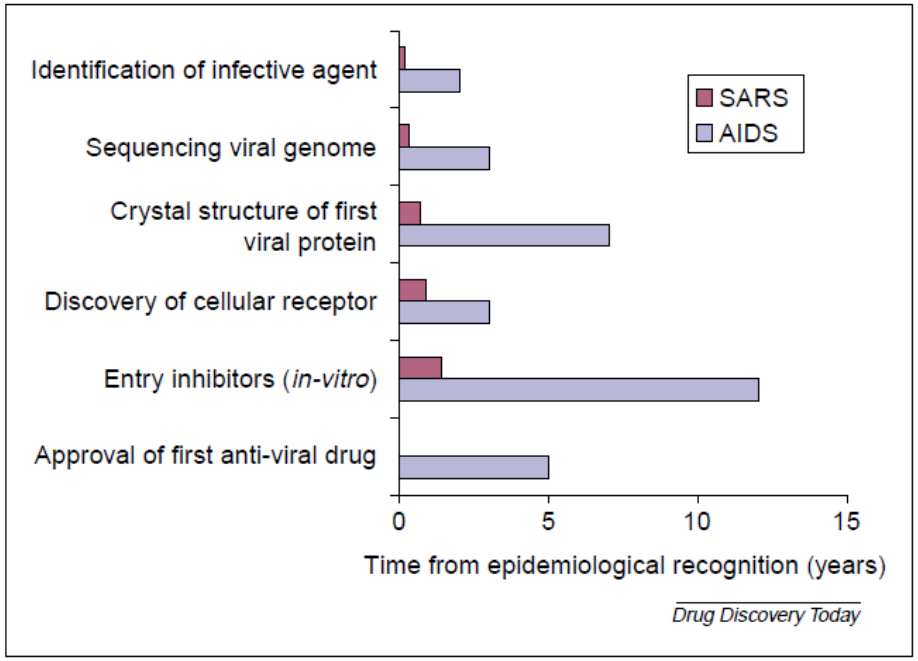
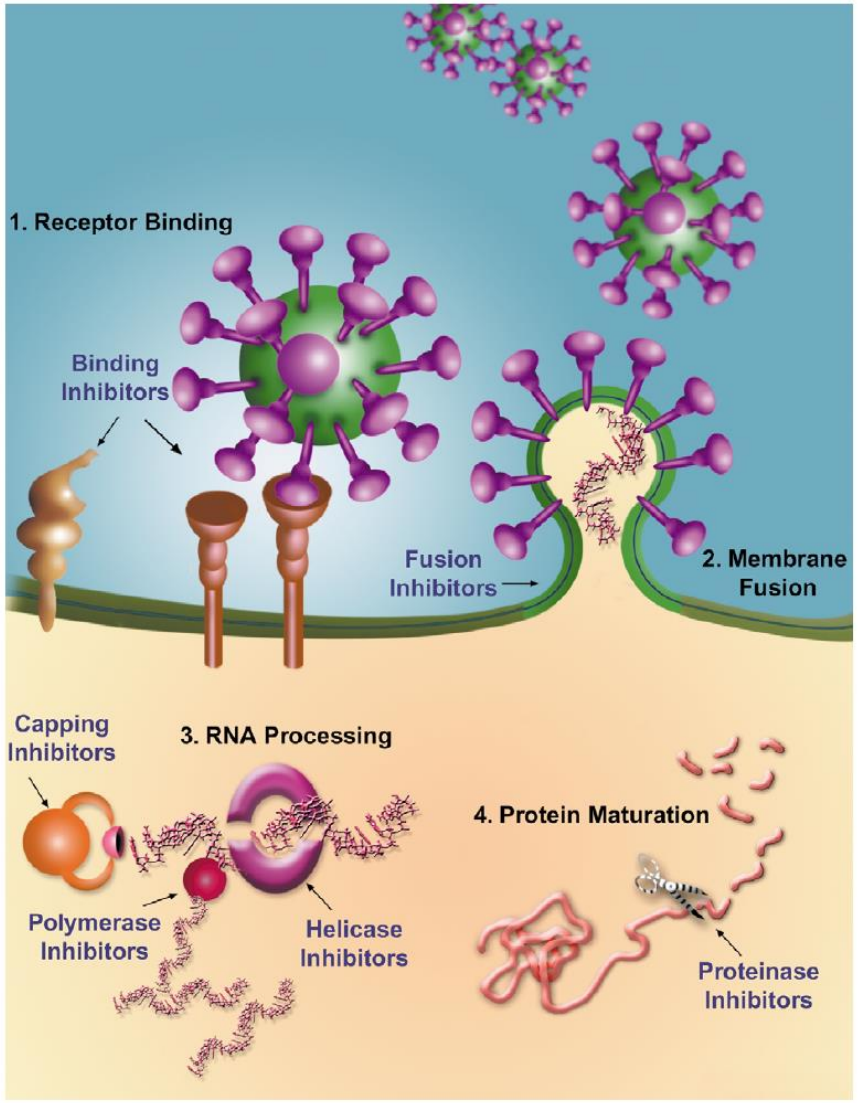


FIGURE 3

A time line comparing key achievements in AIDS and SARS research. Effective international collaborations and technological advances greatly accelerated the understanding of viral diseases. It is anticipated that these research achievements will also lead to faster drug discovery and development.



The SARS-CoV life cycle is vulnerable to therapeutic intervention in several places. (1) Virus binding to cellular receptors. Outside the cell, blocking the interaction of SARS-CoV with the cellular receptor will prevent the virus from attaching to host cells. Co-receptor antagonists will prevent the initiation of the next step. **(2)** Membrane fusion of the virion with the host. Fusion inhibitors will block merging of the viral membrane with the host cell membrane. **(3)** Viral RNA processing. Within the cell, transcription and multiplication of the viral RNA can be blocked by polymerase and helicase inhibitors. Translation of the viral proteins might be inhibited by blocking mRNA capping. **(4)** Protein maturation. Viral proteins will not mature in the presence of protease inhibitors, rendering them useless.

New horizons for antiviral drug discovery from virus–host protein interaction networks

Benoît de Chassey[1,2], Laurène Meyniel-Schicklin[1,2], Anne Aublin-Gex[1,2], Patrice André[1,2,3] and Vincent Lotteau[1,2,3]

Viruses are recurrent socio economical and health problems each year worldwide. Current drugs are mainly directed against viral components and select resistant strains that urge the need to develop new antiviral therapeutics. High-throughput screening technologies now allow to draw comprehensive genome-wide maps of physical and genetic virus–host interactions. This has been done recently for several viruses such as HIV, HCV, DENV and FLUAV and revealed a wealth of potential antiviral cellular targets. Systems-level analysis of virus–host protein networks and subnetworks begins to uncover several specific points of intervention for a human centered drug development. We present here this new paradigm in antiviral drug discovery together with the first promising antiviral molecules.

Table 1

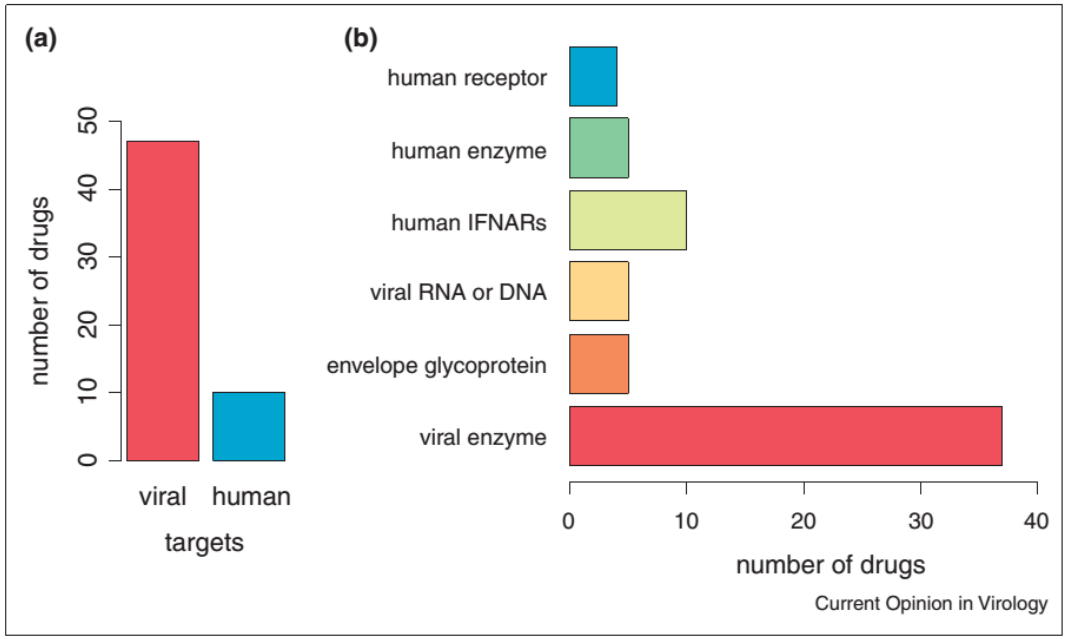
List of relevant interaction databases

A. Databases providing lists of protein–protein interactions between viruses and human

Database	Website	Manual curation	Last release date
VirHostNet	http://pbildb1.univ-lyon1.fr/virhostnet	x	2010/11
VirusMINT	http://mint.bio.uniroma2.it/virusmint	x	2008/09
HPIDB	http://agbase.msstate.edu/hpi/main.html		2012/04
Phisto	http://www.phisto.boun.edu.tr		2012/03
BIND	http://bind.ca	x	2006/07
IntAct	http://www.ebi.ac.uk/intact	x	2012/06
PIG	Discontinued		–

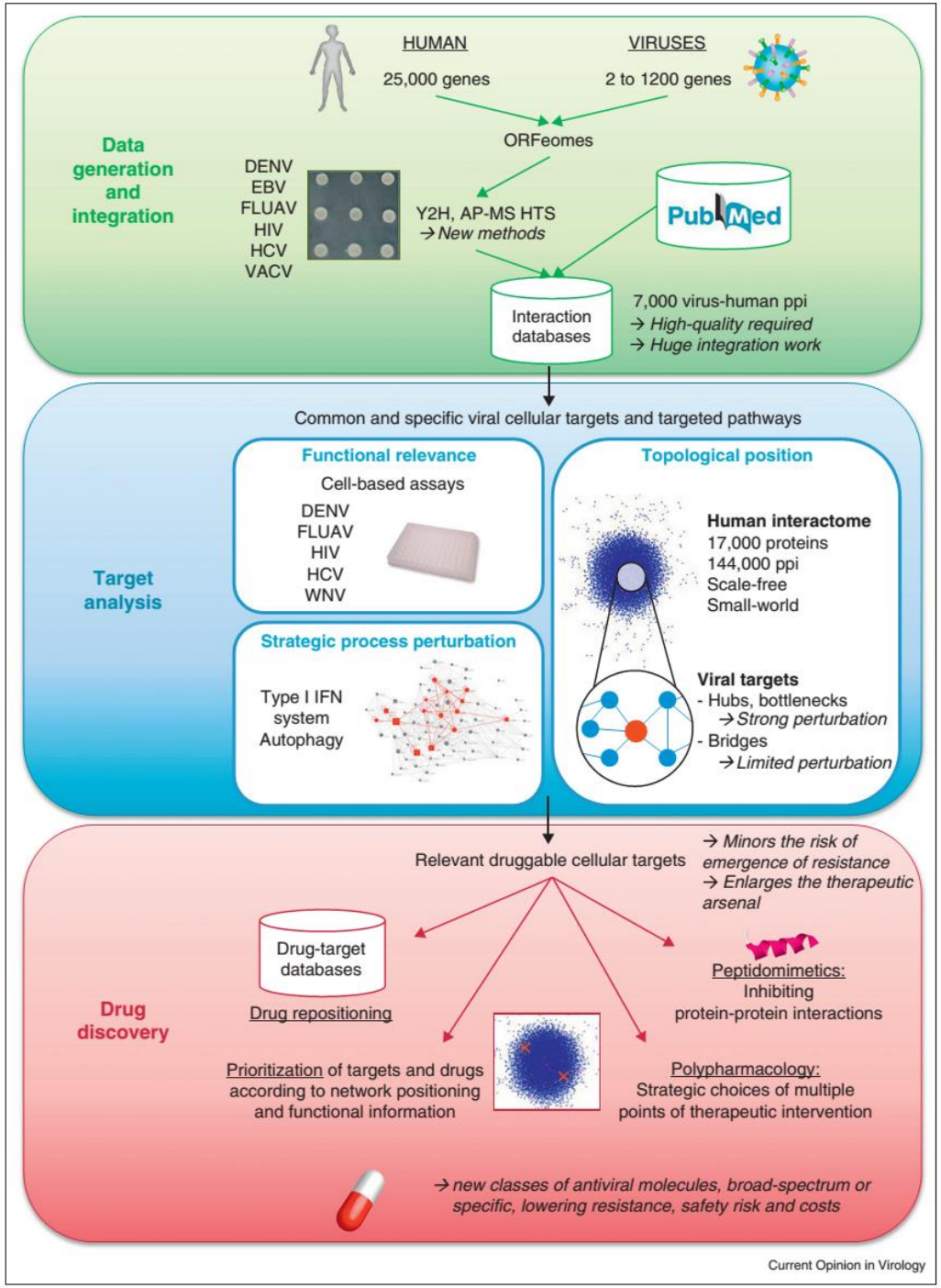
B. Databases providing lists of drug-target interactions

Database	Website	Last release date
DrugBank	www.drugbank.ca	2011/01
Therapeutic Target DB	http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp	2011/08
PharmGKB	www.pharmgkb.org	2012/06
STITCH	http://stitch.embl.de	2012/04
SuperTarget	http://insilico.charite.de/supertarget	2011/11
ChEMBL	www.ebi.ac.uk/chembldb	2012/02



Current FDA-approved antiviral drugs and their targets. (a) Number of FDA-approved antiviral drugs targeting viral and human cellular elements. All interferon molecules have been gathered as a single molecule. (b) Distribution of FDA-approved antiviral drugs according to the nature of their target: human receptor, enzyme or IFNARs or viral RNA, DNA, envelope glycoprotein or enzyme. (Source DrugBank)

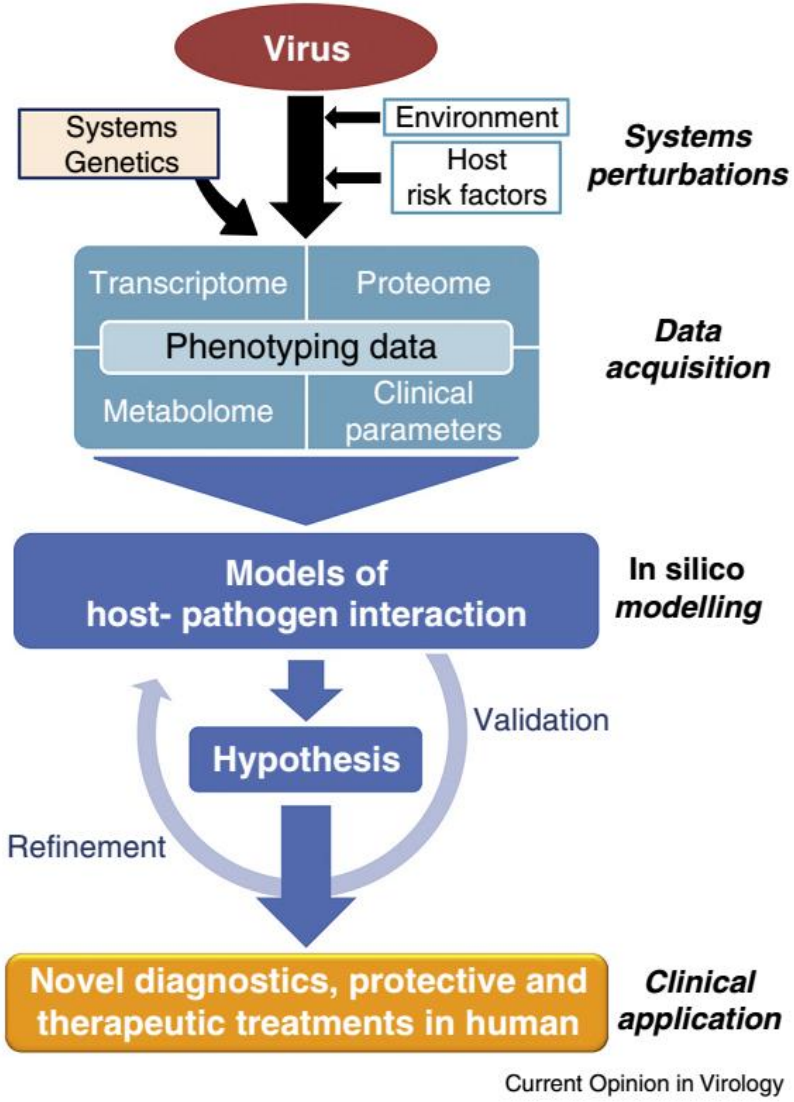
Overview of the innovative strategy from virus–host interaction data generation to the rational discovery of new antiviral drug candidates



Systems biology and systems genetics — novel innovative approaches to study host–pathogen interactions during influenza infection

Heike Kollmus[1,2], Esther Wilk[1,2] and Klaus Schughart[1,2,3]

Influenza represents a serious threat to public health with thousands of deaths each year. A deeper understanding of the host–pathogen interactions is urgently needed to evaluate individual and population risks for severe influenza disease and to identify new therapeutic targets. Here, we review recent progress in large scale omics technologies, systems genetics as well as new mathematical and computational developments that are now in place to apply a systems biology approach for a comprehensive description of the multidimensional host response to influenza infection. In addition, we describe how results from experimental animal models can be translated to humans, and we discuss some of the future challenges ahead.



Systems analysis approach to study host–pathogen interactions during influenza infection.