Concise Review: New Paradigms for Down Syndrome Research Using Induced Pluripotent Stem Cells: Tackling Complex Human Genetic Disease

JAMES A. BRIGGS, ELIZABETH A. MASON, DMITRY A. OVCHINNIKOV, CHRISTINE A. WELLS, ERNST J. WOLVETANG

Key Words. Chromosomal aberrations • Reprogramming • Induced pluripotent stem cells • Neural differentiation

Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St Lucia, Queensland, Australia

Correspondence: Ernst J. Wolvetang, Ph.D., Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St Lucia, Queensland 4072, Australia. Telephone: 61-7-3346 3894; Fax: 61-7-3346 3973; E-Mail: e.wolvetang@uq.edu.au

ABSTRACT

Down syndrome (DS) is a complex developmental disorder with diverse pathologies that affect multiple tissues and organ systems. Clear mechanistic description of how trisomy of chromosome 21 gives rise to most DS pathologies is currently lacking and is limited to a few examples of dosage-sensitive trisomic genes with large phenotypic effects. The recent advent of cellular reprogramming technology offers a promising way forward, by allowing derivation of patient-derived human cell types in vitro. We present general strategies that integrate genomics technologies and induced pluripotent stem cells to identify molecular networks driving different aspects of DS pathogenesis and describe experimental approaches to validate the causal requirement of candidate network defects for particular cellular phenotypes. This overall approach should be applicable to many poorly understood complex human genetic diseases, whose pathogenic mechanisms might involve the combined effects of many genes. STEM CELLS TRANSLATIONAL MEDICINE *2013;2:175–184*

INTRODUCTION

Down syndrome (DS) is a complex disease caused by trisomy of human chromosome 21 (HSA21) that occurs in about 1 in 750 live births [1]. In most cases the extra HSA21 is the result of a maternal segregation error in meiosis I [2, 3] or, in rare cases, Robertsonian translocations (4%) or mosaicism (1%) [4]. Individuals with DS exhibit a large range of pathologies, some of which are highly penetrant, such as cognitive defects, premature Alzheimer disease and aging, and dysmorphic facial features; as well as pathologies that are highly variable, such as cardiac defects, which occur in 50% of cases [5]. DS individuals also have an increased risk of duodenal stenosis (250 times normal risk), increased incidence of Hirschsprung disease and leukemia (30 times and 300 times normal risk, respectively), and reduced incidence of solid tumors [5].

Although HSA21 is considered to host approximately 350 protein-coding genes, for many years the prevailing hypothesis was that DS was the result of dosage sensitivity of only a small subset of these genes, which are clustered together in one chromosomal region denoted the Down syndrome critical region (DSCR) [6–11]. However, chromosomal engineering experiments in mice [12] and, more recently, genotype-phenotype analysis of human individuals

with partial trisomy for regions of HSA21 [13, 14] have now demonstrated that the complete DSCR is not necessary or sufficient for most DS phenotypes. Rather, these analyses have revealed regions linked to individual DS phenotypes that are distributed over large regions of chromosome 21, suggesting that there are potentially many causative genes in DS [13] (summarized in Fig. 1).

The new zeitgeist proposes that DS phenotypes arise from a complex interplay of these many causative genes, which can exert pathological influence at various stages of development or maturity [15]. Under this hypothesis, it has been suggested that the genetic mechanisms responsible for DS phenotypes can be subdivided into the following (from [15]): small sets of dosage-sensitive genes, such as APP, in early onset Alzheimer disease pathology [16, 17]; interacting genes that drive major effects, such as DSCR1 and DYRK1A in heart abnormalities and solid cancer risk [18, 19]; or the collective influence of many coincident small effects. Although this latter mechanism has not been formally demonstrated, small sets of causative genes responsible for most DS phenotypes have resisted identification, and it is increasingly becoming acknowledged that coincident small effects are major drivers of complex disease pathogenesis [20–23]. Moreover, this possibility is supported by the observations that most trisomic genes in DS
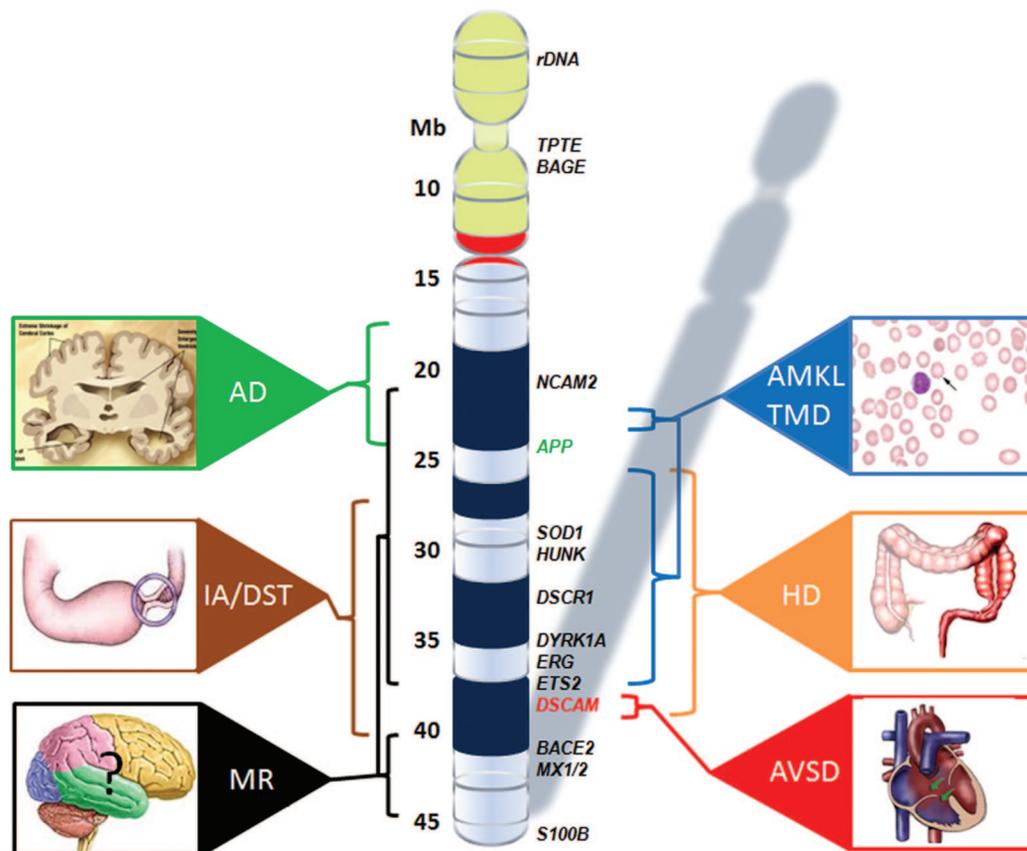
Figure 1. HSA21 regions linked to particular Down syndrome (DS) phenotypes. Analysis of human segmental trisomies [13] have identified several dispersed HSA21 regions linked to particular DS phenotypes, including AVSD, AD, IA/DST, MR, HD, and AMKL/TMD. These chromosomal regions are relatively large and overlap incompletely, suggesting the existence of potentially many causative genes in DS (well-known examples indicated at their approximate locations). This is contrary to the long-held notion that genes responsible for DS may be clustered into a single chromosomal region denoted the DSCR [6–11]. The centromere is shown in red, the HSA21 short arm in yellow, and the HSA21 long arm in blue, with light and dark regions reflecting G-banding. Numbers represent distance in Mb from the distal end of the HSA21 short arm. Abbreviations: AD, Alzheimer disease; AMKL/TMD, acute megakaryoblastic leukemia/transient myeloproliferative disorder; AVSD, atrioventricular septal stenosis; DSCR, Down syndrome critical region; HD, Hirschsprung disease; IA/DST, imperforate anus/duodenal stenosis; Mb, megabases; MR, mental retardation; rDNA, recombinant DNA.

are indeed overexpressed in multiple tissues [24–29] and that gene deregulation in DS further impacts up to one-third of disomic genes [29–31]. In addition to dosage effects, this gene deregulation seems to result from complex cis- and trans-, intra- and interchromosomal regulatory interactions, including some that may impose genome-wide effects.

As with many complex diseases, most DS phenotypes manifest with variable expressivity. This could be due in part to allelic variation in modifier genes, with only certain combinations of alleles exerting pathological effects, or chromatin modifications such as parental imprinting, which may alter the expression of particular modifier genes [15, 30]. These effects contribute to a complex interplay between genetic, developmental, and external parameters, which collectively determine phenotype expressivity (summarized in Fig. 2).

THE "NEAREST" MOUSE MODELS OF DS

Mouse models are currently the primary tools used to study DS etiology and have contributed much of our current understanding of DS pathogenesis [32, 33]. These models are powerful because they provide researchers with access to developing tissues and are compatible with a wealth of technologies that allow the

roles of particular genes in establishing disease phenotypes to be interrogated. It has become clear, however, that for complex diseases such as DS, with potentially many causative genes, generating mouse models with the exact genetic defect found in the human disorder might not be possible, resulting in an imperfect recapitulation of disease etiology and potentially confounding species-specific disease mechanisms [34]. It is additionally concerning that for many DS pathologies (such as cognitive impairment) the affected mouse and human developmental processes differ significantly [35].

The most sophisticated mouse models of DS include mice that contain single-chromosome duplications for the three mouse chromosomal regions orthologous to HSA21 [36] or contain an extra copy of the complete HSA21 (Tc1 mice) [37]. Triple segmental duplication DS mice models display cognitive impairment, including reduced hippocampal long-term potentiation, and hydrocephalus [36]. However, these mice lack B-amyloid plaque deposition, which is highly penetrant in DS [36]. It is thought that beyond mouse-human brain differences, such discrepancies could be because mouse-human orthology of HSA21 is not complete [38, 39].

Tc1 mice somewhat circumvent the issue of incomplete mouse-human orthology for modeling DS by carrying a complete
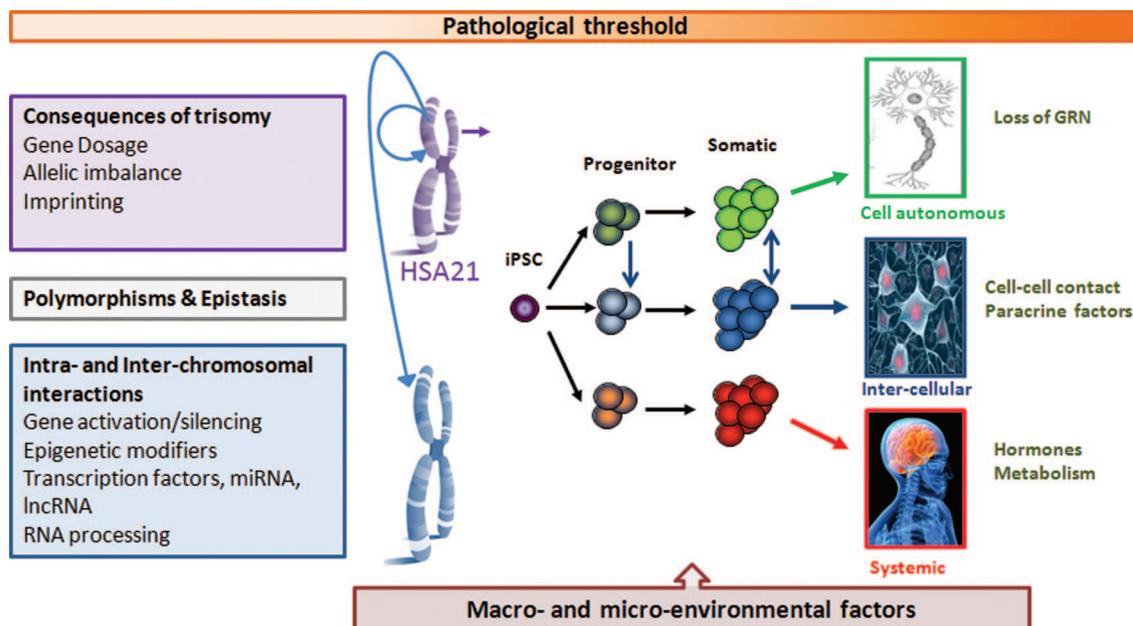
Figure 2. Mechanisms underlying etiology and variable expressivity of Down syndrome (DS) phenotypes. Down syndrome phenotypes are classically understood to arise from gene dosage of HSA21 genes (purple arrow), which can be uniquely influenced in different DS individuals by allelic variation or differential imprinting. It is now becoming clear that subsets of HSA21 genes can in fact often deviate significantly from 1.5-fold overexpression [24−29] and that trisomy of HSA21 additionally results in significant genome-wide transcriptome deregulation [29−31], seeming to suggest the existence of additional complex intra- and interchromosomal genetic interactions that may also contribute to DS pathogenesis (blue arrows). Candidate drivers of such mechanisms include HSA21 epigenetic modifiers, transcription factors, or noncoding RNAs. Notably, the complexity of these interactions presents abundant additional opportunities for polymorphism and epistasis to contribute to interindividual differences in gene deregulation in DS, and may be a major driver of variable phenotype expressivity. Gene deregulation in DS underpins a complex etiology that potentially involves multiple cell-autonomous, intercellular and systemic mechanisms, some of which may additionally be influenced by macro- and microenvironmental factors, and that together exert pathological influence at various stages of development or maturity such that normal physiological processes approach a pathological threshold. Abbreviations: miRNA, microRNA; iPSC, induced pluripotent stem cell; lncRNA, long noncoding RNA.

copy of HSA21 [37]. However, the utility of these mice is limited by the variable contribution of humanized ES cells to mouse chimeras and by the fact that they are sterile, limiting analysis to mice that are mosaic for the additional copy of human HSA21 [34]. Nevertheless, these mice were found to recapitulate many, but still not all, of the brain-, heart-, and craniofacial development-related DS traits observed in humans [37]. Other than intrinsic developmental differences between mice and humans, this could be because human genes placed in a mouse cellular context will be regulated and behave differently than they would in a human [33].

A ROLE FOR DS INDUCED PLURIPOTENT STEM CELLS

Reprogramming of somatic cells of disease-affected patients to a pluripotent state, generating disease-specific induced pluripotent stem cells (iPSCs), has recently emerged as an alternative paradigm for modeling human genetic disease in vitro [40, 41]. In combination with directed differentiation protocols, these iPSC models theoretically make it possible to derive any given disease-affected human cell-type [42]. Fluorescence-activated cell sorting purification protocols further allow isolation of highly reproducible populations of diseased and wild-type cells [43], which may then be interrogated by the library of tools currently used for the manipulation of human cells in vitro. Because of this, disease-specific iPSCs not only supersede much of the utility offered by mouse disease models but also do so within a human cellular context that is able to faithfully capture complex genetic disease genotypes, making them a particularly powerful tool for investigating the mechanisms of

complex genetic diseases such as DS. Disease-specific iPSCs also have the potential to increase our understanding of the genetic (recapitulated by iPSCs) versus environmental (not recapitulated) contributions to complex disease phenotypes.

Using iPSCs to Model DS and Other Complex Diseases

An increasing array of disease-specific iPSC lines have been generated [44], but to date their utility in disease modeling has mainly been demonstrated for single-gene disorders [40, 45–48]. We recently demonstrated that neurally differentiated DS iPSCs are able to recapitulate several well-established DS phenotypes, including increased gliogenesis and sensitivity to oxidative stress-induced apoptosis, the latter of which could be pharmacologically prevented [49]. Another group has also recently shown that DS iPSC-derived neurons recapitulate Alzheimer disease-associated amyloid and tau pathologies in vitro [50], which seems to demonstrate that iPSC disease models are also able to recapitulate degenerative phenotypes over relatively short time scales in vitro.

Schizophrenia is another complex disorder that has recently been successfully modeled using iPSCs [51]. Schizophrenia iPSC-derived neurons were reported to display reduced neuronal connectivity (assessed by an elegant in vitro assay measuring the trans-synaptic transfer of a modified rabies virus), reduced neurite extensions, reduced PSD95 gene and protein expression, and reduced glutamate receptor expression, metrics consistent with phenotypes observed in post-mortem schizophrenia brains [51]. Notably, some of these phenotypes in schizophrenia iPSC-derived neurons

could be partially rescued with the known antipsychotic drug loxapine, suggesting that these cells will additionally have utility in drug screening [51]. Taken together, these studies demonstrate that iPSCs are able to recapitulate both DS and other complex disease phenotypes, including those with both early and delayed disease onset, and that disease-specific iPSCs allow testing of corrective therapeutics.

Confounding Factors and Limitations When Using iPSC Models of Complex Disease

It is tempting to speculate that disease-specific iPSCs will become a core component of future investigations into the mechanisms of complex human genetic disease in general. However, before too much expectation is placed on these models for disease mechanism discovery, researchers need to be aware of several potential technical confounders and other limitations. An obvious limitation is that some disease phenotypes arise from interactions between multiple cell types and involve three-dimensional or non-cell-autonomous parameters that can only be modeled in vitro (if at all) by well-designed organoid or coculture assays, which are technically challenging to engineer. Capturing phenotypes that arise from a combination of genetic background and environmental influences (lifestyle, diet, etc.) over prolonged periods of time will also remain difficult to model in vitro. Further to these inherent limitations, significant gaps in directed differentiation protocols currently prevent the derivation of sufficiently pure populations of cells for many lineages [42] and are susceptible to potentially confounding variability in differentiation potential, between human embryonic stem cell or iPSC lines (including from the same individual/genotype) [52]. It is also known that the process of reprogramming can generate genetic and epigenetic variation between iPSC lines [53–57] and that iPSCs seem to retain an epigenetic memory of their cell type of origin [58–61], both of which can have a confounding influence on postreprogramming developmental trajectories and other cellular processes. Another consequence of epigenetic memory is that disease-associated epigenetic aberrations from donor somatic tissues could plausibly be preserved through reprogramming and could introduce artifactual epigenetic aberrations in patient-derived iPSC lines, which would not normally be present in pluripotent cells. For these reasons, when using iPSC disease models it is important to generate multiple lines from prudently selected tissues of multiple affected and nonaffected (ideally from siblings or parents) individuals and to prioritize the use of nonintegrating reprogramming technologies, which have been shown to generate less reprogramming-associated genetic and epigenetic variability [62]. Although this is a significant technical and financial burden, it will help to reduce noise in analyses from genetic variation and experimental artifacts, both maximizing power to identify pathogenic pathways and reducing the likelihood of discovering false disease mechanisms.

INTEGRATING GENOMICS TECHNOLOGIES AND iPSCs TO UNCOVER THE MOLECULAR NETWORKS DRIVING DS PHENOTYPES

A picture of many complex diseases such as DS is emerging in which pathological processes can result from perturbed behaviors of biological networks, driven by the collective effects of many risk genes acting in concert within related network modules [20–23]. This is in contrast to single-gene disorders or more simple DS genetic mechanisms, where altered function or dysregulation of small sets of genes is sufficient for phenotype manifestation, and necessitates a different approach to identifying causative complex disease mechanisms. If we are to understand the more complex genetic mechanisms of diseases such as DS, the questions become (a) how are we going to acquire a complete picture of the genetic defect caused by trisomy 21? (b) how are we going to predict the causative genetic mechanisms from this large data set? and (c) how are we going to prove that these mechanisms are indeed causative?

Genomics Technologies to Capture Complex-Disease-Associated Transcriptome Deregulation

Technologies such as microarrays [63, 64] and, more recently, RNA sequencing [65, 66] have emerged as powerful tools allowing quantification of genome-wide gene expression. Microarray studies on peripheral blood cells or post-mortem material have confirmed that most HSA21 genes are overexpressed by 1.5-fold in DS, reflecting gene dosage [24–27, 29]. It was these studies that also surprisingly identified significant global transcriptome deregulation in DS [29, 30]. Our own transcriptome study of DS iPSCs confirmed that thousands of genes are deregulated across the genome even in pluripotent DS cells [49], further suggesting that this global deregulation may be involved in establishing DS developmental defects from the earliest stages of development. Global transcriptome deregulation in DS is likely to result from HSA21 transcription factors, epigenetic machinery, and noncoding RNAs acting in trans. For example, overexpression of HSA21 microRNAs (miR-155 and miR-802) leads to progressive silencing of MeCP2, CREB1, and MEF2C in DS post-mortem brain specimens [67]. Application of microarrays and RNA sequencing technologies across periods of development in DS can also provide insight into developmental disease mechanisms, and the dynamics of transcriptome deregulation in DS. Such an approach was used to identify Notch signaling and cell migration abnormalities during DS cerebellar development and demonstrated that overexpression of trisomic genes was remarkably stable through development [68].

Profiling genome-wide gene expression both in tissues of interest and across periods of development promises to produce a comprehensive annotation of the transcriptome deregulation responsible for DS phenotypes. However, it remains a major challenge to effectively interpret these data in a way that will elucidate the mechanisms that are *causative* in DS and therefore appropriate for therapeutic targeting.

Predicting Pathogenic Disease Mechanisms

It is becoming clear that many phenotypes of complex diseases such as DS arise from potentially unique gene deregulation that converges on particular pathogenic processes [20–23]. This limits the power of classic approaches for identifying disease mechanisms from transcriptome data, which might aim to identify mis-expressed gene sets common to all patients or patients with a particular phenotype. A more promising strategy is to use pathway or gene-set enrichment analysis methods (summarized in Table 1). These methods do not rely on common gene sets and are thus well suited to reveal deregulated processes for complex disease mechanism discovery [69, 70]. Such approaches have elucidated defects in dopamine receptor signaling, insulin-like growth factor 1 signaling pathways, and linked proteasomal and mitochondrial pathways from Parkinson disease patient brain

Table 1. Tools to analyze genome-wide gene expression data for complex disease mechanism discovery

Approach	Rank gene expression	Identify groups	Gene set analysis
Purpose	Arrange genes in rank order based on a test statistic and significance [121]. For experimental design see Churchill [122], and for analysis approaches see Slonim [123].	Group data at the sample or gene level with (supervised) or without (unsupervised) previous knowledge or classification schemes.	Identify overrepresented groups of genes (pathway analyses) or functional classes (ontology analysis). For review of current approaches see Khatri and Draghici [75].
Statistical tools	**t test/p value:** Identifies significant change between repeated measures of the same variable in two groups. **Analysis of variance/F-statistic:** Generalizes the t test to multiple sample groups. **B-statistic:** Represents the log of odds that the gene of interest is differentially expressed [124]. **Fold change:** Ratio of the expression levels of a gene between groups [125].	**PCA:** Reduces the dimensionality of the data, while conserving as much variance as possible [126, 127]. **Hierarchical cluster:** Closely related samples/genes are clustered by a similarity metric [64] (e.g., Spearman or Pearson correlation). **Fuzzy clustering:** Samples may belong to more than one cluster, and ranked by membership levels.	**GSEA:** Determines whether members of a gene set are primarily at the top/bottom of the full ranked gene list [69]. **Attract:** Identifies patterns in well-defined KEGG [128] pathways that distinguish groups, and decomposes pathways into meta-genes representing the patterns. This is used to identify groups of highly correlated genes from the whole data set [73]. **GO:** Identifies functional classes that are overrepresented in a gene list. Functional classes are derived from the Gene Ontology (GO) project.
Software packages	**MeV** [129, 130] (multiexperiment viewer) has a comprehensive suite of statistical tools for a variety of data types, including microarray. **GenePattern** [131] provides free access to a number of tools for gene expression analysis. **R bioconductor** packages (e.g. lumi [132], limma [124]) for microarray data processing and differential expression analyses. **SAM** (significance analysis of microarrays) is a statistical technique for finding significant genes in a set of microarray experiments [125].	**Genesis** [133] is a suite of analysis, clustering, and visualization tools for microarray data (allows both supervised and unsupervised clustering). **MATISSE** [134] (module analysis via topology of interactions and similarity sets) detects functional modules by using interaction networks and expression data (supervised).	**GenePattern** [131], **MeV** [129, 130]. **MAPPFinder** [135] integrates annotations of the GO Project with the free software package GenMAPP [136]. **DAVID** [137] provides a comprehensive set of functional annotation tools for identifying enriched GO terms. **PANTHER** [138] allows mapping of multiple gene lists to PANTHER molecular function and biological pathways. **Cytoscape** [139, 140] visualizes and analyzes molecular interaction networks and integrates these networks with annotations, GO categories, and gene expression profiles. **IPA** (http://www.ingenuity.com) is a subscription-based package that integrates data from a variety of experimental platforms providing insight into the molecular and chemical interactions, cellular phenotypes, and disease processes.
Advantages	Identifies, on average, which genes contribute to overall phenotypic differences.	Groups of genes that share similar expression patterns may also share biological functions [141] (e.g. common regulators [142]). Identifying patterns of gene expression that correlate with experimental artifacts can be used to filter samples or genes. Sample clustering can lead discovery of disease subclasses [143].	GSEA considers all genes in an experiment, not just those above an arbitrary cutoff, and allows flexibility in the selection of gene sets and comparisons between different platforms. Biochemical pathway databases (such as KEGG) provide a higher level of resolution of gene-pathway relationships than categorical definitions alone. Attract searches the whole data set to identify groups of genes showing patterns of expression that are highly correlated with the meta-genes from known pathways. This presents an opportunity for discovery [74] and considers all genes in the experiment, and is not limited to those with functional annotations. The GO project (http://www.geneontology.org) is actively curated by an international consortium. The project provides an ontology of defined terms representing gene product properties (cellular component, molecular function, and biological processes).
Disadvantages	**Sensitivity:** With too few sample replicates, it is difficult to obtain accurate estimates of variance. **Multiplicity:** Thousands of hypotheses are tested simultaneously, increasing the chance of false positives. The statistic often requires post hoc adjustment for this reason [144]. **Concordance:** Ranked gene lists often do not replicate between different experiments [15, 145]. The criteria used to define differentially expressed genes can have a dramatic impact on the overlap of the resulting gene lists. Significance thresholds are set arbitrarily, and interpretations of the statistic require data to be sampled from a normal distribution. Normalization and transformation are used for data not meeting these assumptions.	Clustering approaches rely on the presumption that most variation in the data can be explained by a small number of variables.	Gene sets are usually comprised of either members of well-defined pathways or genes with functional annotation. A small number of genes contribute to these annotations, due to the limited amount of functional information available. Ontological analyses are limited to the three GO functional categories, developed primarily using known protein domains to predict gene function, for which there are always exceptions. Protein domains with poor functional information or genes lacking known protein domains are therefore under-represented. For genes that are involved in several biological processes, equal weight is given to each process, such that it is not always possible to identify the most relevant functional category using ontological analysis alone.

gene expression data [71, 72]. Another approach is to reverse the traditional pathway-analysis workflow of mapping mis-expressed genes to annotated pathways and to instead begin by identifying pathways that best distinguish between phenotypes of interest using methods such as "attract" [73]. "Attract" is also able to decompose the most informative expression profiles into "synexpression" patterns and allow the identification of novel genes that had not previously been annotated to a pathway of interest but that nonetheless share the same expression profile and therefore might be under shared regulatory control or contribute to similar biological processes. This approach has been used to implicate novel cell signaling pathways in complex diseases such as Parkinson disease and schizophrenia, and it has also identified disease-associated differences in expression variance of core biological networks, defining disease-associated alterations in network constraints [74].

As a cautionary note, however, most of these annotation-based methods are still being actively developed, with several sources of bias, including type I and type II statistical error, acting to potentially confound interpretation and reduce reproducibility of results [69]. For example, most publicly available pathway analysis tools generally rely on public databases of gene

annotation, which are often biased toward biological processes that have been studied in more detail than others (e.g., apoptosis) [75]. This reduces the discovery component of these analyses. Moreover, a recent study found that random gene expression signatures were commonly significantly associated with breast cancer outcome by standard gene expression signature analysis methods, thus implying that previously identified associations by these methods are likely spurious and nonreflective of bona fide disease mechanisms [76], emphasizing the essential requirement of assessing the reproducibility of transcriptome studies. Nonetheless, with refinement, pathway-based methods for interpretation of gene expression data are likely to prove particularly powerful in investigations into complex diseases such as DS, because of their ability to detect instances where small changes in the expression of a group of genes is likely to exert an amplified effect via their enrichment as a group within a particular biological pathway. Such pathways would warrant functional experimentation and may constitute appropriate targets for therapeutic intervention.

Another approach is to look for evidence of common causes for large sets of transcriptome deregulation in DS, such as cis- or trans-regulatory effects of HSA21 gene products, including transcription factors, epigenetic modifiers, and noncoding RNAs [77, 78]. This is possible through screening for binding motif enrichment within sets of deregulated genes [79] or for overlap of transcriptome deregulation with differential methylation or chromatin modification profiles across the genome [80–83]. Such analysis will provide insight into the as yet unexplored contribution of epigenetic mechanisms to DS pathogenesis. Identification of genes that are drivers of global transcriptome deregulation in DS is useful because these are likely to be particularly appropriate therapeutic targets, as their normalization could potentially correct coordinated pathogenic processes.

LINKING NETWORK DEFECTS WITH DISEASE PHENOTYPES USING iPSCs

Study Design

For any predicted pathogenic complex disease mechanism, the next step is to experimentally demonstrate the necessity and/or sufficiency of that mechanism for specific disease features. To do this using iPSC disease models, this first requires that some disease feature can be captured by an in vitro cellular phenotype. Second, the candidate causative mechanism(s) must be predicted from omics data generated from the relevant cell type and developmental stage, since a mechanism that is pathogenic in one setting may not be in another. Once these criteria are met, the reproducible nature of iPSC-derived models leaves investigators well placed to carry out functional experiments. The goal is to experimentally modulate candidate pathogenic mechanisms and examine the consequences on phenotype manifestation, to establish whether the mechanism is causative in pathogenesis. Normalizing a particular disease mechanism and rescuing the phenotype demonstrates the necessity of that aberration for the phenotype. Alternatively, if recapitulating a molecular aberration in wild-type cells recreates a disease phenotype, this proves the sufficiency of that mechanism for the phenotype. Importantly, the ability to normalize phenotypes by specific modulation of molecular aberrations is likely to constitute a starting point for therapeutic strategies.

Approaches to Proving Causality of Predicted Pathogenic Mechanisms

A number of versatile technologies are available for the specific and regulated over- and underexpression of transcripts of interest. The most established of these include RNA interference-based approaches and overexpression of exogenous transgenes, both of which may be delivered in a variety of ways depending on the particular experimental requirements [84–88]. High degrees of control over the timing of gene knockdown and overexpression are achievable using inducible expression systems [89], allowing simple assessment of the roles of particular genes during narrow developmental windows or periods following a particular stimulus (maybe to simulate a physiological function). One limitation of these systems, however, is a lack of control over the magnitude of transcript knockdown or overexpression. This is particularly limiting when interrogating complex genetic disease mechanisms, since gene deregulation in these disorders is often expected to be subtle and to potentially involve many related gene products.

More sophisticated control of gene expression modulation is likely to be possible with emerging transcription activator-like effector (TALE)-based technologies [90, 91]. These technologies allow transcriptional activator and repressor modules to be targeted to specific loci anywhere in the genome [90, 91]. Importantly, by substituting activator or repressor modules with different activities, or altering TALE concentrations, it should be possible to achieve a high degree of control over the magnitude of change produced in gene expression. More farsighted possibilities for the subtle modulation of gene regulatory networks, such as those that use synthetic network components to modify the behavior/logic of particular gene regulatory modules [92, 93], might prove particularly useful in modeling and normalizing more elaborate pathogenic genetic mechanisms for some complex-diseases. In some cases, it might also be effective to target gene products, for example by using small molecules or recombinant proteins, to directly examine the contribution of deregulation of these gene products to transcriptome deregulation and pathogenesis [94]. By combining multiple modules/technologies, it should be possible to sensibly modulate complex gene deregulation phenotypes. Once this can be achieved we can ask whether a particular network defect is indeed necessary or sufficient for a particular phenotype.

In cases where large sets of deregulated genes appear to result from trans-regulatory effects of individual deregulated transcription factors, epigenetic modifiers, or noncoding RNAs, technologies such as chromatin immunoprecipitation sequencing [95, 96] and capture hybridization analysis of RNA targets/chromatin isolation by RNA purification [97, 98] may be useful to demonstrate binding of the trans-acting factor to deregulated target loci. Normalizing the expression of the trans-acting factor(s) (by the above approaches) could then be used to confirm the functional importance of these genes in regulating their targets and contributing to disease phenotype manifestation.

Generating Genotype-Phenotype Maps of DS Using Chromosomal Engineering of DS iPSCs

Chromosomal engineering using DS iPSCs could allow assessment of the role of trisomy of discrete chromosomal regions

in pathogenesis of DS cellular phenotypes. TAL effector nucleases (TALENs) allow targeted cleavage of almost any site in the human genome [90, 91] and have been used for chromosomal engineering of iPSCs [99]. We are currently pursuing a chromosomal engineering approach that allows deletion of large regions of HSA21, adapted from a method used in mice [12, 100]. The method introduces two cassettes bordering the region to be deleted, containing both tandemly oriented Cre recombinase *loxP* substrate sites for gene excision and positive and negative genetic selection markers, allowing simple selection of successfully excised regional HSA21 deletion DS iPSC clones [101]. This will be particularly informative when rationally implemented to test genotype-phenotype maps generated from human segmental trisomy patients and corresponding clinical data [13]. Combination of regional deletion DS iPSCs with unmodified DS iPSC lines, and whole-HSA21 deletion DS iPSC lines, would offer perfect experimental controls, with identical genetic backgrounds.

PERSPECTIVES AND FUTURE DIRECTIONS: iPSC-BASED PARADIGMS FOR INVESTIGATION INTO COMPLEX DISEASE ETIOLOGY

The strategy for research into complex disease presented in this review is primarily motivated by appreciation of the limitations of reductionist-type approaches that focus on individual proteins or signaling pathways in the setting of complex disease. These studies have dominated human genetic disease research for the last two decades, and their limitations are well evidenced by the lack progress to date in understanding the mechanisms of, and designing treatments for, complex human genetic diseases in general. Many complex diseases such as DS are thought to arise in part from the collective contributions of potentially small perturbations in large numbers of causative genes. Moreover, in DS causative genes need not be limited to HSA21, because of the ostensibly large influence of trans-regulatory effects on global transcriptome deregulation. To account for this we describe strategies to capture systems-level genetic deregulation, and we describe approaches that are able to predict mechanisms that are likely to be pathogenic from this profile. Importantly, these methods are sensitive to small-magnitude changes in large sets of functionally related genes. We further describe how iPSCs can be used in conjunction with a variety of established and emerging technologies to functionally interrogate these predicted pathogenic genetic mechanisms, to formally demonstrate the necessity and sufficiency of these mechanisms for pathogenesis of particular disease phenotypes. iPSCs have enabled this paradigm of research by providing a renewable and reproducible source of disease relevant human cells, theoretically representing any cell type or developmental stage, and the impact of iPSCs is amplified by the apparent inability of mouse models to faithfully recapitulate complex human genetic diseases.

There are a number of limitations to this strategy as it is presented that will need to be overcome to yield the true potential of iPSC-based research into complex disease. First, the strategy we have presented is not directly sensitive to alterations in the function of particular transcripts or proteins; rather, it only detects changes in their expression. Combination of DNA se-

quencing technologies with diseased donor cell lines could catalogue genome-wide mutations/single-nucleotide polymorphisms enriched in diseased cells, as done in genome-wide association studies, to aid identification of potential functional as well as expression-level alterations [102]. However, although techniques able to predict the mechanistic consequences of sequence changes to protein function are showing some promise [103], they are prone to error and have not yet been applied to noncoding RNAs. This limitation will need to be overcome if we are to reduce the requirement for time-consuming and costly experimental validation of the mechanistic influence of large numbers of disease-associated sequence variants.

Second, many complex disease mechanisms may be rooted in protein or metabolic network functional defects [20–22]. Emerging technologies are beginning to allow systems-level profiling of these networks [104–108], including post-translational protein modifications [109, 110] and protein complex stoichiometries [111, 112], even with spatial and high-throughput temporal resolution [113, 114], and within human tissues and cell lines [115–119]. Importantly, once this is possible, an overall approach similar to that which we have presented here for gene expression should be applicable, but instead centered on proteome/metabolome features or even multiomics networks [120]. It is also worth noting that although gene expression profiling cannot detect functional transcript changes or changes in protein/metabolic network function directly, in many cases it may be possible to infer such functional changes based on the structure of gene regulatory network alterations [94].

Finally, strategies to combat challenges associated with modeling complex genetic diseases will need to be implemented (as discussed above), and progress is still required in developing directed differentiation and cell sorting protocols. Nonetheless, a combination of disease-specific iPSCs with increasingly sophisticated systems-level technologies, including components allowing profiling of complex systems, interpretation of the resulting data, and the functional manipulation of relevant biological networks, promises to produce a lasting impact on the field of complex human genetic disease research and lay a foundation for the future development of therapeutic strategies.

AUTHOR CONTRIBUTIONS

J.A.B.: manuscript writing, conception and design, data analysis and interpretation, collection and/or assembly of data; E.A.M. and D.A.O.: collection and/or assembly of data, manuscript writing; C.A.W.: manuscript writing, data analysis and interpretation, collection and/or assembly of data; E.J.W.: manuscript writing, conception and design, data analysis and interpretation, collection and/or assembly of data, final approval of manuscript.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The authors indicate no potential conflicts of interest.

REFERENCES

1 Parker SE, Mai CT, Canfield MA et al. Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004–2006. Birth Defects Res A Clin Mol Teratol 2010;88:1008–1016.

2 Yoon PW, Freeman SB, Sherman SL et al. Advanced maternal age and the risk of Down syndrome characterized by the meiotic stage of chromosomal error: A population-based study. Am J Hum Genet 1996;58:628–633.

3 Hassold T, Hall H, Hunt P. The origin of human aneuploidy: Where we have been, where we are going. Hum Mol Genet 2007;16 Spec No. 2:R203–R208.

4 Petersen MB, Mikkelsen M. Nondisjunction in trisomy 21: Origin and mechanisms. Cytogenet Cell Genet 2000;91:199–203.

5 Korenberg JR, Chen XN, Schipper R et al. Down syndrome phenotypes: The consequences of chromosomal imbalance. Proc Natl Acad Sci USA 1994;91:4997–5001.

6 Delabar JM, Theophile D, Rahmani Z et al. Molecular mapping of twenty-four features of Down syndrome on chromosome 21. Eur J Hum Genet 1993;1:114–124.

7 Epstein CJ, Avraham KB, Lovett M et al. Transgenic mice with increased Cu/Zn-superoxide dismutase activity: Animal model of dosage effects in Down syndrome. Proc Natl Acad Sci USA 1987;84:8044–8048.

8 Smith DJ, Stevens ME, Sudanagunta SP et al. Functional screening of 2 Mb of human chromosome 21q22.2 in transgenic mice implicates minibrain in learning defects associated with Down syndrome. Nat Genet 1997;16:28–36.

9 Niebuhr E. Down's syndrome. The possibility of a pathogenetic segment on chromosome no. 21. Humangenetik 1974;21:99–101.

10 McCormick MK, Schinzel A, Petersen MB et al. Molecular genetic approach to the characterization of the "Down syndrome region" of chromosome 21. Genomics 1989;5:325–331.

11 Korenberg JR. Molecular mapping of the Down syndrome phenotype. Prog Clin Biol Res 1990;360:105–115.

12 Olson LE, Richtsmeier JT, Leszl J et al. A chromosome 21 critical region does not cause specific Down syndrome phenotypes. Science 2004;306:687–690.

13 Korbel JO, Tirosh-Wagner T, Urban AE et al. The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. Proc Natl Acad Sci USA 2009;106:12031–12036.

14 Lyle R, Bena F, Gagos S et al. Genotype-phenotype correlations in Down syndrome identified by array CGH in 30 cases of partial trisomy and partial monosomy chromosome 21. Eur J Hum Genet 2009;17:454–466.

15 Roper RJ, Reeves RH. Understanding the basis for Down syndrome phenotypes. PLoS Genet 2006;2:e50.

16 Cataldo AM, Petanceska S, Peterhoff CM et al. App gene dosage modulates endosomal abnormalities of Alzheimer's disease in a segmental trisomy 16 mouse model of down syndrome. J Neurosci 2003;23:6788–6792.

17 Rovelet-Lecrux A, Hannequin D, Raux G et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 2006;38:24–26.

18 Arron JR, Winslow MM, Polleri A et al. NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. Nature 2006;441:595–600.

19 Baek KH, Zaslavsky A, Lynch RC et al. Down's syndrome suppression of tumour growth and the role of the calcineurin inhibitor DSCR1. Nature 2009;459:1126–1130.

20 Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature 2009;461:218–223.

21 Loscalzo J, Kohane I, Barabasi AL. Human disease classification in the postgenomic era: A complex systems approach to human pathobiology. Mol Syst Biol 2007;3:124.

22 Schadt EE, Lum PY. Thematic review series: Systems biology approaches to metabolic and cardiovascular disorders. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. J Lipid Res 2006;47:2601–2613.

23 Eichler EE, Flint J, Gibson G et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 2010;11:446–450.

24 Mao R, Zielke CL, Zielke HR et al. Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain. Genomics 2003;81:457–467.

25 FitzPatrick DR, Ramsay J, McGill NI et al. Transcriptome analysis of human autosomal trisomy. Hum Mol Genet 2002;11:3249–3256.

26 Lockstone HE, Harris LW, Swatton JE et al. Gene expression profiling in the adult Down syndrome brain. Genomics 2007;90:647–660.

27 Aït Yahya-Graison E, Aubert J, Dauphinot L et al. Classification of human chromosome 21 gene-expression variations in Down syndrome: Impact on disease phenotypes. Am J Hum Genet 2007;81:475–491.

28 Kahlem P, Sultan M, Herwig R et al. Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of down syndrome. Genome Res 2004;14:1258–1267.

29 Mao R, Wang X, Spitznagel EL, Jr., et al. Primary and secondary transcriptional effects in the developing human Down syndrome brain and heart. Genome Biol 2005;6:R107.

30 Antonarakis SE, Lyle R, Dermitzakis ET et al. Chromosome 21 and down syndrome: From genomics to pathophysiology. Nat Rev Genet 2004;5:725–738.

31 Saran NG, Pletcher MT, Natale JE et al. Global disruption of the cerebellar transcriptome in a Down syndrome mouse model. Hum Mol Genet 2003;12:2013–2019.

32 Haydar TF, Reeves RH. Trisomy 21 and early brain development. Trends Neurosci 2012;35:81–91.

33 Reeves RH. Down syndrome mouse models are looking up. Trends Mol Med 2006;12:237–240.

34 Devoy A, Bunton-Stasyshyn RK, Tybulewicz VL et al. Genomically humanized mice: Technologies and promises. Nat Rev Genet 2011;13:14–20.

35 Watase K, Zoghbi HY. Modelling brain diseases in mice: The challenges of design and analysis. Nat Rev Genet 2003;4:296–307.

36 Yu T, Li Z, Jia Z et al. A mouse model of Down syndrome trisomic for all human chromosome 21 syntenic regions. Hum Mol Genet 2010;19:2780–2791.

37 O'Doherty A, Ruf S, Mulligan C et al. An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. Science 2005;309:2033–2037.

38 Gardiner K, Fortna A, Bechtel L et al. Mouse models of Down syndrome: How useful can they be? Comparison of the gene content of human chromosome 21 with orthologous mouse genomic regions. Gene 2003;318:137–147.

39 Waterston RH, Lindblad-Toh K, Birney E et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520–562.

40 Zhu H, Lensch MW, Cahan P et al. Investigating monogenic and complex diseases with pluripotent stem cells. Nat Rev Genet 2011;12:266–275.

41 Grskovic M, Javaherian A, Strulovici B et al. Induced pluripotent stem cells: Opportunities for disease modelling and drug discovery. Nat Rev Drug Discov 2011;10:915–929.

42 Irion S, Nostro MC, Kattman SJ et al. Directed differentiation of pluripotent stem cells: From developmental biology to therapeutic applications. Cold Spring Harb Symp Quant Biol 2008;73:101–110.

43 Laslett AL, Grimmond S, Gardiner B et al. Transcriptional analysis of early lineage commitment in human embryonic stem cells. BMC Dev Biol 2007;7:12.

44 Park IH, Arora N, Huo H et al. Disease-specific induced pluripotent stem cells. Cell 2008;134:877–886.

45 Marchetto MC, Carromeu C, Acab A et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. Cell 2010;143:527–539.

46 Kim KY, Hysolli E, Park IH. Neuronal maturation defect in induced pluripotent stem cells from patients with Rett syndrome. Proc Natl Acad Sci USA 2011;108:14169–14174.

47 Chailangkarn T, Acab A, Muotri AR. Modeling neurodevelopmental disorders using human neurons. Curr Opin Neurobiol 2012;22:785–790.

48 Muotri AR, Marchetto MC, Coufal NG et al. L1 retrotransposition in neurons is modulated by MeCP2. Nature 2010;468:443–446.

49 Briggs JA, Sun J, Shepherd J et al. Integration-free induced pluripotent stem cells model genetic and neural developmental features of Down syndrome etiology. STEM CELLS 2013 (in press).

50 Shi Y, Kirwan P, Smith J et al. A human stem cell model of early Alzheimer's disease pathology in Down syndrome. Sci Transl Med 2012;4:124ra129.

51 Brennand KJ, Simone A, Jou J et al. Modelling schizophrenia using human induced pluripotent stem cells. Nature 2011;473:221–225.

52 Osafune K, Caron L, Borowiak M et al. Marked differences in differentiation propensity among human embryonic stem cell lines. Nat Biotechnol 2008;26:313–315.

53 Lister R, Pelizzola M, Kida YS et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 2011;471:68–73.

54 Ohm JE, Mali P, Van Neste L et al. Cancer-related epigenome changes associated with reprogramming to induced pluripotent stem cells. Cancer Res 2010;70:7662–7673.

55 Miura K, Okada Y, Aoi T et al. Variation in the safety of induced pluripotent stem cell lines. Nat Biotechnol 2009;27:743–745.

56 Hussein SM, Batada NN, Vuoristo S et al. Copy number variation and selection during reprogramming to pluripotency. Nature 2011; 471:58–62.

57 Laurent LC, Ulitsky I, Slavin I et al. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. Cell Stem Cell 2011;8:106–118.

58 Polo JM, Liu S, Figueroa ME et al. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol 2010;28: 848–855.

59 Bar-Nur O, Russ HA, Efrat S et al. Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells. Cell Stem Cell 2011;9:17–23.

60 Ohi Y, Qin H, Hong C et al. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat Cell Biol 2011;13:541–549.

61 Kim K, Doi A, Wen B et al. Epigenetic memory in induced pluripotent stem cells. Nature 2010;467:285–290.

62 Cheng L, Hansen NF, Zhao L et al. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. Cell Stem Cell 2012;10:337–344.

63 Schena M, Shalon D, Davis RW et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467–470.

64 Eisen MB, Spellman PT, Brown PO et al. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–14868.

65 Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63.

66 Mortazavi A, Williams BA, McCue K et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 2008;5: 621–628.

67 Kuhn DE, Nuovo GJ, Terry AV, Jr., et al. Chromosome 21-derived microRNAs provide an etiological basis for aberrant protein expression in human Down syndrome brains. J Biol Chem 2010;285:1529–1543.

68 Dauphinot L, Lyle R, Rivals I et al. The cerebellar transcriptome during postnatal development of the Ts1Cje mouse, a segmental trisomy model for Down syndrome. Hum Mol Genet 2005;14:373–384.

69 Ramanan VK, Shen L, Moore JH et al. Pathway analysis of genomic data: Concepts, methods, and prospects for future development. Trends Genet 2012;28:323–332.

70 Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–15550.

71 Duke DC, Moran LB, Kalaitzakis ME et al. Transcriptome analysis reveals link between proteasomal and mitochondrial pathways in Parkinson's disease. Neurogenetics 2006;7: 139–148.

72 Sutherland GT, Matigian NA, Chalk AM et al. A cross-study transcriptional analysis of Parkinson's disease. PLoS One 2009;4:e4955.

73 Mar JC, Matigian NA, Quackenbush J et al. attract: A method for identifying core pathways that define cellular phenotypes. PLoS One 2011;6:e25445.

74 Mar JC, Matigian NA, Mackay-Sim A et al. Variance of gene expression identifies altered network constraints in neurological disease. PLoS Genet 2011;7:e1002207.

75 Khatri P, Draghici S. Ontological analysis of gene expression data: Current tools, limitations, and open problems. Bioinformatics 2005;21:3587–3595.

76 Venet D, Dumont JE, Detours V. Most random gene expression signatures are significantly associated with breast cancer outcome. PLoS Comput Biol 2011;7:e1002240.

77 Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: Insights into functions. Nat Rev Genet 2009;10:155–159.

78 Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet 2006;15 Spec No 1:R17–R29.

79 Tompa M, Li N, Bailey TL et al. Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 2005;23:137–144.

80 Laird PW. Principles and challenges of genomewide DNA methylation analysis. Nat Rev Genet 2010;11:191–203.

81 Schones DE, Zhao K. Genome-wide approaches to studying chromatin modifications. Nat Rev Genet 2008;9:179–191.

82 Meissner A, Mikkelsen TS, Gu H et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 2008; 454:766–770.

83 Mikkelsen TS, Ku M, Jaffe DB et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:553–560.

84 Luo D, Saltzman WM. Synthetic DNA delivery systems. Nat Biotechnol 2000;18:33–37.

85 Rubinson DA, Dillon CP, Kwiatkowski AV et al. A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. Nat Genet 2003;33:401–406.

86 Mäe M, Langel U. Cell-penetrating peptides as vectors for peptide, protein and oligonucleotide delivery. Curr Opin Pharmacol 2006;6:509–514.

87 Chu TC, Twu KY, Ellington AD et al. Aptamer mediated siRNA delivery. Nucleic Acids Res 2006;34:e73.

88 Shen C, Buck AK, Liu X et al. Gene silencing by adenovirus-delivered siRNA. FEBS Lett 2003;539:111–114.

89 Saez E, No D, West A et al. Inducible gene expression in mammalian cells and transgenic mice. Curr Opin Biotechnol 1997;8:608–616.

90 Sanjana NE, Cong L, Zhou Y et al. A transcription activator-like effector toolbox for genome engineering. Nat Protoc 2012;7:171–192.

91 Bogdanove AJ, Voytas DF. TAL effectors: Customizable proteins for DNA targeting. Science 2011;333:1843–1846.

92 Khalil AS, Collins JJ. Synthetic biology: Applications come of age. Nat Rev Genet 2010; 11:367–379.

93 Culler SJ, Hoff KG, Smolke CD. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. Science 2010;330:1251–1255.

94 Lamb J, Crawford ED, Peck D et al. The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 2006;313:1929–1935.

95 Park PJ. ChIP-seq: Advantages and challenges of a maturing technology. Nat Rev Genet 2009;10:669–680.

96 Valouev A, Johnson DS, Sundquist A et al. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 2008;5:829–834.

97 Chu C, Qu K, Zhong FL et al. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 2011;44:667–678.

98 Simon MD, Wang CI, Kharchenko PV et al. The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci USA 2011;108:20497–20502.

99 Hockemeyer D, Wang H, Kiani S et al. Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol 2011;29:731–734.

100 Ramírez-Solis R, Liu P, Bradley A. Chromosome engineering in mice. Nature 1995; 378:720–724.

101 van der Weyden L, Shaw-Smith C, Bradley A. Chromosome engineering in ES cells. Methods Mol Biol 2009;530:49–77.

102 Nuzhdin SV, Friesen ML, McIntyre LM. Genotype-phenotype mapping in a post-GWAS world. Trends Genet 2012;28:421–426.

103 Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol 2007;8:995–1005.

104 Ghaemmaghami S, Huh WK, Bower K et al. Global analysis of protein expression in yeast. Nature 2003;425:737–741.

105 Huh WK, Falvo JV, Gerke LC et al. Global analysis of protein localization in budding yeast. Nature 2003;425:686–691.

106 Kislinger T, Cox B, Kannan A et al. Global survey of organ and organelle protein expression in mouse: Combined proteomic and transcriptomic profiling. Cell 2006;125:173–186.

107 Herrgård MJ, Swainston N, Dobson P et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 2008;26:1155–1160.

108 Burgard AP, Nikolaev EV, Schilling CH et al. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res 2004;14:301–312.

109 Sacco F, Gherardini PF, Paoluzi S et al. Mapping the human phosphatome on growth pathways. Mol Syst Biol 2012;8:603.

110 Minguez P, Parca L, Diella F et al. Deciphering a global network of functionally associated post-translational modifications. Mol Syst Biol 2012;8:599.

111 Yang B, Wu YJ, Zhu M et al. Identification of cross-linked peptides from complex samples. Nat Methods 2012;9:904–906.

112 Walzthoeni T, Claassen M, Leitner A et al. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. Nat Methods 2012;9:901–903.

113 Kristensen AR, Gsponer J, Foster LJ. A high-throughput approach for measuring temporal changes in the interactome. Nat Methods 2012;9:907–909.

114 Tkach JM, Yimit A, Lee AY et al. Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. Nat Cell Biol 2012;14:966–976.

115 Pontén F, Gry M, Fagerberg L et al. A global view of protein expression in human cells, tissues, and organs. Mol Syst Biol 2009;5:337.

116 Beck M, Schmidt A, Malmstroem J et al. The quantitative proteome of a human cell line. Mol Syst Biol 2011;7:549.

117 Munoz J, Low TY, Kok YJ et al. The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. Mol Syst Biol 2011;7:550.

118 Havugimana PC, Hart GT, Nepusz T et al. A census of human soluble protein complexes. Cell 2012;150:1068–1081.

119 Rual JF, Venkatesan K, Hao T et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature 2005;437:1173–1178.

120 Ideker T, Krogan NJ. Differential network biology. Mol Syst Biol 2012;8:565.

121 Parmigiani G, Garrett ES, Irizarry RA et al. The analysis of gene expression data: An overview of methods and software. In: Parmigiani G, Garrett ES, Irizarry RA et al., eds. The Analysis of Gene Expression Data: An Overview of Methods and Software. New York, NY: Springer, 2003:1–45.

122 Churchill G. Fundamentals of experimental design for cDNA microarray. Nat Genet 2002;32:490–495.

123 Slonim D. From patterns to pathways: Gene expression data analysis comes of age. Nat Genet 2002;32:502–508.

124 Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article3.

125 Tusher V, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;98:5116–5121.

126 Ringner M. What is principal component analysis? Nat Biotechnol 2008;26:1–2.

127 Jolliffe IT. Principal Component Analysis. New York, NY: Springer, 2002.

128 Kanehisa M. The KEGG database. Novartis Found Symp 2002;247:244–252.

129 Saeed AI, Bhagabati NK, Braisted JC et al. TM4 microarray software suite. Methods Enzymol 2006;411:134–193.

130 Saeed AI, Sharov V, White J et al. TM4: A free, open-source system for microarray data management and analysis. Biotechniques 2003;34:374–378.

131 Reich M, Liefeld T, Gould J et al. GenePattern 2.0. Nat Genet 2006;38:500–501.

132 Du P, Kibbe WA, Lin SM. lumi: A pipeline for processing Illumina microarray. Bioinformatics 2008;24:1547–1548.

133 Sturn A, Quackenbush J, Trajanoski Z. Genesis: Cluster analysis of microarray data. Bioinformatics 2002;18:207–208.

134 Ulitsky I, Shamir R. Identification of functional modules using network topology and high-throughput data. BMC Syst Biol 2007;1:8.

135 Doniger SW, Salomonis N, Dahlquist KD et al. MAPPFinder: Using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. Genome Biol 2003;4:R7.

136 Salomonis, Hanspers K, Zambon A et al. GENMAPP 2: New features and resources for pathway analysis. BMC Bioinformatics 2007;8:217.

137 Dennis G, Jr., Sherman BT, Hosack DA et al. DAVID: Database for Annotation, Visualisation, and Integrated Discovery. Genome Biol 2003;4:P3.

138 Thomas PD, Campbell MJ, Kejariwal A et al. PANTHER: A library of protein families and subfamilies indexed by function. Genome Res 2003;13:2129–2141.

139 Shannon P, Markiel A, Ozier O et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–2504.

140 Yeung N, Cline MS, Kuchinsky A et al. Exploring biological networks with Cytoscape software. Curr Protoc Bioinformatics 2008; Chapter 8:Unit 8.13.

141 Wu LF, Hughes TR, Davierwala AP et al. Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. Nat Genet 2002;31:255–265.

142 Allocco DJ, Kohane IS, Butte AJ. Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 2004;5:18.

143 Golub TR, Slonim DK, Tamayo P et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 1999;286:531–537.

144 Storey JD, Tibshirani R. SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In: Parmigiani G, Garrett ES, Irizarry RA et al., eds. The Analysis of Gene Expression Data: Methods and Software. New York, NY: Springer, 2003:272–290.

145 MAQC Consortium; Shi L, Reid LH, Jones WD et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 2006;24:1151–1162.