MENTAL RETARDATION AND DEVELOPMENTAL DISABILITIES
RESEARCH REVIEWS *13: 199–206 (2007)*

GENETIC MECHANISMS INVOLVED IN THE PHENOTYPE OF DOWN SYNDROME

David Patterson[1,2]*

[1]Eleanor Roosevelt Institute, University of Denver, Denver, Colorado

[2]Department of Biological Sciences, University of Denver, Denver, Colorado

Down syndrome (DS) is the most common genetic cause of significant intellectual disability in the human population, occurring in roughly 1 in 700 live births. The ultimate cause of DS is trisomy of all or part of the set of genes located on chromosome 21. How this trisomy leads to the phenotype of DS is unclear. The completion of the DNA sequencing and annotation of the long arm of chromosome 21 was a critical step towards understanding the genetics of the phenotype. However, annotation of the chromosome continues and the functions of many genes on chromosome 21 remain uncertain. Recent findings about the structure of the human genome and of chromosome 21, in particular, and studies on mechanisms of gene regulation indicate that various genetic mechanisms may be contributors to the phenotype of DS and to the variability of the phenotype. These include variability of gene expression, the activity of transcription factors both encoded on chromosome 21 and encoded elsewhere in the genome, copy number polymorphisms, the function of conserved nongenic regions, microRNA activities, RNA editing, and perhaps DNA methylation. In this manuscript, we describe current knowledge about these genetic complexities and their likely importance in the context of DS. We identify gaps in current knowledge and suggest priorities to fill these gaps.
MRDD Research Reviews 2007;13:199–206.

Key Words: Down syndrome; gene regulation; genome structure and function; genotype–phenotype correlations

THE CHROMOSOMAL ABNORMALITY LEADING TO DOWN SYNDROME

Down syndrome (DS), or Trisomy 21, is the most common genetic cause of significant intellectual disability, occurring in ~1 in 733 live births [Centers for Disease Control and Prevention, 2006]. Trisomy of the entire chromosome is present in about 95% of cases. In a small fraction of cases, trisomy of part of chromosome 21 (HSA21) leads to DS.

THE PHENOTYPIC VARIABILITY OF DS

Certain aspects of the phenotype appear to occur in every person with DS, whereas other traits are more highly variable [Antonarakis et al., 2004]. Even for phenotypes that occur in every individual, there is variability in expression. For example, cognitive impairment is ubiquitous but ranges from mild to moderate. Congenital heart defects, by contrast, occur in only 40% of those with the disorder, and its severity also varies. Certain other conditions such as duodenal atresia, Hirschsprung disease, and acute megakaryocytic leukemia are increased in DS compared to the general population. To some extent this variability reflects the variability seen in the human population. It is not clear whether the population with DS shows a wider variability than that seen in the population as a whole. If the variability is greater, it is not clear whether there are specific genes, either on HSA21 or elsewhere in the genome that cause this increased variability.

THE GENES ON HSA21 AND THE PHENOTYPE OF DS

An essential step in understanding how the genes on HSA21 lead to DS was the publication of the virtually complete DNA sequence of the long arm of HSA21 (21q) and the annotation of this sequence [Hattori et al., 2000]. Annotation of HSA21 is an ongoing process. Currently about 430 genes and gene models have been identified (http://chr21db.cudenver.edu/). An important point here is that the functions of about 45% of the genes on HSA21 are uncertain [Kahlem, 2006]. In addition, as discussed later, the sequence of HSA21, and indeed the entire human genome, has revealed unexpected complexity in genome structure and in gene regulation that will have profound implications for understanding genotype/phenotype relationships. Each of these will be discussed, and examples based on experiments relevant to HSA21 specifically will be used to elucidate the significance of these observations.

Two primary hypotheses exist regarding the relationship of the genes on HSA21 to the phenotype. The first, the gene dosage hypothesis, is that elevated expression of specific genes on HSA21 leads to specific aspects of the phenotype [Korenberg et al., 1990]. This hypothesis does not make any predictions regarding how many or which genes on HSA21 may be involved in any specific aspect of the phenotype, but does predict that specific genes or subsets of genes can be related to specific aspects of DS. The second hypothesis is the genetic homeostasis hypothesis. This hypothesis states that it is expression

Grant sponsor: NINDS; Grant number: NS51539; Grant sponsor: The Itkin Family Foundation.
*Correspondence to: David Patterson, Ph.D., Eleanor Roosevelt Institute, 1899 Gaylord St., Denver, CO 80206. E-mail: dpatter2@du.edu
Received 10 August 2007; Accepted 13 August 2007
Published online in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/mrdd.20162

of extra genes, regardless of their identity or function that leads to phenotypic alterations [Shapiro, 1983]. Clearly, numerous combinations of these hypotheses are possible. For example, it could be that overexpression of a set of genes on HSA21 could have a direct connection to some aspects of the phenotype of DS, while expression of the extra genes could in itself also disturb normal development. Recent experiments favor a modified gene dosage hypothesis. Some genes, for example, the amyloid precursor protein (*APP*) gene, may indeed play a significant role in defining the phenotype of DS [Salehi et al., 2006]. Moreover, some regions of HSA21 may be relevant for some features of DS [Ronan et al., 2007]. Experiments with mouse models of DS show that the region of HSA21 commonly referred to as the DS Critical Region (DSCR) may be necessary, but not sufficient, for appearance of some phenotypes of DS [Olson et al., 2004a,b; Olson et al., 2007].

LEVELS OF GENE EXPRESSION IN DS

The simplest model for levels of gene expression in DS would predict that each gene on chromosome 21 would be expressed at 150% of the levels seen in euploid individuals. A number of investigators have examined this question both in tissues or cell lines from individuals with DS and in mice trisomic for regions of Mmu16 homologous to HSA21. Several methods have been used, including DNA microarray analysis, serial analysis of gene expression (SAGE), and, less commonly, qRT-PCR or proteomic approaches. These approaches have the advantage that in principle large numbers of genes can be analyzed, but they may lack discriminatory power. In general, the outcomes of these studies seem to support the hypothesis that many, but not all, HSA21 genes are expressed at the expected 150% of normal, although there are numerous exceptions, including genes on HSA21 that are not overexpressed or overexpressed by more than 150% and overexpression of genes on other chromosomes. In some studies, only overall global chromosome 21 gene expression could be concluded to be overexpressed. That is, the set of HSA21 genes taken together is overexpressed by ~150%, but overexpression of individual genes could not be assessed in a statistically significant way [Mao et al., 2003].

In a some cases, expression levels of individual HSA21 genes appeared to differ in different tissues and cells [Li et al., 2006]. In some tissues, HSA21 genes are overexpressed while in others these same genes are not be overexpressed. A similar situation exists in the Ts65Dn mouse model of DS. In this case, expression of the trisomic gene GABPA protein is elevated only in brain and skeletal muscle [O'Leary et al., 2004]. Clearly, an answer to whether or not each HSA21 gene is overexpressed, to what extent, in what tissues, and during what developmental stage is an exceedingly important information to have when attempting genotype–phenotype correlations.

Variability of Gene Expression as a Factor in the DS Phenotype

Antonarakis et al. hypothesize that in any human population there will be variation in the expression levels of many genes [Antonarakis et al., 2004; Deutsch et al., 2005]. Further, they argue that if the variation in expression in the euploid population of a gene on HSA21 exceeds the 1.5-fold expression elevation expected by the gene dosage hypothesis, then this gene may not be related to the phenotype or may only be related to the phenotypes of DS that are not universally present. However, if the expression level of a particular HSA21 gene is tightly regulated in the euploid population, then a change in expression of 1.5-fold might be directly related to the more consistently present phenotypes of DS, for example, intellectual disability. They present evidence that variation of expression levels of a significant fraction of genes on HSA21 is indeed greater than 2-fold in the euploid population, raising the possibility that these genes may not be causal for the more universally present features of DS [Deutsch et al., 2005]. Stranger et al. [2005], using HSA21 (and a region of chromosome 20) as a representative genomic region, presented evidence for widespread gene expression variability. Importantly, they also concluded that most of the gene expression variability is located in the 1 Mb region proximal to the gene of interest.

. . . the sequence of HSA21, and indeed the entire human genome, has revealed unexpected complexity in genome structure and in gene regulation that will have profound implications for understanding genotype/phenotype relationships.

MECHANISMS OF VARIABILITY IN GENE EXPRESSION

Copy Number Variations

Copy number variations (CNVs) can be defined as regions of the genome, often several kb or larger, that vary in copy number from individual to individual [Iafrate et al., 2004; Sebat et al., 2004; Hegele, 2007]. Often, these CNVs contain genes. A recent study presents evidence that duplication of a region of HSA21 including the *APP* gene can lead to familial Alzheimer's disease (AD) in five separate families. In these families, the duplication ranged from 0.58 to over 6 Mb [Rovelet-Lecrux et al., 2006]. Importantly, a detailed phenotypic analysis of the five families in which this duplication was observed demonstrated that there was no clinical evidence of DS (other than the appearance of AD) [Cabrejo et al., 2006]. This observation is consistent with the idea that trisomy of *APP* may explain, at least in part, the observation that persons with DS virtually always develop the neuropathology associated with AD, but that trisomy of *APP* is not sufficient to cause DS. However, there are four other genes in the smallest duplicated region reported in these families, and the level of gene expression was not reported. Thus, the conclusion that *APP* is the responsible gene is attractive, but not yet definitively proven.

A recent report indicates that duplication of a 4.3 Mb region of HSA21 containing slightly more than 30 genes causes a DS phenotype in three family members [Ronan et al., 2007]. These family members had the facial gestalt of DS but had mild cognitive disability. Again, however, no measurement of the level of overexpression of the genes in the duplicated region was reported. Moreover, two genes hypothesized to be involved in the phenotype of DS, *DSCR1* and *DSCAM*, were not included in the duplicated region. Thus, it appears that duplication

of this region, while contributing to the phenotype of DS, is probably not sufficient to cause all the features of DS. Although these data implicate CNVs as a reason for variation in gene expression, a recent study estimates that single nucleotide polymorphisms detect about 84% of gene expression variability while CNVs capture about 17% [Stranger et al., 2007]. This is consistent with comparisons of promoter mutations between human HSA21 genes and chimpanzee chromosome 22 (the chimpanzee chromosome homologous to HSA21) genes, which concluded that the vast majority of alterations in promoter regions were single nucleotide changes [Watanabe et al., 2004].

Transcription Factors and Variability of Expression

There are at least 25 genes on HSA21 that directly or indirectly regulate gene transcription [Gardiner, 2006]. Many if not most of these function as parts of multiprotein complexes. Alterations in the relative abundance of members of a protein complex may affect its activity in ways that are difficult to predict. Moreover, many genes are regulated by sets of transcription factors acting in concert. For example, the HSA21 gene encoding the reduced folate carrier (*RFC*), the major protein responsible for transport of reduced folates, shows marked tissue and developmental specificity of expression and response to environmental folate levels [Liu et al., 2005; Payton et al., 2007]. The *RFC* gene has up to five major 5′ untranslated regions [Payton et al., 2007]. Each of these 5′ UTRs appears to be regulated by a different cassette of transcription factors [Whetstine and Matherly, 2001; Payton et al., 2005; Liu et al., 2006]. So far HSA21 encoded transcription factors have not been implicated directly in regulation of *RFC*. However, in many cases, genes on HSA21 modulate (or have been hypothesized to modulate) transcription factors that do regulate transcription of *RFC* [Whetstine and Matherly, 2001; Payton et al., 2005; Gardiner, 2006; Liu et al., 2006]. There are up to 14 *RFC* transcripts for which different 5′ UTRs are fused to a common coding sequence. Transfection studies into HeLa cells of individual 5′ UTR *RFC* constructs demonstrates a range of steady-state reduced folate carrier proteins and transcripts that seem to be functions of relative transcript stability and translation efficiency [Payton et al., 2007].

This regulatory system illustrates a number of key points. It is a concrete example of the complexity of regulation of genes on HSA21 both by other genes on HSA21 and by genes on other chromosomes that interact with these HSA21 genes. It demonstrates the complexity of regulation at both the transcriptional and posttranscriptional level. It demonstrates the complex nature of tissue specificity of gene expression. Also, there is the important point that expression of *RFC* responds to environmental levels of folate.

Measurements of gene expression by assessment of mRNA levels do not always correspond to equivalent alterations in cognate protein levels. Moreover, elevations of protein levels must result in significant alterations in metabolic, structural, or other functional activities, or they will not result in phenotypic alterations. Thus, the measurements of alterations along the continuum from genotype to phenotype are likely to be more relevant the closer they are on this continuum to the ultimate phenotypic consequences.

An intriguing observation is that the transcription factor GATA1 may regulate the A1/A2 *RFC* promoter [Liu et al., 2006]. Mutations in this transcription factor are essentially always found in persons with DS who develop a transient leukemoid disorder (TMD) and/or acute megakaryocytic leukemia (AMKL), a form of leukemia ~500-fold more common in individuals with DS than in children without DS. Individuals with AMKL without DS in general do not have these mutations in GATA1 [Vyas and Crispino, 2007]. Although GATA1 is not encoded on HSA21, it does apparently affect expression of the HSA21 encoded transcription factor BACH1 [Vyas and Crispino, 2007]. AMKL associated with DS shows increased sensitivity to chemotherapy, which may be related to altered transcription of enzymes that metabolize the drugs cytosine arabinoside and daunorubicin [Muntean et al., 2006]. Thus, the concept is developing that occurrence of AMKL and/or TMD in persons with DS may require three genetic events, mutations in GATA1, trisomy of chromosome 21, and an as yet unidentified third event. Trisomy of HSA21 may give a selective advantage to hematopoietic cells with a GATA1 mutation [Vyas and Crispino, 2007]. The genetic mechanism by which this occurs remains unknown.

Conserved Noncoding Regions

As the essentially complete genomic sequences of mammals in addition to humans have become available, genomic sequence comparisons have revealed the existence of a large number of conserved non-coding regions [Dermitzakis et al., 2005]. Conserved noncoding regions (CNCs) are not repetitive sequences and the vast majority are not transcribed [Dermitzakis et al., 2005]. HSA21 was the first chromosome completely analyzed for CNCs, and 2,262 were identified. Possible functions for CNCs include roles as cis- or trans-regulatory regions or elements required for chromatin or chromosome structure [Dermitzakis et al., 2004].

POST-TRANSCRIPTIONAL REGULATION

As the discussion above demonstrates, assessments of transcript levels are crucial for understanding DS and other conditions. However, post-transcriptional regulatory mechanisms clearly play important roles, since transcript levels do not always accurately reflect the levels of the proteins they encode. Some unexpected mechanisms of posttranscriptional regulation have been revealed by functional genomics studies.

MicroRNAs

MicroRNAs (MiRNAs) are endogenous small RNA molecules of about 22 nt derived from larger transcripts

[Bartel, 2004; Kosik and Krichevsky, 2005]. Hundreds of miRNAs have been identified [Nam et al., 2005]. They bind to the 3′ UTRs of mRNA molecules and can interfere with mRNA translation or subject mRNA to degradation [Bartel, 2004; Bartel and Chen, 2004; Fahr et al., 2005; Maas et al., 2006]. An intriguing hypothesis is that miRNAs serve to dampen translation of mRNAs rather than have an all or none effect. In this way, they may be able to exert more rapid and subtle modifications of protein levels than transcriptional control mechanisms [Bartel and Chen, 2004; Kosik and Krichevsky, 2005]. At least five putative miRNAs are encoded by genes on HSA21.

Recent work shows that miRNAs constitute up to 5% of human genes [Niwa and Slack, 2007]. Each miRNA may contribute to the regulation of expression of hundreds of mRNAs. It appears that miRNAs may be particularly relevant to regulation of genes in the nervous system, genes that control expression of other genes, e.g., transcription factor genes, or genes that control signaling. There is evidence suggesting that this mechanism of gene regulation is relatively rare for genes that are widely expressed in many cell types, and that may be essential for cellular survival, for example, metabolic genes [Cui et al., 2007; Gaidatzis et al., 2007].

RNA Editing

RNA editing is a process in which adenosine in RNA is deaminated to inosine. Inosine acts as guanosine during mRNA translation, so editing can result in alterations in protein sequence. A to I editing has been associated with human disease [Maas et al., 2006; Pokharel and Beal, 2006]. RNA editing is apparently important for normal functioning of the nervous system [Pokharel and Beal, 2006]. The extent of RNA editing in humans is not known.

An interesting convergence of miRNA and RNA editing appears to be emerging, with the initial example involving HSA21. RNA editing is carried out by enzymes known as ADARs (A Deaminases that Act on RNA). ADAR2 is encoded by a gene on HSA21. Clearly, ADARs change the protein coding of mRNAs, but the A to I editing also changes the base pairing characteristics of RNA. It has been hypothesized that this editing could alter the binding of a miRNA to its mRNA target. Kawahara et al. [2007] examined the consequences of editing of the *miR-376* RNA by ADAR2. They found that one of the genes regulated by *miR-376* RNA encodes phosphoribosyl pyrophosphate synthetase 1 (PRPS1), an enzyme catalyzing a rate-limiting step of de novo purine nucleotide synthesis. Mice in which *ADAR2* had been silenced by targeted mutagenesis did not appropriately edit *miR-376* RNA. PRPS1 gene expression was elevated by a factor of 2 in a tissue specific manner in *ADAR2* null mice compared to normal control mice. Elevation occurred in brain but not in liver. The expression of *miR-476* RNA was similar in brain (cortex) and liver, but editing was detected in wild-type mouse cortex but not liver. These investigators extended these studies in a very important way. Because PRPS1 is a rate-limiting step for de novo purine synthesis, increased levels of this enzyme should increase purine synthesis. Indeed this was found. Uric acid is a catabolite of purines. In *ADAR2* null mice, urate levels in cortex were elevated by a factor of 2, whereas in liver, urate levels were equivalent in wild-type and *ADAR2* null mice.

This experiment allows a number of very important inferences. It demonstrates a metabolic consequence of a change in expression of an enzyme, PRPS1, controlled by editing of a miRNA. This is a crucial point, since unless alteration in the level of a protein results in a metabolic (or other functional) alteration, the protein alteration may be of no phenotypic consequence. Another interesting implication of this experiment is that it demonstrates that miRNA posttranslational regulation of gene activity may be a mechanism for precise regulation of gene activity. Importantly, persons with DS generally have urate levels of 150% of those seen in euploid individuals [Pant et al., 1968]. If ADAR2 levels are elevated in persons with DS in a dosage-dependent manner, one might expect decreased levels of urate. Clearly, miRNA regulation is of considerable significance, but it must work in concert with other mechanisms of regulation of the end consequences of trisomy.

DNA METHYLATION

DNA methylation is strictly speaking an epigenetic change, not a genetic change. However, it clearly plays a role in gene expression, and genes on HSA21 play a role in the phenomenon. Moreover, DNA methylation is critical for meiosis, and recent experiments suggest that it may play a major role in memory formation, so a brief discussion is included here.

Methylation of the 5′ position of cytosines located in cytosine-guanosine dinucleotides resulting in m^5C is the primary, if not the only, modification of mammalian DNA [Turek-Plewa and Jagodzinksi, 2005]. Generally, DNA methylation is thought to play a role in long-term silencing of gene expression, and to be critical for mammalian development [Turek-Plewa and Jagodzinksi 2005; Shames et al., 2007]. In some cases, tissue-specific DNA methylation may be important for tissue specificity of gene expression in mice [Song et al., 2005]. In cancer, global genome hypomethylation is often observed coupled with hypermethylation of specific genes, for example, tumor suppressor genes [Shames et al., 2007]. It has been hypothesized that altered dietary folates can affect DNA methylation, and correlations have been drawn between folate deficiency and increased risk for many types of cancer [Kim, 2005]. As discussed earlier, the gene encoding the reduced folate carrier, considered the primary protein responsible for cellular internalization of folates, is located on HSA21 [Moscow et al., 1995]. Thus, one could hypothesize that elevated levels of the reduced folate carrier protein might enhance cellular uptake of folates mitigating the effects of low folate, including the effects of low folate on DNA methylation. This might be related to the observed low incidence of solid tumors in individuals with DS [Patja et al., 2006]. Recently, demethylation of methylated miRNA resulting in its activation has been observed in cancer cells [Lujambio et al., 2007]. This finding implies a link between regulation of gene expression by DNA methylation and miRNA.

DNA methylation is carried out by a family of *DN*A *m*ethyl*t*ransferase (DNMT) enzymes. A member of this gene family, *DNMT3L*, is located on HSA21 [Aapola et al., 2000]. The DNMT3L protein has significant homology to the DNMT3 DNA methyltransferases, but no known DNA methyltransferase activity of its own. However, it interacts directly with DNMT3A. *DNMT3L* knockout mice are viable, but sterile. Study of these mice has shown that DNMT3L is required for establishment of maternal imprints during oogenesis. Moreover, it is necessary for normal spermatogenesis [Hata et al., 2006]. It is difficult to extrapolate from knockout mice to the possible effects of trisomy of the

DNMT3L gene in persons with DS. However, one could speculate that overexpression of this gene might alter meiotic DNA methylation and be related to aberrant chromosomal segregation and also to the male sterility seen in men with DS.

In a very provocative publication Miller and Sweatt [2007] present evidence that DNA methylation and demethylation dynamically regulate memory formation. These investigators demonstrate that fear conditioning is associated with upregulation of expression of *DNMT3A* and *DNMT3B* [Miller and Sweatt, 2007]. They present evidence that chemical inhibition of DNMT activity blocks contextual fear conditioning in mice and that contextual fear conditioning is associated with methylation and silencing of the memory suppressor gene protein phosphatase 1 (*PP1*) and with DNA demethylation and transcription of the memory promoting gene *REELIN* in the mouse hippocampus, a region of the brain affected in DS and required for this task in mice. Apparently, dynamic DNA methylation changes occur in the hippocampus of mice during training for fear conditioning. These results require that DNA methylation be a dynamic process, at least in the mouse hippocampus during memory formation [Miller and Sweatt, 2007]. If these observations are confirmed and extended, they will require a re-evaluation of DNA methylation as a relatively stable phenomenon and also will implicate DNA methylation/demethylation in memory formation, a process compromised in persons with DS. It is interesting to speculate that the DNMT3L protein, which interacts with DNMT3A, might be involved in this process.

INTERACTIONS BETWEEN GENES

Overexpression of individual genes cannot be considered in isolation when one considers phenotype–genotype correlations. Thus, recent evidence demonstrates that trisomy of two or more genes acts synergistically to affect phenotype. A well-developed example of this type of synergy is the accumulating evidence that elevated expression of the *DYRK1A* and *DSCR1* genes on HSA21 act synergistically to prevent the nuclear localization of the NFATc transcription factor [Arron et al., 2006]. Thus, trisomy of either of these genes individually may not be sufficient to cause this dysregulation, but trisomy of both is sufficient to cause dysregulation of this developmentally significant pathway.

MOUSE MODELS

The Rationale for Various Types of Mouse Models

The rationale for the use of mouse models to study the etiology of DS has been discussed at length recently [e.g., Patterson and Costa, 2005], and the reader is referred to the manuscripts by Drs. Mobley and Reeves in this issue for a discussion of these. It is worth stressing some points here. The intrinsic value of mouse models is that mice genetically and biochemically are quite similar to humans. This is important when one is attempting to determine the genetics and biochemistry of any human disease. It is relatively easy to manipulate the mouse genome in precise ways. Moreover, extensively validated tests of learning and memory in mice exist [Crawley, 2007]. Thus, it is possible to study these phenotypes in mice. Moreover, many experiments cannot be done in humans or in simpler systems like tissue culture.

Overexpression of individual genes cannot be considered in isolation when one considers phenotype–genotype correlations. Thus, recent evidence demonstrates that trisomy of two or more genes may act synergistically to affect phenotype.

Even for homologous genes, it may be that the functions of those genes are not quite the same in mice and humans, or that their regulation will be different in the two species. An example is folate/one-carbon/trans-sulfuration. It now appears that the tissue specificity of expression of critical genes in this biological system is different in mice and in humans [Whetstine et al., 2002; Liu et al., 2005; Butler et al., 2006]. This may imply that these pathways serve slightly different functions in humans and in mice. Another example is the purine metabolic pathway. In this case, the end products of purine catabolism are different. Humans accumulate uric acid as the end product of this pathway, whereas mice metabolize urate further. Individuals with DS have elevated urate levels of essentially 150% [Pant et al., 1968]. This suggests that a gene dosage effect may be responsible. In particular, a gene encoding an enzyme carrying out three steps of de novo purine synthesis, *GART* (*phosphoribosylglycineamide transformylase*), is on HSA21, and trisomy of this gene has been hypothesized to be responsible for urate over accumulation in people with DS [Brodsky et al., 1997]. However, other genes on HSA21 may play a role here, for example, *ADAR2*, as discussed earlier. Therefore, results from studying the mice must eventually be validated for relevance to humans.

FUTURE DIRECTIONS

Defining the Function of All Genes on Chromosome 21

Although the genetic basis for the phenotype of DS is complex, the fact remains that a critical step is the identification of the function of the proteins encoded by all the genes on HSA21. One approach which should be pursued vigorously is the production of mice in which the genes of unknown function are disrupted by targeted mutagenesis, so called knockout mice. In fact, there are world-wide efforts underway to produce knockout mice for all mouse genes and to produce a library of mouse embryonic stem cells in which each mouse gene has been knocked out so that individual investigators can produce knockout mice in genes of interest [Austin et al., 2004; Nord et al., 2006]. These efforts should be encouraged to consider genes on HSA21 as a high priority.

The production of additional transgenic mice may help in this regard, but, as many investigators have pointed out, mice transgenic for individual genes may not accurately reflect the effects of trisomy of these genes in the context of trisomy of many other genes. Some investigators have suggested the production of mice transgenic for all members of a suspected biological pathway as an alternative to single gene transgenics, or at least production of transgenic animals that are transgenic for all of the genes of a particular pathway that are located on HSA21

[Gardiner, 2006; Gardiner and Costa, 2006]. In fact, even identification of which metabolic/biological pathways particular proteins participate in is a challenge. Thus, studies underway to determine protein–protein interactions (the interactome) must be continued and intensified [e.g., Lage et al., 2007].The choice of what type of transgenic mouse is most informative is not yet settled and at least to some extent depends on the experimental goals. For example, an issue that needs to be considered is whether or not it is more appropriate to produce a mouse trisomic for a mouse gene or into which an extra copy of the human gene has been inserted. Another consideration is whether the gene should be under the control of its endogenous control elements. This is likely to be the most relevant to DS, but it may be difficult to discern an effect of the transgene.

Mice trisomic for segments of the mouse genome homologous to segments of human HSA21 have been exceedingly useful both in elucidating specific phenotypes and in testing fundamental genetic hypotheses about DS, for example, the function of the so-called DSCR. These studies must continue and expand. In particular, mice trisomic for chromosomes homologous to regions of HSA21 not present in existing mouse models need to be produced. Efforts are underway in this regard, and success in producing mice trisomic for the HSA21 syntenic region on mouse chromosome 16 has recently been reported [Li et al., 2007]. These mice have cardiovascular and gastrointestinal abnormalities similar to those seen in individuals with DS and not seen in other mouse models of DS. Construction of mice syntenic for the mouse chromosome 10 and 17 regions is urgently needed.

After years of intense effort, a mouse strain that is trisomic for almost all of human HSA21 has been produced [O'Doherty et al., 2005]. So far, the mice reported in the literature are all chimeric. Thus, two problems need to be addressed for these mice to be optimally useful, the chimerism and the missing segment of HSA21. It is not clear whether trisomy of this region of HSA21 is incompatible with viability or whether a nonchimeric mouse would be viable. Comparison of these mice with mice trisomic for the homologous regions of mouse chromosomes should be undertaken. Such studies will be important to determine which mouse models are best suited for any particular experimental goal and will aid in the interpretation of results.

Determination of Which HSA21 Genes are Overexpressed, to What Extent, and in Which Tissues and Developmental Stages

Once the functions of HSA21 genes are defined, obtaining information about the level, tissue specificity, and developmental specificity of expression of these genes is essential. Studies are urgently needed to obtain this information. The technology to obtain this information exists, but it is often expensive, and difficult to accomplish in humans because of the lack of availability of appropriate tissue samples. Development of cell and tissue banks from individuals with DS and their families would aid this and many other studies of DS immensely.

Gene Expression Variability

The true extent of gene expression variability has not yet been determined, although available evidence suggests that it is quite significant. The fact that such significant variability in gene expression can be observed in the euploid human population demonstrates the complexity of drawing conclusions based on overexpression of trisomic genes. Also, the issue of the relevance of measurements of variability of gene expression needs to be carefully assessed for each gene. For example, a gene reported to have very high (perhaps 40-fold) variability in expression from individual to individual, *APP*, has also been strongly implicated in DS [Deutsch et al., 2005; Rovelet-Lecrux et al., 2006; Salehi et al., 2006]. Clearly, the overall extent of gene expression variability needs to be ascertained.

Definition of the True Extent and Significance of CNVs, CNCs, and Other Alterations in Genome Structure, for Example, Methylation

The functional significance of the unexpected variation in genome structure and the significance of conserved noncoding regions must be assessed to fully understand how the DS phenotype arises [Antonarakis and Epstein, 2006]. Moreover, it would seem that studies reporting phenotypic effects of copy number duplications or deletions should report also on whether alterations in gene expression of the duplicated or deleted genes occur. This problem is, of course, not unique to HSA21. The Encyclopedia of DNA Elements (ENCODE) Project is designed to identify all functional DNA elements of the human genome [ENCODE Project Consortium, 2004; Thomas et al., 2006]. A pilot phase involves the analysis of 1% of the human genome. About 2.2 Mb of the initial target DNA is located on HSA21 [Antonarakis and Epstein, 2006]. This project should provide key information regarding the significance of various structural and sequence variations in the DNA of HSA21. This pilot study is already completed [Giardine et al., 2007].

The possible role of DNA methylation as a dynamic process that influences memory formation in ways possibly relevant to DS needs to be explored. Fortunately, methods are being developed to assess DNA methylation on a genome-wide basis, and some of these are being applied to analysis of HSA21 [Yamada et al., 2004; Schumacher et al., 2006].

Tissue and Developmental Stage Variability

Many genes are apparently expressed in elevated levels and essentially according to gene dosage in DS and in mouse models of trisomy. However, tissue specificity and temporal specificity of altered expression has been observed [O'Leary et al., 2004; Rachidi et al., 2005; Butler et al., 2006; Li et al., 2006]. The relevance of this specificity has been discussed eloquently in recent reviews [Antonarakis and Epstein, 2006; Roper and Reeves, 2006]. The extent of this specificity needs to be determined. Fortunately, efforts are underway in this regard [Gitton et al., 2002; Reymond et al., 2002]. It may be that overexpression of the same gene can have different consequences at different developmental stages, and that during some developmental stages overexpression of a particular gene may cause developmental abnormalities while overexpression of that same gene at a different developmental stage may compensate for developmental anomalies associated with DS [Head et al., 2007]. In some cases, overexpression of a gene may be limited to a particular developmental stage and to particular cell types during this developmental stage yet still have life-long effects on the phenotype. On the other hand, there may be genes whose continued overexpression may be required to maintain their effect on phenotype. Clearly, these patterns of gene overexpression have profound consequences for rational design of therapies based on gene function. Additional levels of

specificity may also be important, for example, subcellular localization of protein expression. Information of this type should be extremely useful in understanding protein function and interactions, and elegant studies have been initiated to obtain this information [Hu et al., 2006].

The Genome–Transcriptome–Proteome–Interactome–Metabolome–Phenome Continuum

It is becoming well recognized that increased gene expression as measured by transcript levels does not always accurately predict increased levels of protein for corresponding transcripts. Perhaps equally problematic is that increased levels of particular proteins may not have the expected, or even any, phenotypic consequences. As discussed earlier, proteins that interact with other proteins, when elevated, may or may not have consequences for the interacting proteins, and thus may or may not have functional consequences. A less well-recognized but important complexity is that many enzymes are highly regulated, for example, by end-product inhibition, phosphorylation, stability, and other mechanisms. Thus, simple detection of an elevated level of an enzyme protein may not accurately reflect increased cognate enzyme activity and certainly may not accurately reflect altered metabolism. Metabolic flux analysis may be of use to assess the relevance of any alterations in enzyme activity levels [Cornish-Bowden and Cardenas, 2000]. This may be especially a problem when human proteins are expressed in transgenic mice [Spiegel, 2007; Patterson, unpublished results].

These considerations give strong impetus to metabolomic studies of transgenic and trisomic mice and where possible individuals with DS. Metabolomics is a relatively new endeavor but solidly based on decades of biochemistry. Nonetheless, new methodologies are required to carry out the detailed analyses required. Here, two complementary but different approaches to metabolomic analysis need to be undertaken. One is an undirected approach in which changes in the metabolome of individuals are assessed globally without preconceived hypotheses regarding what metabolic pathways might be involved. This approach is beginning to show promise with other human disorders and should be applied to DS [Kaddurah-Daouk et al., 2004; Rozen et al., 2005]. On the other hand, compelling hypotheses about the possible role of particular metabolic pathways can be tested by metabolomic approaches [Bagley and Selhub, 2000]. A combination of introduction of specific mutations into the genome of mice coupled with metabolomic analysis of the effects of the mutations may yield unexpected results and presents an important opportunity. Again, it is important to remember that alterations in metabolites may or may not have phenotypic consequences. This could be because increases or decreases of a particular metabolite may truly have no phenotypic consequence or because we are currently unable to discern the phenotypic consequence.

Enhanced bioinformatics approaches are likely to be required to integrate the results obtained from the studies described earlier and additional studies that are likely to emerge.

Ultimately, the goals of all these efforts should be to develop knowledge necessary to design therapies to ameliorate the abnormalities caused by trisomy of human chromosome 21. Fortunately, the methods are available or under development to achieve immense progress to understand the various genetic factors that influence the variable phenotypes seen in individuals with DS. This information will need to be integrated fully with the other aspects of the research agenda discussed in this volume. ■

REFERENCES

Aapola U, Kawasaki K, Scott HS, et al. 2000. Isolation and initial characterization of a novel zinc finger gene. DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 65:293–298.

Antonarakis SE, Epstein CJ. 2006. The challenge of Down syndrome. Trends Mol Med 12: 473–479.

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

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

Austin CP, Battey JF, Bradley A, et al. 2004. The Knockout Mouse Project. Nat Genet 36: 921–924.

Bagley PJ, Selhub J. 2000. Analysis of folate form distribution by affinity followed by reversed-phase chromatography with electrochemical detection. Clin Chem 46:404–411.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281–297.

Bartel DP, Chen C-Z. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet 5:396–400.

Brodsky G, Barnes T, Bleskan J, et al. 1997. The human GARS-AIRS-GART gene encodes two proteins which are differentially expressed during human brain development and temporally overexpressed in cerebellum of individuals with Down syndrome. Hum Mol Genet 6:2043–2050.

Butler C, Bowersox J, Forbes S, et al. 2006. The production of transgenic mice expressing human cystathionine β-synthase to study Down syndrome. Behav Genet 36:429–438.

Cabrejo L, Guyant-Marechal L, Laquerriere A, et al. 2006. Phenotype associated with APP duplication in five families. Brain 129: 2966–2976.

Centers for Disease Control and Prevention. 2006. Improved national prevalence estimates for 18 selected major birth defects—United States, 1999–2001. Morb Mortal Wkly 54:1301–1305.

Cornish-Bowden A, Cardenas ML. 2000. From genome to cellular phenotype—a role for metabolic flux analysis? Nat Biotechnol 16: 267–268.

Crawley JN. 2007. What's wrong with my mouse? Hoboken, NJ: Wiley.

Cui Q, Yu Z, Pan Y, et al. 2007. MicroRNAs preferentially target the genes with high transcriptional regulation complexity. Biochem Biophys Res Commun 352:733–738.

Dermitzakis ET, Kirkness E, Schwarz S, et al. 2004. Comparison of human chromosome 21 conserved nongenic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment. Genome Res 14: 852–859.

Dermitzakis ET, Reymond A, Antonarakis SE. 2005. Conserved non-genic sequences—an unexpected feature of mammalian genomes. Nat Rev Genet 6:151–157.

Deutsch S, Lyle R, Dermitzakis ET, et al. 2005. Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. Hum Mol Genet 14:3741–3749.

ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia of DNA Elements) Project. Science 306:636–640.

Fahr KK-L, Grimson A, Jan C, et al. 2005. The widespread impact of mammalian microRNAs on mRNA repression and evolution. Science 310:1817–1821.

Gaidatzis D, van Nimwegen E, Hausser J, et al. 2007. Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC Bioinform 8:69.

Gardiner K. 2006. Transcriptional dysregulation in Down syndrome: predictions for altered protein complex stoichiometries and post-translational modification, and consequences for learning/behavior genes ELK, CREB, and the estrogen and glucocorticoid receptors. Behav Genet 36:439–453.

Gardiner K, Costa ACS. 2006. The proteins of chromosome 21. Am J Med Genet C 142: 196–205.

Giardine B, Riemer C, Hefferon T, et al. 2007. PhenCode: connecting ENCODE data with mutations and phenotype. Hum Mutat 28: 554–562.

Gitton Y, Dahmane N, Baik S, et al. 2002. A gene expression map of human chromosome 21 orthologues in the mouse. Nature 420: 586–590.

Hata K, Kusumi M, Yokomine T, et al. 2006. Meiotic and epigenetic aberrations in the Dnmt3l-deficient male germ cells. Mol Reprod Dev 73:116–122.

Hattori M, Fujiyama A, Taylor TD, et al. 2000. The DNA sequence of human chromosome 21. Nature 405:311–319.

Head E, Lott IT, Patterson D, et al. 2007. Possible compensatory events in adult Down syndrome brain prior to the development of Alzheimer disease neuropathology. Targets for nonpharmacological intervention. J Alzhheimer's Dis 11:61–76.
Hegele RA. 2007. Copy-number variations add a new layer of complexity in the human genome. CMAJ 176:441–442.
Hu Y-H, Warnatz H-J, Vanhecke D, et al. 2006. Cell array-based intracellular localization screening reveals novel functional features of human chromosome 21 proteins. BMC Genomics 7:155.
Iafrate AJ, Feuk L, Rivera MN, et al. 2004. Detection of large-scale variation in the human genome. Nat Genet 36:949–951.
Kaddurah-Daouk R, Beecher C, Kristal BS, et al. 2004. Bioanalytical advances for metabolomics and metabolic profiling. PharmaGenomics 4:46–52.
Kahlem P. 2006. Gene-dosage effect on chromosome 21 transcriptome in Trisomy 21: implication in Down syndrome cognitive disorders. Behav Genet 36:416–428.
Kawahara Y, Zinshteyn B, Sethupathy P, et al. 2007. Redirection of silencing targets by adenosine-to-inosine editing of mRNAs. Science 315:1137–1140.
Kim Y. 2005. Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. J Nutr 135:2703–2709.
Korenberg JR, Kawashima H, Pulst SM, et al. 1990. Molecular definition of a region of chromosome 21 that causes features of the Down syndrome phenotype. Am J Hum Genet 47:236–246.
Kosik KS, Krichevsky AM. 2005. The elegance of the microRNAs: a neuronal perspective. Neuron 47:779–782.
Lage K, Karlberg EO, Storling ZM, et al. 2007. A human phenome–interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 25:309–316.
Li CM, Guo M, Salas M, et al. 2006. Cell type-specific over-expression of chromosome 21 genes in fibroblasts and fetal hearts with trisomy 21. BMC Med Genet 7:24.
Li Z, Yu T, Morishima M, et al. 2007. Duplication of the entire 22.9-Mb human chromosome 21 syntenic region on mouse chromosome 16 causes cardiovascular and gastrointestinal abnormalities. Hum Mol Genet 16: 1359–1366.
Liu M, Ge Y, Cabelof DC, et al. 2005. Structure and regulation of the murine reduced folate carrier gene. J Biol Chem 280:5588–5597.
Liu M, Ge Y, Payton SG, et al. 2006. Transcriptional regulation of the human reduced folate carrier in childhood aculte lymphoblastic leukemia cells. Clin Cancer Res 12: 608–616.
Lujambio A, Ropero S, Ballestar E, et al. 2007. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res 67:1424–1429.
Maas S, Kawahara Y, Tamburro K, et al. 2006. A-to-I RNA editing and human disease. RNA Biol 3:1–9.
Mao R, Zielke CL, Zielke HR, et al. 2003. Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain. Genomics 81:457–467.
Miller CA, Sweatt JD. 2007. Covalent modification of DNA regulates memory formation. Neuron 53:857–869.
Moscow JA, Gong M, He R, et al. 1995. Isolation of a gene encoding a human reduced folate carrier (RFC1) and analysis of its expression in transport-deficient, methotrexate-resistant human breast cancer cells. Cancer Res 55:3790–3794.
Muntean AG, Ge Y, Tauw JW, et al. 2006. Transcription factor GATA-1 and Down syndrome leukemogenesis. Leuk Lymphoma 47:986–997.
Nam J-W, Shin K-R, Han J, et al. 2005. Human microRNA prediction through a probabilistic co-learning model of sequence and structure. Nucleic Acids Res 33:3570–3581.
Niwa R, Slack FJ. 2007. The evolution of animal microRNA function. Curr Opin Genet Dev 17:1–6.
Nord AS, Chang PJ, Conklin BR, et al. 2006. The international gene trap consortium website: a portal to all publicly available gene trap cel lines in mouse. Nucleic Acids Res 34:D642–D648.
O'Doherty A, Ruf S, Mulligan C, et al. 2005. An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. Science 309:2033–2037.
O'Leary DA, Pritchard MA, Xu D, et al. 2004. Tissue-specific overexpression of the HSA21 gene GABPA: implications for DS. Biochim Biophys Acta 1739:81–87.
Olson LE, Richtsmeier JT, Leszl J, et al. 2004a. A chromosome 21 "critical region" does not cause specific Down syndrome phenotypes. Science 306:687–690.
Olson LE, Roper RJ, Baxter LL, et al. 2004b. Down syndrome mouse models Ts65Dn, Ts1Cje, and Ms1Cje/Ts65Dn exhibit variable severity of cerebellar phenotypes. Dev Dyn 230:581–589.
Olson LE, Roper RJ, Sengstaken CL, et al. 2007. Trisomy for the Down syndrome critical region is necessary but not sufficient for brain phenotypes of trisomic mice. Hum Mol Genet 16:774–782.
Pant SS, Moser HW, Krane SM. 1968. Hyperuricemia in Down's syndrome. J Clin Endocrinol Metab 28:472–478.
Patja K, Pukkala E, Sund R, et al. 2006. Cancer incidence of persons with Down syndrome in Finland: a population-based study. Int J Cancer 118:1769–1772.
Patterson D, Costa ACS. 2005. Down syndrome and genetics—a case of linked histories. Nat Rev Genet 6:137–147.
Payton SG, Haska CL, Flatley RM, et al. 2007. Effects of 5′ untranslated region diversity on the posttranscriptional regulation of the human reduced folate carrier. Biochim Biophys Acta 1769:131–138.
Payton SG, Whetstine JR, Ge Y, et al. 2005. Transcriptional regulation of the human reduced folate carrier promoter C: synergistic transactivation by Sp1 and C/EBP β and identification of a downstream repressor. Biochim Biophys Acta 1727:45–57.
Pokharel S, Beal PA. 2006. High-throughput screening for functional adenosine to inosine RNA editing systems. ACS Chem Biol 1:761–765.
Rachidi M, Lopes C, Charron G, et al. 2005. Spatial and temporal localization during embryonic and fetal human development of the transcription factor SIM2 in brain regions altered in Down syndrome. Int J Dev Neurosci 23:475–484.
Reymond A, Marigo V, Yaylaoglu MB, et al. 2002. Human chromosome 21 gene expression atlas in the mouse. Nature 420:582–586.
Ronan A, Fagan K, Christie L, et al. 2007. Familial 4.3 Mb duplication of 21q22 sheds new light on the Down syndrome critical region. J Med Genet 44:448–451.
Roper RJ, Reeves RH. 2006. Understanding the basis for Down syndrome phenotypes. PloS Genet 2:0231–0236.
Rovelet-Lecrux A, Hannequin D, Raux G, et al. 2006. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24–26.
Rozen S, Cudkowicz ME, Bogdanov M, et al. 2005. Metabolomic analysis and signatures in motor neuron disease. Metabolomics 1:101–108.
Salehi A, Delcroix J-D, Belichenko PV, et al. 2006. Increased App expression in a mouse model of Down's syndrome disrupts NGF transport and causes cholinergic degeneration. Neuron 51:1–14.
Schumacher A, Kapranov P, Kaminsky Z, et al. 2006. Microarray based DNA methylation profiling: technology and applications. Nucleic Acids Res 34:528–542.
Sebat J, Lakshmi B, Troge J, et al. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.
Shames DS, Minna JD, Gazdar AF. 2007. DNA methylation in health, disease, and cancer. Curr Mol Med 7:85–102.
Shapiro BL. 1983. Down syndrome—a disruption of homeostasis. Am J Med Genet 14:241–269.
Song F, Smith JF, Kimura MT, et al. 2005. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci USA 102: 3336–3341.
Spiegel EK. 2007. Psychomotor deficits in mice transgenic for a mutant adenylosuccinate lyase associated with autism in humans, Ph.D. thesis, University of Colorado at Denver and Health Sciences Center, Colorado.
Stranger BE, Forrest MS, Clark AG, et al. 2005. Genome-wide associations of gene expression variation in humans. PLoS Genet 1: 695–704.
Stranger BE, Forrest MS, Dunning M, et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315:848–853.
Thomas DJ, Rosenbloom KR, Clawson H, et al. 2006. The ENCODE Project at UC Santa Cruz. Nucleic Acids Res 35:D663–D667.
Turek-Plewa J, Jagodzinski PP. 2005. The role of mammalian DNA methyltransferases in the regulation of gene expression. Cell Mol Biol Lett 10:631–647.
Vyas P, Crispino JD. 2007. Molecular insights into Down syndrome-associated leukemia. Curr Opin Pediatr 19:9–14.
Watanabe H, Fujiyama A, Hattori M, et al. 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature 249: 382–388.
Whetstine JR, Flatley RM, Matherly LH. 2002. The human reduced folate carrier gene is ubiquitously and differentially expressed in normal human tissues: identification of seven non-coding exons and characterization of a novel promoter. Biochem J 367:629–640.
Whetstine JR, Matherly LH. 2001. The basal promoters of the human reduced folate carrier gene are regulated by a GC-box and a cAMP-response element/AP-1-like element. J Biol Chem 276:6350–6358.
Yamada Y, Watanabe H, Miura F, et al. 2004. A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. Genome Res 14:247–266.