

The Year in Human and Medical Genetics

Highlights of 2007–2008

Moyra Smith

*University of California, Irvine, Pediatrics, College of Health Sciences,
Irvine, California, USA*

The breadth and scope of new information makes difficult a selection of topics to be included in a limited review of highlights of the year. Admittedly, the choices are idiosyncratic. The eight topics presented here are (1) structural and copy number variants in the human genome; (2) progress in defining genetic factors in the etiology of schizophrenia; (3) microRNAs in central nervous system development and function; (4) progress in elucidation of risk factors for complex common disorders through large scale association studies; (5) epigenetics and the epigenomic era; (6) reprogramming of somatic cell nuclei to generate pluripotent stem cells; (7) new concepts regarding factors involved in sexual differentiation; and (8) Duffy blood group antigens: new concepts, and new discoveries on the role of these antigens in malaria and HIV–AIDS.

Key words: human genome; copy number variant or variation (CNV); schizophrenia; microRNA (miRNA); epigenetics; pluripotent stem cell; sexual differentiation; Duffy antigen; malaria; HIV–AIDS

Review Topics

The rapid expansion of knowledge in human and medical genetics is a source of excitement and joy to those of us who work in these areas. However, the breadth and scope of new information makes difficult the selection of topics to include in a limited review of highlights of the year. I have selected eight topics and realize that the choice is somewhat idiosyncratic.

Topics selected for review:

1. Structural and copy number variants (CNVs) in the human genome.
 - Copy number variation (also given by CNV in the literature) and gene expression
 - Copy number variants in disorders associated with abnormal immune response
 - Copy number variants in neurological and neurobehavioral disorders
2. Progress in defining genetic factors in the etiology of schizophrenia
 - Alternate hypotheses concerning the origin of schizophrenia
 - *De novo* copy number changes in schizophrenia
 - Evidence for rare copy number variants under negative selective pressure in schizophrenia etiology
 - Evidence for roles for genome-wide burden of copy number variants and rare variants at specific loci in schizophrenia pathogenesis
 - Analysis of the role of chromosome 22q11.2 deletion in schizophrenia
3. MicroRNAs (miRNAs) in central nervous system development and function
4. Progress in elucidation of risk factors for complex common disorders through large-scale association studies
5. Epigenetics and the epigenomic era
 - Epigenetic mechanisms
 - Epigenome projects and genome-wide chromatin analysis

Address for correspondence: Moyra Smith, University of California—
Irvine, Pediatrics, College of Health Sciences, Irvine, California, USA.
Voice: 949 824-7469. Email: dmsmith@uci.edu

- Therapeutic interventions that target chromatin remodeling or DNA methylation
 - Epigenetic mechanisms and the Barker hypothesis
6. Reprogramming of somatic cell nuclei to generate pluripotent stem cells
 - Factor-based reprogramming
 - Somatic cell nuclear transfer into oocytes
 7. New concepts regarding factors involved in sexual differentiation
 8. Duffy blood group antigens: new concepts and new discoveries on the role of these antigens in malaria and HIV–AIDS.

Structural and Copy Number Variants in the Human Genome

During the past two years significant progress was made in further defining the genomic architecture and extent of copy number variation in the human genome. Researchers are in the early stages of gathering information on the extent to which copy number variation impacts levels of gene expression. Examples of such studies discussed below include analysis of levels of salivary amylase and polymorphic variation of amylase gene copy number and evidence that copy number variation plays a significant role in the etiology of immune response. There is also evidence of the role of copy number variation in the pathogenesis of neurological and neurobehavioral disorders.

In order to define the extent of genomic copy number variation and the mechanisms involved in the generation of these variants, Kidd and colleagues¹ cloned and sequenced structural variants greater than 8 kb in length present in eight different individuals of Asian, European, or African ancestry. These studies led to generation of maps of 1695 sites of structural variation. Data derived from these studies indicated that nonallelic homologous recombination was the mechanism for structural and copy

number variants in 48% of cases. Their data revealed that recombination between segmental duplications in the genome is more common than L1- or Alu transposon-mediated events. Kidd and colleagues identified inversions and insertions in specific regions associated with recurrent structural variation. They noted that a future goal of sequence analysis will be identification of haplotypes associated with specific structural variants.¹

Copy Number Variation and Gene Expression

Key questions remain regarding the extent to which copy number variations impact quantity of gene product produced. Population-specific differences exist in the copy number of salivary amylase genes *AMY1*, and in the levels of the *AMY1*-encoded protein in saliva. Furthermore there is a positive correlation with concentration of salivary amylase protein and starch content of the diet.² On average, individuals in populations who have a high starch diet (e.g., Japanese, Hazda) have higher salivary amylase levels than individuals in populations where the starch content of the diet is low (e.g., Yakut and Biaka). Perry and colleagues note² that variation in salivary protein levels is only partly explained by variation in gene copy number. They attributed the observed variation in part to other genetic factors and to environmental factors.

Copy Number Variation Associated with Abnormal Immune Response

HIV–AIDS Susceptibility and Polymorphisms

HIV infection in humans is associated with immune activation and loss of CD4 T cells. There is evidence that the CCR5 cytokine receptor on CD4 T cells plays an important role in HIV infection. The CCR5 receptor interacts with the HIV virus. The CCR5 receptor also interacts with the cytokine CCL3L1 (chemokine (C-C motif) ligand 3-like 1). This

interaction then blocks entry of HIV. The CCR5 and CCL3L1 polymorphisms apparently interact to influence HIV infection and response. The 35 base-pair deletion polymorphism in CCR5 impairs viral entry. HIV entry through CCR5 is also impaired by high copy variants of CCL3L1.³

Willcocks⁴ demonstrated a role in immune complex clearance for copy number variation in FCGR3B (receptor for Fc fragment of IgG). They noted an association of low FCGR3B copy number and occurrence of systemic lupus erythematosus. This report follows on the study of Aitman and colleagues,⁵ who reported low FCGR3B copy number variants (CNVs) predisposed to glomerulonephritis in patients with systemic lupus erythematosus.

Beta Defensin Gene Copy Number in Psoriasis

Defensin-encoding gene clusters occur on chromosome 20 and on 8p23. Defensins are small antimicrobial peptides that act as cytokines. The DEFB1 gene is separated from the other seven beta defensin genes that are encoded on a segment that varies in copy number in different individuals at between 2 and 12 copies.⁶ DEFB3 located within this segment encodes a protein that is expressed in skin. The DEFB4 gene also maps in this cluster and it encodes a protein hBD2 that is induced during the inflammatory response in skin.

Psoriasis is a chronic inflammatory skin disease; genetic and environmental factors play a role in its etiology. Hollox and colleagues⁷ undertook a study of defensin gene copy number in patients with psoriasis. They studied Dutch and German patients and controls; the total number of patients was 498 and the number of controls was 577. They reported a significant association between higher copy number of beta defensin genes and risk of psoriasis. The mean beta defensin gene copy number was 4 in both cases and control. The relative risk of psoriasis in individuals with 2 copies was 0.31,

with 4 copies, it was 0.84, and in individuals with 6 copies of the beta defensin gene cluster it was 1.69.

Hollox and colleagues⁷ postulated that high resting levels of beta defensins or high induced levels following an environmental trigger may be psoriasis risk factors.

Copy Number Variations in Neurological and Neurobehavioral Disorders

Blauw and colleagues⁸ carried out studies in 406 patients with sporadic ALS and 404 controls to search for copy number changes. They identified 155 gene-containing DNA segments that were deleted exclusively in patients. In analyzing function of proteins encoded by genes in these regions they determined that many are involved in oxidative phosphorylation, regulation of actin cytoskeleton, and interaction between cytokines and receptors. On the basis of their study they postulated that multiple rare deletions may play important roles in pathogenesis of sporadic ALS.

The role of copy number variations in schizophrenia is discussed below. Evidence for the importance of copy number variations in autism is presented in a separate chapter in this volume.

Summary

1. The extent of structural variation in the genome, as revealed by microarray studies was not previously anticipated.
2. More studies will be required to determine the impact of this structural variation on gene expression.
3. It is possible that these variants may impact phenotype more than single nucleotide changes.
4. Structural variants and single nucleotide polymorphisms likely play roles in common complex diseases and will need to be taken into account in association studies.
5. Further studies will be required to determine whether or not the frequency of CNVs in the human genome is increasing

and to define the cause of this altered frequency.

Progress in Defining Genetic Factors in Schizophrenia Etiology

Alison Abbott⁹ reviewed the apparent lack of progress in identifying genes that play key roles in the etiology of schizophrenia. Apparently, neuroscientists, psychiatrists, and geneticists identify different causes for lack of progress. Geneticists express concern about inconsistent diagnostic categories and the possibility that patients selected for study may not even have the same disorder. A number of scientists are skeptical as to whether inclusion of large numbers of patients will solve ambiguities in association studies, particularly in light of lack of specificity of symptoms used to diagnose the specific psychiatric disorder selected for analysis.

Alternate Hypotheses Concerning the Origin of Schizophrenia

McClellan and King¹⁰ emphasized the importance of the common disease/rare variant, genetic heterogeneity model of schizophrenia, particularly in light of the fact that most patients with schizophrenia have no close relative affected with that disorder. They noted, however, that the common disease/common allele, and common disease/rare allele models of disease need not be mutually exclusive. They refer to Alzheimer's disease that is induced by rare mutations in amyloid precursor protein and rare mutations in presenilin 1 and 2. However, there is also evidence that a common allele APOE4 significantly alters risk.

In their review Burmeister and colleagues¹¹ emphasized that progress in identifying important genes involved in psychiatric disease is likely to be dependent on more attention being focused on unique families, rare variants, and specific environmental factors, and on incorporating emerging knowledge of bio-

logical function into analyses. They identified the following key problem factors: imprecise specification of traits, locus heterogeneity, and incomplete penetrance and interaction with nongenetic factors. They noted that analysis of endophenotypes and rare families and concentration on pathways may be particularly important.

De Novo Copy Number Changes in Schizophrenia

In a review of their studies in schizophrenia Walsh and colleagues¹² expressed the view that individually rare mutations that alter genes in neurodevelopmental pathways contribute to etiology of this disorder. This is an alternate hypothesis to the common disease, common allele hypothesis that drives association studies. Walsh and colleagues proposed that highly penetrant, individually rare mutations of recent origin and perhaps specific to specific individuals or families may be the key etiological factors in schizophrenia. They postulated that the same severe mutations may each occur only in a few patients, and they searched for structural variations,¹² i.e., deletions and duplications greater than 100 kb in size, in DNA samples from 150 individual patients with schizophrenia and in DNA from 268 ancestry-matched controls. Their studies revealed that the frequency of known common CNVs did not differ between cases and controls. However, unique structural variants (not previously described) that altered genes occurred with 4X higher frequency in cases with disease onset before 18 years of age than in controls. There was no difference between the case and control groups with respect to duplications and deletions that do not alter genes. Virtually every structural variant they identified in cases with childhood onset schizophrenia was unique to a specific individual. Of particular interest was the finding that in their study, genes involved in structural and copy number changes were frequently active in pathways important for brain development.

Walsh and colleagues¹² identified structural variations in genes that encode neuregulin, and ERK/MAPK tyrosine kinase signaling pathway proteins in genes that encoded proteins involved in long-term potentiation, axonal guidance, integrin signaling, and glutamate receptor signaling. In controls structural variations did not impact genes in these pathways. They characterized mutant transcripts in two of the structurally altered genes. They noted that the mutant transcript generated by a deletion within the ERBB4 gene would, if translated, give rise to a protein that lacked an intracellular kinase domain. The ERBB4 gene encodes a transmembrane tyrosine kinase receptor for neuregulins. Disruption of function of the neuregulin receptor would impact neuronal migration and synaptic transmission. They observed a deletion that encompassed parts of the exons within the SLC1A3 and SKP2 genes. They noted that this deletion led to a chimeric transcript with different function. The SLC1A3 gene encodes a protein that functions as a glutamate transporter. SKP2 encodes an S-phase kinase. Walsh and colleagues emphasized that any gene that harbors a large structural variant, particularly a deletion, is a candidate for screening for smaller deletions and mutations in other individuals.

Xu and colleagues¹³ carried out a study of large-scale genomic variants in 1077 triads (proband and biological parents). Their sample included 152 sporadic cases of schizophrenia with no family history of this disorder in first- or second-degree relatives. They also studied 48 probands and parents with a family history of schizophrenia in first- or second-degree relatives. They analyzed a control sample comprising 159 unaffected individuals and their parents. All studies were carried out on the same population. Xu and colleagues used the Affymetrix 5.0 chip microarray. They considered variants that encompassed 10 consecutive probe sets. They identified *de novo* copy number changes, including an approximately equal number of duplications and deletions, in 10% of sporadic cases of schizophrenia and in 1.5%

of control individuals. There was no increase in frequency of *de novo* copy number variants in cases with a family history of schizophrenia relative to controls. The frequency of rare structural variants inherited from unaffected parents was 30% in sporadic cases and 20% in controls.

Among the *de novo* deletions identified were deletions of chromosome 22q11.21 in three cases. These deletions led to hemizyosity for approximately 27 genes, including the DGCR8 locus that encodes a microRNA (miRNA) processing gene. This gene is of particular interest. In one patient deletion on chromosome 14q32.1 led to hemizyosity for the DICER1 gene that also plays a role in miRNA processing. MicroRNA processing and the DGCR8 gene will be discussed further below.

Another interesting *de novo* deletion discovered in the Xu and colleagues study¹³ was microdeletion of exons 2 through 11 of RAPGEF6 (Rap guanine nucleotide exchange factor), also known as PDZGEF2 on chromosome 5q31.1. This chromosome region was previously shown to be linked to, and associated with, schizophrenia. Fanous and colleagues¹⁴ identified a haplotype within a 758-kb interval that contained the SPEC2/PDZGEF2/ACSL6 genes and segregated with schizophrenia in high density schizophrenia families from Ireland. The high linkage disequilibrium in this region prevented them from distinguishing which gene generated the association signal.

Together the studies of Walsh and colleagues and Xu and colleagues provide^{12,13} evidence for the importance of copy number variation in schizophrenia, particularly in sporadic cases.

Evidence for Rare CNVs Under Negative Selective Pressure in Schizophrenia Etiology

Stefansson and colleagues¹⁵ proposed that rare copy number variants under negative selection may account for a large fraction of the overall risk in schizophrenia. They noted that this hypothesis would explain why common variants that confer risk have not been

identified. In the first phase of their study they set out to identify recurrent CNVs in a control population. They analyzed a control sample of 2160 trios and 5558 parent offspring pairs. In this sample they identified 66 *de novo* CNVs; of these, 23 were flanked by low copy repeats (LCRs). They then carried out analysis of these 66 CNVs in a sample of 1433 patients with schizophrenia and related psychoses and in 33,250 controls. For 3 of the 66 CNVs that showed large deletions, there was nominal evidence that they occurred more frequently in schizophrenia patients than in controls ($P < 0.05$). These three deletions were then analyzed in additional samples so that a total of 3285 affected patients and 7951 controls were examined.

In the combined sample the three deletions occurred with significantly greater frequency in patients with schizophrenia and psychosis than in controls. Significance values for frequency differences were for the 1q21.1 deletion $P = 2.9 \times 10^{-5}$, for the 15q11.2 deletion $P = 6 \times 10^{-4}$, and for the 15q13.3 deletion $P = 5.3 \times 10^{-4}$. When broader diagnostic criteria were used including schizoaffective disorder, schizophreniform disorder, and unspecified psychosis, only the 1q21.1 deletion remained significant ($P = 2.32 \times 10^{-5}$).

It is important to note that these deletions also occurred in the control sample but the frequency was three- to tenfold less in controls than in patients.

The chromosome 1q21.1 deletion ranged in size between 1.35 and 2.1 Mb.¹⁵ Stefansson and colleagues noted that in all cases the connexin gene *GJA8* is deleted. Association of this gene and schizophrenia was previously reported. On chromosome 15q11.2, the gene these investigators consider most likely to be important in etiology of schizophrenia is *CYFIP1*, which encodes a protein that interacts with the *FMRP* protein encoded by fragile X gene and with the signaling protein *RhoGTPase Rac1*. *CYFIP1* is not imprinted. The gene within 15q13.3 that they considered most significant in schizophrenia etiology is *CHRNA7*

that encodes the nicotinic receptor that is targeted to axons by *neuregulin1*, a gene previously implicated in schizophrenia.

Stefansson and colleagues¹⁵ proposed that the same mutation could increase the risk of a broad range of clinical psychopathologies. They emphasized the importance of identifying all recurrent copy number variations under negative selection.

Roles for Genome-Wide Burden of CNVs and Rare Variants at Specific Loci in Schizophrenia Pathogenesis

In 2008 the International Schizophrenia Consortium¹⁶ reported that results of their CNV analyses were consistent with a model of schizophrenia pathogenesis that includes the effects of genome-wide CNV burden and the presence of rare but highly penetrant CNVs at specific loci. They used SNP polymorphisms and Affymetrix 5.0 and 6.0 arrays to identify copy number variants in 3391 samples from schizophrenia patients and 3181 samples from controls. Evidence in support of the increased genome-wide burden of CNVs in schizophrenia included the fact that gene-carrying CNVs occurred more frequently in cases; the frequency ratio in cases versus controls was 1.41 to 1.0. In this study investigators encountered more deletions larger than 500 kb in cases than in controls. With respect to rare deletions in specific regions, this study identified deletions in 22q11.2, 15q13.3, and 1q21.2 that occurred more frequently in cases than in controls.

Investigators in the International Schizophrenia Consortium study emphasize that protective and environmental factors likely act in consort with CNVs to influence occurrence of schizophrenia manifestations.

Analysis of Chromosome 22q11.2 Deletion in Schizophrenia

A number of different studies have confirmed that the 22q11.2 deletion is an

important etiological factor in a subset of cases of schizophrenia. It is interesting to note that a single specific causative gene that maps within this region has not definitively been identified. An exciting new study points to the role of a miRNA processing enzyme encoded by DGCR8 as playing a key role. Stark and colleagues¹⁷ engineered a mouse strain that carried a hemizygous deletion that spanned a 1.3-mb segment on mouse chromosome 16, homologous to the human 22q11.2 deletion. The deletion encompasses 27 genes. The engineered heterozygous mice Df(16)A^{+/-} showed deficits in hippocampal connectivity with reduction in size and number of dendritic spines. Transcriptome analysis with strict criteria revealed decreased expression of genes in the deletion regions. They noted that a number of mitochondrial expressed genes are encoded in 22q11.2. These include SLC25A1 (solute carrier citrate transporter), TXNRD2 (thioredoxin reductase), MRPL140 (mitochondrial ribosomal protein L40), and PRODH (proline dehydrogenase).

Transcriptome analyses in the Df(16)A^{+/-} mice also revealed evidence for differential expression of genes that did not map in the deleted region. These included unique transcripts that were differentially expressed in the hippocampus and/or frontal cortex. Expression changes common to both the hippocampus and prefrontal cortex included altered expression of miRNAs. Stark and colleagues noted that there was a consistent upregulation of specific miRNA transcripts. The apparent upregulation was attributed to increased quantities of pri-miRNA species due to a bottleneck imposed by deficiency of Dgcr8 (Di George critical region 8) and decreased miRNA processing.

Stark and colleagues subsequently generated Dgcr8-deficient mice.¹⁷ The heterozygous Dgcr8 deficient mice Dgcr8^{+/-} showed altered levels of miRNA as seen in the Df(16)A^{+/-} mice. The Dgcr8 deficient mice (Dgcr8^{+/-}) and the Df(16)A^{+/-} mice manifested impaired performance on tasks requiring spatial working memory.

These studies provide interesting and significant evidence that the key gene in 22q11.2 that impacts neurobehavioral and cognitive phenotype may be a gene that plays a role in miRNA processing.

Risk Factors for Emergence of Psychiatric Symptoms in 22q11.2 Deletion Patients

It is interesting to note that one-third of patients with 22q11.2 deletion develop psychiatric symptoms. Psychotic symptoms similar to schizophrenia develop in adolescence or early adult life. Gothelf and colleagues¹⁸ studied a group of 60 children that included 31 with 22q11.2 deletion syndrome and 21 with idiopathic developmental delay matched for IQ. They determined that psychiatric symptoms developed in one-third of the 22q11.2 deletion patients. Risk factors identified included high anxiety, low IQ, and the presence on the remaining chromosome 22 of the low activity allele of COMT (catechol methyl transferase (COMT met)). They proposed that early recognition of 22q11.2 deletion patients at particularly high risk for psychosis will permit early intervention and decrease disability.

Consideration of Gene Networks and Epistasis in Schizophrenia

Talkowski and colleagues¹⁹ analyzed polymorphisms in 18 genes in the dopaminergic pathway in patients with schizophrenia. The impetus for these studies was the observation of the clinical efficacy of antipsychotic drugs that have affinity for the dopamine receptors. In a series of samples from the USA and a series from Bulgaria, the researchers found evidence of association between schizophrenia and polymorphisms at four different loci: SL6A3 (dopamine transporter), DRD3, SLC18A2 (vesicular mono-amine transporter), and COMT (Catechol-O-methyltransferase). COMT catalyzes the transfer of a methyl group from S-adenosylmethionine to catecholamines, including the neurotransmitters dopamine, epinephrine, and norepinephrine.

This O-methylation reaction results in degradation of the catecholamine neurotransmitters. Talkowski and colleagues then tested for epistatic interactions. Their epistasis study indicated that variations at SLC6A3 are important determinants of schizophrenia and that variants at SLC18A2, DRD3, and COMT contribute to risk.

Summary

1. Recent studies by Walsh and colleagues¹² and Xu and colleagues¹³ indicate that structural genomic variations, particularly deletions, may play a major role in schizophrenia. Many of the changes may be rare and in different individuals different changes occur.
2. It is possible, as discussed by McLellan and King,¹⁰ that two hypotheses regarding etiology, i.e., the common disease/common variant hypothesis, and the common disease/rare variant hypothesis are both relevant.
3. It is important to bear in mind that even in cases where a major causative variant is present, e.g., chromosome 22q11 deletion, this variant does not inevitably lead to development of psychosis; other genetic and environmental factors apparently play a role.
4. Large-scale studies on *de novo* copy number changes in subjects with schizophrenia and in controls by Stefansson and colleagues¹⁵ provided confirmation for the hypothesis that recurrent deletions in the genome for which there is negative selection plays an important role in the etiology of schizophrenia.
5. Studies on schizophrenia subjects and controls reported by the International Schizophrenia Consortium provided evidence that both the total burden of genomic copy number variation and the occurrence of individually rare, highly penetrant genomic deletions play important roles in schizophrenia pathogenesis.
6. It is likely that considerable progress will

be made in defining schizophrenia pathogenesis through investigations of gene networks and analysis of gene interactions.

MicroRNAs and the Central Nervous System

Biogenesis and function of microRNAs (miRNAs) in brain have been intensely studied during the past two years by Barbato and colleagues²⁰ MicroRNAs are abundantly expressed in the central nervous system (CNS), and there is a growing body of evidence that they play a role in embryonic life, in subsequent development and in functioning of the mature nervous system. Primary miRNAs (pri-miRNAs) are transcribed usually by RNA polymerase 2 and they contain 5' cap structures and poly-A tails. MiRNAs may be derived from introns or exons of protein-coding genes; they may be transcribed in antisense orientation from coding sequences or from non-protein-coding sequence. Within the nucleus the ribonuclease RNase III Drosha processes pri-miRNAs to produce pre-miRNAs that are transported to the cytoplasm through the activities of exportin 5 and RanGTP. Within the cytoplasm they are further processed by Dicer ribonuclease and nucleotides of approximately 18–24 bases are generated. These miRNA duplexes are then incorporated into multiprotein RNA complexes: RNA-induced silencing complexes RISC. This complex then binds to the target mRNA complex that manifests sequence complementarity to the miRNA. Following binding the miRNA RISC complex blocks mRNA translation or leads to mRNA degradation. MicroRNAs are therefore small regulatory RNA molecules. MiRNA production is highly regulated and there is evidence of developmental and tissue-specific patterns of regulation. This may occur at the level of miRNA transcription or at the level of processing of pri-miRNA or pre-mRNA to mature miRNA. There is new evidence that indicates that, under specific conditions and in specific

phases of the cell cycle, miRNAs may upregulate translation.²¹

In recent years much effort was invested in identification of microRNA targets and determination of effect on protein synthesis of miRNA target binding to mRNA. Selbach and colleagues²² devised a method to directly measure the effect on miRNA expression on protein synthesis on a genome-wide scale. Their method involved addition of different forms of amino acid isotopes to culture medium in pulse labeling experiments. They used transfection to introduce a specific miRNA into the cell. Results of their studies revealed that microRNAs downregulate mRNA levels and reduce protein translation. Most important, they demonstrated that a specific miRNA, e.g., miLet7b, can reduce expression levels of hundreds of proteins. This effect may be direct or indirect.

Studies are ongoing to identify brain-specific microRNAs. MiRNA133b is present in high amounts in brain and is deficient in patients with Parkinson's disease. A microRNA that has high nucleotide conservation from *C. elegans* to humans is miRNA124a; it is expressed in embryonic and mature central nervous system. A number of investigators have demonstrated the importance of miRNA in the development of the central nervous system. Neuronal cell-specific miRNAs act to reduce availability of the neuron restrictive silencing factor NRSF/REST. In non-neuronal cells this factor inhibits production of neuronal gene products. Neuronal development is in part dependent upon repression of the REST silencing transcription factor by miRNA124a. The latter also downregulates a number of other mRNAs. MIRNA 134 is a brain-specific miRNA that localizes near synaptic sites in dendrites and impacts width of dendrites.²³

There is evidence that the ribonuclease Dicer plays a very important role in nervous system morphogenesis, and in the mature nervous system Dicer has functions other than miRNA processing, e.g., it plays a role in heterochromatin processing.²⁴ Dicer is abun-

dantly expressed in dendritic spines and in postsynaptic densities. In lower organisms, e.g., zebra fish, Dicer mutations lead to altered brain morphogenesis. In mice Dicer plays an essential role in maintenance of dopamine neurons.

Correct functioning of the CNS is dependent on specific synapse formation and functioning. Bicker and Schrott²⁵ noted that local protein synthesis at synapses likely contributes to synaptic maturation and plasticity. Experimental inactivation of Dicer ribonuclease led to neuronal degradation and disrupted cellular and tissue morphogenesis. With respect to neuronal plasticity and function Bicker and Schrott emphasized the important role played by miRNA 132. CREB (cyclic AMP response element binding protein) regulates production of miRNA132 that in turn reduces translation of mRNA for a GTPase factor that inhibits neurite outgrowth. Studies of Kandel's group²⁶ provided evidence that modification of synaptic function and structure both play key roles in the generation and storage of memory. MicroRNAs impact synaptic strength and functioning since they impact dendrite development and synaptic activity.

There is increasing evidence that microRNAs play a role in specific neurodevelopmental disorders, including fragile X mental retardation. Li and colleagues²⁷ reported interaction between FMRP and the miRNA pathway Perkins and colleagues²⁸ reported that levels of miRNA are reduced in the prefrontal cortex of patients with schizophrenia. As noted above, Stark and colleagues¹⁷ reported that a key deficiency in patients with chromosome 22q11.2 deletion is deficiency of the gene DHCR8. The product of this gene interacts with the ribonuclease DROSHA and impacts microRNA processing.

Summary

1. It will become increasingly important when one defines the role of specific genomic segment changes in etiology of disease to identify within that segment,

miRNAs and factors that lead to changes in expression of genes.

2. This may be particularly important in psychiatric diseases, given evidence of the abundance of miRNAs in brain and their important role in neurodevelopment.

Progress in Elucidation of Risk Factors for Complex Common Disorders Through Large-Scale Association Studies

During 2007 and 2008 results of association studies that included analyses of very large numbers of patients with specific complex diseases were published in the The Wellcome Trust Case Control Consortium Study.²⁹ For each of seven major diseases, 2000 affected individuals were studied and results were compared to allele frequencies in a shared cohort of 3000 unaffected controls. Affected individuals and controls were from UK. The most powerful association was found in coronary heart disease, where specific alleles in marker rs1333049 on chromosome 9p21.3 yielded association with a significance value of $P = 1.8 \times 10^{-14}$. This marker is not within a gene but is adjacent to a region that contains two cyclin-dependent kinase inhibitors and the MTAP (methylthioadenosine phosphorylase) gene. A moderately strong association with coronary artery disease and a methylene tetrahydrofolate dehydrogenase MTHFD1L was found, significance value $P = 6.3 \times 10^{-6}$.

A number of markers were identified that showed significant association with Crohn's disease; several of these markers were identified in prior studies. The most significant association was with a marker within the interleukin receptor 23, IL23R ($P = 6.5 \times 10^{-13}$).

Two other associations of markers with Crohn's disease are of interest in terms of their functional significance. One highly significant association ($P = 2.1 \times 10^{-13}$) was with a marker on chromosome 5p13.1. This marker is described as being located in a gene desert but

has been shown to affect expression of a specific protein encoded by PTGER4 (prostaglandin E receptor 4). Crohn's disease was also associated with a SNP rs10761659, $P = 2.7 \times 10^{-7}$. This SNP is in an intergenic noncoding region on chromosome 10.

In the The Wellcome Trust Case Control Consortium study the previously described association of type 1 diabetes with the major histocompatibility locus was confirmed and MHC marker rs927234 revealed an association significance 2.42×10^{-134} . In addition, four other markers were associated with type 1 diabetes and had very high significant values. Five completely different loci were associated with type 2 diabetes.

Summary

1. The Wellcome study demonstrated the importance of very large samples sizes in harnessing the value of association studies.
2. The study report emphasized the fact that most of the disease-associated alleles have a modest effect on risk for the specific disorder.

Epigenetics and the Epigenomic Era Epigenetic Mechanisms

Epigenetic mechanisms include DNA methylation and changes in chromatin composition and RNA silencing. Baylin and Schuebel³⁰ drew attention to the new era opening to biologists, the epigenomic era. They noted that during this era biologists would be able to gain greater understanding of processes involved in generating cellular diversity and development. Epigenetic changes also play a key role in determining cell and tissue renewal and response of the organism to environmental variations. Furthermore there is evidence that both genetic and epigenetic factors are involved in cancer etiology.

Chromatin is made of DNA nucleosomes, around which DNA strands are wound, and nucleosomes are interconnected by DNA strands.

Nucleosomes are composed of histone subunits, including histone H3 and H4, and H2A and H2B. Histone 1 binds to the nucleosomes and to linker DNA strands between nucleosomes.³¹ Nucleosomes may be tightly compacted or they may be more widely spaced so that the interconnecting DNA strands are more exposed, i.e., chromatin structure is “open.” During chromatin remodeling histone complexes move a short distance along the DNA strand. This movement causes different DNA sequences to lie in the open region between nucleosome beads. Histone modification plays a key role in chromatin modification. Essential to chromatin structure are enzymes that methylate or demethylate DNA and enzymes that modify histones through addition of methyl, acetyl, or phosphate groups and enzymes that remove these modifying groups from DNA. Histone acetyltransferases promote acetylation of histone and this leads to an open chromatin structure and promotes transcription. Acetylation requires acetyl coenzyme A. Histone deacetylases remove acetyl groups from histone tails leading to denser chromatin structure and reduced transcription.

Epigenome Project and Genome-Wide Chromatin Analysis

In the epigenome projects investigators have begun comprehensive analyses to document the range of modifications of histones and to correlate these with DNA packaging and gene expression. They have also undertaken genome-wide analysis of DNA methylation. In many experiments to date chromatin is immune-precipitated prior to analysis. There are difficulties in isolation of chromatin fragments due to problems related to specificity of antibodies. Increasingly mass spectrometry will likely be used to isolate chromatin fragments.³²

In 2007 Mikkelsen and colleagues³³ reported results of high throughput profiling of histone modifications in mammalian cells. They analyzed DNA and chromatin immunoprecipitates from different cell types. They reported

that trimethylation of histone H3 lysine 4 and H3 lysine 27 discriminates between expressed genes and repressed genes.

They determined that genome-wide maps of H3 lysine 4 trimethylation and H3 lysine 36 trimethylation may provide a means to discriminate between coding and noncoding genomic DNA. They noted that trimethylation of H3 lysine 9 and histone H4 lysine 20 occurred in satellites, telomeres, and in long terminal repeats in DNA

Barski and colleagues³⁴ reported generation of human genomic maps of monomethylation methylation of histone lysine and arginine residues. They defined characteristics of expressed and repressed gene regions.

Meissner and colleagues³⁵ used representation bisulfite sequencing of DNA for genome scale analysis of nucleotide methylation. Bisulfite sequencing is based on the observation that when DNA is treated with bisulfite, cytosine residues are converted to uracil. However, 5-methylcytosine residues are not converted. Analysis of DNA sequence following bisulfite sequencing yields high-resolution information on the methylation status of the defined segment of DNA used for sequence analysis.

Meissner and colleagues generated DNA methylation maps of CpG islands, transposons and noncoding elements in mouse embryonic stem cells and in eight other tissues. They confirmed that methylation of CpG constitutes a dynamic epigenetics mark that undergoes changes during cell differentiation.

Therapeutic Interventions that Target Chromatin Remodeling or DNA Methylation

Chromatin remodeling is targeted primarily by histone deacetylase (HDAC) inhibitors. A number of different HDAC inhibitors have been used in experiments on animal models of specific genetic diseases to test efficacy. There are also ongoing clinical trials to test HDAC inhibitors such as suberoylanilide hydroxamic

acid (SAHA), sodiumphenylbutyrate, and valproic acid. The histone deacetylase inhibitor sodium phenylbutyrate is in clinical trials for treatment of Huntington's chorea.³⁶ This substance was shown to ameliorate manifestations of the Huntington's phenotype in animal models of the disorder.

The CGG-CCG repeat expansion in the 5' region of the FMR1 gene leads to fragile X mental retardation. Biacso and colleagues³⁷ reported that the class III histone deacetylase SIRT1 plays an important role in deacetylation and repressing expression of genes with this repeat. They propose that inhibitors of SIRT1 be applied to reverse the inhibition of gene expression.

Pandolfo and colleagues³⁸ carried out studies in a mouse model of Friedreich's ataxia with frataxin deficiency due to repeat expansion in the first intron of the gene. They demonstrated that a novel histone deacetylase inhibitor "compound 106" increased histone H3 and H4 acetylation near the frataxin gene promoter and increased gene expression. This compound had low toxicity.

Therapeutic strategies are also in place to modify DNA methylation. In situations where DNA is hypermethylated and reduction in methylation would be advantageous to enhance gene expression (e.g., in cancers where tumor suppressor genes are silenced through DNA methylation) inhibitors of DNMT (DNA methyl transferase) are utilized in clinical trials.³⁹

Methyl donors such as S-adenosylmethionine, methionine, and choline, and increasing intake of cofactor vitamins such as folate and vitamin B12 impact DNA methylation. Kovacheva and colleagues⁴⁰ reported that choline deficiency modulates fetal DNA methylation machinery in a complex fashion that includes hypomethylation of the regulatory CpGs within the DNA methyltransferase gene (*Dnmt1*). This leads to DNMT1 overexpression and results in increased global and gene-specific DNA methylation, e.g., *Igf2* hypermethylation.

Epigenetic Mechanisms and the Barker Hypothesis

Barker⁴¹ provided evidence for the fetal origins of adult disease. His studies revealed the association of low infant birthweight with adult onset coronary heart disease, type 2 diabetes, and obesity. The central concept of the Barker hypothesis is that poor or unbalanced nutrition during fetal life induces specific adaptations that result in reduced growth during fetal life. These adaptations have long-term effects, including a predisposition to adult onset chronic diseases. Hanson and Gluckman⁴² proposed important additions to the concept of the developmental origins of health and disease. They noted that the prenatal challenge may not necessarily be a change or changes that lead to low birthweight. They propose that prenatal challenge may entail exposure to toxic levels of chemical agents or pollutants, e.g., endocrine disruptors and bisphenol A. Another prenatal challenge may be reduced intake of micronutrients.

Jirtle and Skinner⁴³ propose that epigenetic mechanisms play a key role in establishing the link between environmental exposures and phenotype. The epigenetic changes involve changes in DNA methylation and histone modification, particularly in the promoter region of genes.

Yauk and colleagues⁴⁴ reported that hypermethylation of sperm DNA occurs in mice exposed to particulate air pollution in an urban/industrial location. With respect to environmental toxicants investigators have emphasized that when a pregnant female is exposed both the F1 embryo and the F2 generation germ cells are also exposed.

Reprogramming Somatic Nuclei to Generate Pluripotent Stem Cells

In 2007–2008 significant progress was made in development of methodologies to reprogram human and nonhuman primate somatic cells into pluripotent stem cells. In a review of

methods to achieve this reprogramming Byrne⁴⁵ noted that cell fate is determined by the epigenetic code, i.e., DNA methylation and histone modification. While early embryonic cells are pluripotent, subsequent epigenetic modifications result in differentiation to specific cell types. Methodologies developed recently have resulted in the generation of isogenic pluripotent stem cells from somatic cells. These cells may be cultivated outside the body, differentiated into specific cell types, and subsequently used for therapy in the donor from whom the somatic cells were derived. The isogenic cells will not generate an immune response and will not be rejected following transplant. Byrne reviewed methods for epigenetic reprogramming. These include factor-based reprogramming and somatic cell nuclear transfer.

Factor-Based Reprogramming

Studies by Yu and colleagues⁴⁶ revealed that four factors are sufficient to reprogram primate somatic cells to pluripotent stem cells. These factors include Oct4, Sox2, Nanog, and Lin28. Takahashi and colleagues⁴⁷ reported induction of pluripotent stem cells from adult human fibroblasts using four factors Oct4, Sox2, Nanog, and c-Myc. In subsequent experiments Nakagawa and colleagues⁴⁸ reported a modified protocol for generating pluripotent cells that did not require c-Myc. This is an important modification since reactivation of c-Myc virus used in the stem cells increases their tumorigenicity.

Nanog is a factor that is produced in preimplantation embryos in the inner cell mass. It is also produced by cultured embryonic stem cells. It serves as a selection factor for isolation of stem cells. Oct3 and Oct4 are members of the family of octamer-binding transcription factors that contain a POU domain. Sox2 is an SRY related HMG box transcription factor. Lin 28 is the human homolog of a protein isolated in *C. elegans* and defined as a regulator of developmental timing. It displays stage-specific expression during embryonic development and is

also expressed in mature tissues.⁴⁸ These factors were cloned into retroviral vectors and transfected into cells for reprogramming.

Yu and colleagues⁴⁹ reported that the pluripotent stem cells derived from skin fibroblasts and induced by Oct4, Sox2, Nanog, and Lin 28 have normal karyotypes, express telomerase, and cell surface markers characteristic of embryonic stem cells. Importantly, these cells have the capacity to differentiate into cells characteristic of the three primary germ layers of the embryo. Yu and colleagues analyzed methylation status and determined that the Oct4 promoter in the induced pluripotent was similar to that of embryonic cells and different from that of the parental fibroblasts.

An important consideration discussed by these investigators was that use of cloning vectors to introduce these factors into the genome of somatic cells might induce mutations at the integration site.

In addition to concerns that integration into human DNA of vectors that contain reprogramming factors may be harmful, there is also the concern that Oct4 is potentially oncogenic. Byrne⁴⁵ noted that while pluripotent cells derived through these methodologies may not be useful for therapeutic applications they may be useful for research. They could be utilized, for example, to analyze the downstream effects of a specific mutation encountered in human genetic disease in specific differentiated cell and tissue types. Additional research applications include study of drug metabolism in cells with a specific mutation or in individuals with a specific genotype.

Somatic Cell Nuclear Transfer into Oocytes

Somatic cell nuclear transfer into enucleated metaphase II oocytes represents another method of reprogramming to derive pluripotent cells. Byrne and colleagues⁵⁰ reported production of primate embryonic stem cells through use of this method. The oocytes cytoplasm epigenetically reprogrammed the

primate somatic nucleus to pluripotency and resulted in a cell type that could generate all three embryonic tissue types. Byrne reported that somatic cell nuclear transfer to oocytes generates very few pluripotent stem cells; 304 oocytes were used to generate 2 pluripotent stem cell lines.⁴⁵ He noted that one difference between isogenic pluripotent stem cells and embryonic stem cells involves early epigenetic changes. In cells resulting from somatic nuclear transfer the early stage DNA methylation is less complete and levels of histone acetylation are lower than in embryonic stem cells. Further studies have revealed that application of deacetylation inhibitors can induce more efficient epigenetic modification in somatic cell nuclear transfer-derived stem cells. Byrne emphasized the long-term potential of isogenic pluripotent stem cells for autologous transfer back to patients.

Daley and colleagues⁵¹ reported success in generating induced pluripotent stem cells from a number of patients with Mendelian genetic diseases and from patients with complex genetic diseases, e.g., diabetes mellitus type 1. They established cell lines from skin fibroblasts and in some cases, from bone marrow derived Mesenchymal cells. The patient-derived somatic cells were transduced using retroviruses expressing factors Oct4, Sox2, and Klf4 (Kruppel-like factor 4), and in some experiments retroviruses with cMyc and Nanog were used. In a subset of experiments the factor-encoding genes were co-infected into a retrovirus engineered to facilitate doxycycline-inducible gene expression.

Park and colleagues⁵¹ demonstrated pluripotency of colonies generated in the experiment described above. Colonies of the transduced somatic cells generated teratomas with tissues representative of all three embryonic germ layers including endoderm, ectoderm, and mesoderm.

Summary

1. Specific factors involved in the epigenetic reprogramming of somatic cell nu-

clei were identified by several different groups in 2007–2008.

2. In this same period, generation of isogenic pluripotent stem cells through transfer of somatic cell nuclei to oocytes was achieved for primate and human cells.

New Concepts Regarding Factors Involved in Sexual Differentiation

It is now 60 years after the groundbreaking studies of Jost,⁵² who determined that factors produced by the testes play a key role in sexual differentiation. He reported that the production of testosterone was essential for differentiation of external genitalia. He also established that production of anti-Mullerian hormone by the testis led to regression of the Mullerian duct, the requisite structure for development of the female oviduct, uterus, and vagina. In 1990 Sinclair and colleagues⁵³ reported discovery of the SRY, designated the key sex-determining factor produced by the testis. They proposed first that expression of SRY, encoded on the Y chromosome, in the supporting cells of the primitive gonad led to differentiation of Sertoli cells, the testosterone-producing cells. Second, they postulated that in the absence of SRY expression the supporting cells of the primitive gonad gave rise to the follicle cells of the ovary.

There is now evidence from studies on transgenic or knockout mice and from clinical studies that sex determination and sex reversal involves not only SRY genes but also a number of other genes, including autosomal loci. The exact roles of the X and Y chromosomes in sex determination have not yet been fully resolved.

In 1991 Palmer and colleagues⁵⁴ examined gonads in XX/XY chimeric mice and noted that the Sertoli cells in these gonads were predominantly but not exclusively XY. These observations led to a search for other extracellular factors that promoted Sertoli cell differentiation and proliferation. A number of studies on XX

hermaphrodites who had ovotestes and who did not have evidence of SRY challenged the exclusive role of SRY in testicular tissue differentiation.⁵⁵⁻⁵⁷ Another important observation was that SRY gene mutations were encountered in only a small percentage (10%) of XY female sex reversal cases.⁵⁸

The autosomal locus WNT4 encodes a factor required for normal sexual development. Jeays-Ward and colleagues⁵⁹ reported that partial female to male sex reversal occurred in Wnt4 mutant mice. Wnt4 normally represses male development in a female gonad. There is also evidence that Wnt4 enhances male development in an XY gonad.

An important autosomal gene required for sexual development is SOX9. Deletions or mutations in this gene lead to abnormalities in sexual development and skeletal deformities known as campomelic dysplasia. There is some evidence that expression of SOX9 is controlled by SRY. However, studies in SRY null mice have revealed that high levels of expression of Sox9 may be sufficient for male differentiation.⁶⁰ Based on their studies in knockout and transgenic mice Kim and colleagues⁶¹ reported that deletion of either Fgf9 or Fgfr2 causes male to female sex reversal. They noted further that FGFR2 plays a role in Sertoli cell differentiation and proliferation.

Additional clinical evidence for involvement of genes other than SRY and SOX9 in sexual development was published in 2007 and 2008. Temel and colleagues⁶² described a large consanguineous pedigree with nine individuals affected with 46XX testicular disorder of sexual development or ovotesticular disorder of sexual development. SRY was absent and the SOX9 haplotype differed in the affected individuals. Maciel-Guerra and colleagues⁶³ reported results of investigations on monozygotic twins with a disorder of sexual development. In one twin a testis was found in the other twin an ovotestis occurred. SRY was absent and no mutations, deletions, or duplication of SOX9 or DAX were found. The authors concluded that XX maleness XX ovotestes with disorder

of sexual development are manifestations of the same disorder of gonadal development.

Parma and colleagues⁶⁴ described a large consanguineous Italian family with a dermatological condition (palmoplantar hyperkeratosis and skin cancer predisposition) and female to male sex reversal. In this family there were four brothers with palmarplantar keratosis who had XX karyotypes. They carried out genetic linkage analysis and mapped this condition to human chromosome 1p34.3 and then analyzed candidate genes in this region. They identified mutation in the gene RSPO1 that encodes respondin 1. Subsequently they identified in a different individual with the same phenotype who had a homozygous deletion within RSPO1. They concluded that in humans altered expression of the respondin gene results in testis development in XX individuals.

There is now evidence that RSPO1 acts as a key regulator of beta catenin activation. Chasot and colleagues⁶⁵ found that ablation of these two genes leads to differentiation of seminiferous tubules in XX gonads. These investigators demonstrated that XX Rspo1 knockout mice show masculinized gonads and steroidogenesis. In these gonads XX germ cells fail to enter meiosis.

Rspo1 is required for Wnt4 expression. Rspo1 mutation apparently leads to downregulation of Wnt4/ catenin and Sox9 expression may then predominate. Maatouk and colleagues⁶⁶ proposed that in the XX gonad, WNT4 and RSPO1 expression activate the beta catenin canonical signaling pathway. Stabilized beta catenin expression then downregulates Sox9 expression and testis development.

In their review Di Napoli and Capel⁶⁷ concluded that the bipotential gonad is a battleground between two active and opposing signaling pathways that converge on the regulation of the SOX9/ FGF9 loop. They proposed that in early embryonic life in the bipotential gonad, male-promoting SOX9 and FGF9 and female-promoting Wnt4 and Rspo1 are balanced. The subsequent expression of SRY reinforces SOX9 and FGF9, and this drives

testis differentiation. In the XX gonad where SRY is absent WNT4 and RSPO1 promote ovarian differentiation.

True Hermaphroditism: Apparent Increased Incidence in Africa

Krob and colleagues⁶⁸ reported that true hermaphroditism is a rare cause of ambiguous sexual differentiation in Europe and North America, while in Africa it is one of the most common causes. Wiersma⁶⁹ reviewed clinical findings in 85 South African patients with true hermaphroditism. He referenced reports from South Africa indicating that over 51% of patients who presented with ambiguous genitalia were found to be true hermaphrodites. Wiersma noted that the high incidence of true hermaphroditism in the southern African black population is unusual and that no explanation for this high incidence has been forthcoming. Chromosome studies were carried out in 56 of the 85 patients he studied; 50 patients (89%) had a normal 46XX karyotypes and 6 patients were reported as 46XY. They identified no chimeric or mosaic chromosomal patterns. Ovotestes were found in the majority of patients (68 of 85). In some patients gonadal tissue was present that was described as ovarian; in other patients testicular tissue was present.

It is interesting to consider the possibility that mutations in one or more of the genes described above may play a role in true hermaphroditism and that the population frequency of a specific mutation may be higher in specific southern African populations. In this volume Ramsay and colleagues draw attention to the role of chimerism in etiology of some cases of true hermaphroditism.

Summary

1. There is evidence that SRY is but one of the factors important in male sexual differentiation. Furthermore it has become clear that X and Y chromosomes and autosomes are important. In fact at this point in time information on the role of the X

and Y chromosomes in sexual development is not clearly defined.

2. It is interesting to emphasize that important factors in sexual development have been defined through studies on transgenic mice and through clinical patient studies.

Duffy Blood Group Antigens and Malaria and HIV-AIDS Antigens

It is 40 years since Victor McKusick's research group⁷⁰ mapped the gene locus that encoded the Duffy blood group antigen to human chromosome 1. They noted, "We report what we believe is the first assignment of a specific gene locus to a specific autosome in man" (p. 949). It is fitting that we review discoveries about this locus in this issue of *The Year in Human and Medical Genetics* since, sadly, Victor McKusick passed away in 2008.

Evidence that enabled mapping of the Duffy locus to chromosome 1 came from cytogenetic studies. In a three-generation family the Duffy locus was found to segregate with a locus designated UN Uncoiled that led to reduced coiling of heterochromatin in the pericentromeric region of human chromosome 1. Supporting evidence for this chromosomal assignment came from another family where the Duffy blood group segregated with an inversion of chromosome 1.

The Duffy blood group was named for a hemophiliac patient who developed unusual antibodies following multiple transfusions. The blood group antigen that elicited these antibodies was defined as *Duffy a* (Fya+). Subsequently similar but different antibodies were identified in a woman who had had three pregnancies. The antigen that elicited antibodies in this patient was defined as *Duffy b* (Fyb+). The Duffy negative phenotype Fya-b-, is rare in Caucasians and Asians, and it is common in Africans. Race and Sanger⁷¹ reported that the Fya-b- genotype frequency was close to 0.8 in blood donors of African descent; it was close

to 0.01 in Caucasians. Subsequently, additional Duffy antigens were determined Fy3, Fy4, and Fy6.

Chaudhuri and colleagues (1989)⁷² purified an erythrocyte glycoprotein complex using antibody to Fy6. They reported that the complex isolated was multimeric, composed of different subunits and that only one of the subunits reacted with anti-Duffy antibody. They purified this subunit and derived peptides from it. In subsequent experiments Chaudhuri and colleagues⁷³ derived amino acid sequence from these peptides. They deduced nucleotide sequence and designed oligonucleotide primers to screen a cDNA library constructed from bone marrow mRNA. They also analyzed cDNA libraries from kidney, liver, and spleen. In the latter tissues they noted that an apparently full length mRNA could be isolated from Duffy positive and Duffy negative individuals; this indicated that the Duffy negative phenotype was not due to a gene deletion.

Tournamaille and colleagues⁷⁴ determined that a single base substitution G131A led to glycine to aspartate change at amino acid 44 of the Duffy protein and change from a to b specificity. Tournamaille and colleagues⁷⁵ determined that the Duffy negative genotype Fya-b- was determined by a single base substitution T to C in the promoter region of the gene that caused the gene not to be expressed on erythrocytes. This substitution disrupts the binding site for GATA1, an erythroid transcription factor. In Duffy positive and in Duffy negative individuals the Duffy antigen is expressed on endothelial cells of postcapillary venules and is found in liver, spleen, and kidney.

The molecular basis of a phenotype described as Fy^X, where there is weak expression of Fyb, was investigated by Yazdanbaksh and colleagues⁷⁶. This phenotype is absent in the black population and is present in 3.5% of whites. Fy^X: They determined that this variant is due to an Arg89Cys substitution in Fy protein. This substitution in the first intracellular loop leads to a protein that is less stable.

In 1976 Miller⁷⁷ postulated that Duffy antigen is required for infection of erythrocytes by *P. vivax*. They noted that in African populations where Duffy negative status predominates the frequency of this form of malaria is low.

Horuk in 1993⁷⁸ published results of studies on erythrocyte chemokine receptors that bind proinflammatory soluble peptides, including interleukin 8 IL8. These investigators noted that erythrocytes from the majority of African Americans studied did not bind IL8. Individuals whose erythrocytes failed to bind IL8 lacked the Duffy blood group antigen. Their studies indicated that Duffy antigen is a chemokine receptor.

Chakera and colleagues⁷⁹ reported that there is now evidence that Duffy antigen acts as a chemokine receptor and that it binds chemokines of the CC and CXC families. Duffy antigen is now known as Duffy antigen receptor for chemokines (DARC). Duffy encodes an unusual chemokine receptor. Chemokine receptors are most commonly members of the family of rhodopsin-like seven-transmembrane G-coupled receptor proteins (GPCRs). Duffy antigen lacks one specific consensus motif characteristic of GPCRs and there is no evidence that intracellular signaling takes place following binding of chemokine to Duffy antigen receptor. However, expression of this receptor is upregulated on red cells and on vascular endothelial cells during inflammation. The receptor binding face of the DARC protein has little polymorphism whereas the region opposite is highly polymorphic.

The CCR5 chemokine receptor is co-expressed with Duffy antigen (DARC). Chakera and colleagues⁷⁹ used bioluminescent resonant energy transfer studies to demonstrate, first that DARC occurs *in vivo* as a constitutive homodimer. Second, they demonstrated that DARC forms hetero-oligomers with the CCR5 chemokine receptor. The result of DARC CCR5 heterodimerization is inhibition of CCR5 signaling, impaired chemotaxis. DARC is a nonsignaling chemokine receptor that apparently modulates responses to

chemokines. Population genetic studies have revealed that the specific DARC promoter mutation that leads to Duffy negative status may impact predisposition to development of malaria induced by *P. vivax*. Lack of DARC expression on erythrocytes confers resistance to *P. vivax* and is common in African populations.

Earlier studies demonstrated that humoral immune responses against parasite antigens involved in erythrocyte invasion are important. There is now evidence that blood stage infection with *Plasmodium vivax* depends completely on interaction of a specific parasite protein *P. vivax* Duffy-binding protein PvDbp, with Duffy antigen on red blood cells. King and colleagues⁸⁰ determined that naturally acquired inhibitory antibodies that bind to PvDbp region II provide protection against *P. vivax* infection.

In New Guinea four human malaria species induce malaria in children. Artesunate is used to clear malaria parasite infection. Following this treatment the median time to *P. vivax* re-infection is 54 days, and it is 119 days for *P. falciparum*. King and colleagues⁸⁰ compared time to re-infection in individuals with high titers of antibodies against *P. vivax* and in individuals with low titers of such antibodies.⁸⁰ There was a significant delay in *P. vivax* re-infection in individuals who had high titers of antibodies.

A number of investigators have proposed that *P. vivax* Duffy-binding protein PvDBP is an excellent candidate for development of a vaccine. Effects of polymorphism in PvDBP on the activity of antibodies would need to be investigated.⁸¹

In 2008 He and colleagues⁸² reported that DARC influences plasma levels of HIV-1 suppressive and proinflammatory chemokines. They reported that there is a complex interplay between DARC and chemokine receptors CCR5, CCL2, and CXCL8 that impacts initial HIV virus binding and chemokine inflammatory response.

Summary

1. Presenting the history of mapping and analysis of Duffy blood group antigen

serves as a brief retrospective into the methodologies and insights that led to progress in molecular characterization.

2. It also illustrates the importance of taking population genetic differences into account in studies of the epidemiology of infectious diseases and in development of strategies for vaccine development.
3. In the coming years, analysis of the function of antigens and the function of protein identified through genomic analysis will provide interesting and exciting insights into physiology and pathology.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Kidd, J.M. *et al.* 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**: 56–64.
2. Perry, G.H. *et al.* 2007. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**: 1256–1260.
3. Dolan, M.J. *et al.* 2007. CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. *Nat. Immunol.* **8**: 1324–1336.
4. Willcocks, L.C. *et al.* 2008. Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J. Exp. Med.* **205**: 1573–1582.
5. Aitman, T.J. *et al.* 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* **439**: 851–855.
6. Groth, M. *et al.* 2008. High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. *Hum. Mutat.* May 9. [Epub ahead of print].
7. Hollox, E.J. *et al.* 2008. Psoriasis is associated with increased beta-defensin genomic copy number. *Nat. Genet.* **40**: 23–25.
8. Blauw, H.M. *et al.* 2008. Copy-number variation in sporadic amyotrophic lateral sclerosis: A genome-wide screen. *Lancet Neurol.* **7**: 319–326.
9. Abbott, A. 2008. Psychiatric genetics: The brains of the family. *Nature* **454**: 154–157.

10. McClellan, J.M., E. Susser & M.C. King. 2007. Schizophrenia: A common disease caused by multiple rare alleles. *Br. J. Psychiatry* **190**: 194–199.
11. Burmeister, M., M.G. McInnis & S. Zollner. 2008. Psychiatric genetics: Progress amid controversy. *Nat. Rev. Genet.* **9**: 527–540.
12. Walsh, T. *et al.* 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**: 539–543.
13. Xu, B. *et al.* 2008. Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat. Genet.* **40**: 880–885.
14. Fanous, A.H. *et al.* 2007. Association between the 5q31.1 gene neurogenin1 and schizophrenia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **144B**: 207–214.
15. Stefansson, H. *et al.* 2008. Large recurrent microdeletions associated with schizophrenia. *Nature* Jul 30. [Epub ahead of print].
16. Stone, J.L. *et al.* 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* Jul 30. [Epub ahead of print].
17. Stark, K.L. *et al.* 2008. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat. Genet.* **40**: 751–760.
18. Gothelf, D. *et al.* 2007. Risk factors for the emergence of psychotic disorders in adolescents with 22q11.2 deletion syndrome. *Am. J. Psychiatry* **164**: 663–669.
19. Talkowski, M.E. *et al.* 2008. A network of dopaminergic gene variations implicated as risk factors for schizophrenia. *Hum. Mol. Genet.* **17**: 747–758.
20. Barbato, C. *et al.* 2008. Thinking about RNA? MicroRNAs in the brain. *Mamm. Genome*. [Epub ahead of print].
21. Hagen, J.W. & E.C. Lai. 2008. MicroRNA control of cell-cell signaling during development and disease. *Cell Cycle* **7**: 2327–2332.
22. Selbach, M. *et al.* 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**: 58–63.
23. Conaco, C. *et al.* 2006. Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc. Natl. Acad. Sci. USA* **103**: 2422–2427.
24. Gao, F.B. 2008. Posttranscriptional control of neuronal development by microRNA networks. *Trends Neurosci.* **31**: 20–26.
25. Bicker, S. & G. Schratt. 2008. MicroRNAs: Tiny regulators of synapse function in development and disease. *J. Cell Mol. Med.* [Epub ahead of print].
26. Barco, A., C.H. Bailey & E.R. Kandell. 2006. Common molecular mechanisms in explicit and implicit memory. *J. Neurochem.* **97**: 1520–1533.
27. Li, Y., L. Lin & P. Jin. 2008. The microRNA pathway and fragile X mental retardation protein. *Biochim Biophys Acta*. Jul 18. [Epub ahead of print].
28. Perkins, D.O. *et al.* 2007. microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. *Genome. Biol.* **8**: R27.
29. Burton, P.R. 2007. Wellcome Trust Case Control Consortium Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* **39**: 1329–1337.
30. Baylin, S.B. & K.E. Schuebel. 2007. Genomic biology: The epigenomic era opens. *Nature* **448**: 548–549.
31. Felsenfeld, G. & M. Groudine. 2003. Controlling the double helix. *Nature* **421**: 448–453.
32. Bonetta, L. 2008. Epigenomics: Detailed analysis. *Nature* **454**: 795–798.
33. Mikkelsen, T.S. *et al.* 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553–560.
34. Barski, A. *et al.* 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837.
35. Meissner, A. *et al.* 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**: 766–770.
36. Hogarth, P., L. Lovrecic & D. Krainc. 2007. Sodium phenylbutyrate in Huntington's disease: A dose-finding study. *Mov. Disord.* **22**: 1962–1964.
37. Biacsi, R., D. Kumari & K. Usdin. 2008. SIRT1 inhibition alleviates gene silencing in Fragile X mental retardation syndrome. *PLoS Genet.* **4**: e1000017.
38. Rai, M. *et al.* 2008. HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. *PLoS ONE* **3**: e1958.
39. Jung, Y. *et al.* 2007. Potential advantages of DNA methyltransferase 1 (DNMT1)-targeted inhibition for cancer therapy. *J. Mol. Med.* **85**: 1137–1148.
40. Kovacheva, V.P. *et al.* 2007. Gestational choline deficiency causes global and Igf2 gene DNA hypermethylation by up-regulation of Dnmt1 expression. *J. Biol. Chem.* **282**: 31777–31788.
41. Barker, D.J. 1995. The fetal and infant origins of disease. *Eur. J. Clin. Invest.* **25**: 457–463.
42. Hanson, M.A. & P.D. Gluckman. 2008. Developmental origins of health and disease: New insights. *Basic Clin. Pharmacol. Toxicol.* **102**: 90–93.
43. Jirtle, R.L. & M.K. Skinner. 2007. Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* **8**: 253–262.
44. Yauk, C. *et al.* 2008. Germ-line mutations, DNA damage, and global hypermethylation in mice exposed to particulate air pollution in an urban/industrial location. *Proc. Natl. Acad. Sci. USA* **105**: 605–610.
45. Byrne, J.A. 2008. Generation of isogenic pluripotent stem cells. *Hum. Mol. Genet.* **17**: R37–R41.

46. Yu, J. *et al.* 2007. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**: 1917–1920.
47. Takahashi, K. *et al.* 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**: 861–872.
48. Nakagawa, M. *et al.* 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26**: 101–106.
49. Yu, J. & J.A. Thomson. 2008. Pluripotent stem cell lines. *Genes. Dev.* **22**: 1987–1997.
50. Byrne, J.A. *et al.* 2007. Producing primate embryonic stem cells by somatic cell nuclear transfer. *Nature* **450**: 497–502.
51. Park, I.H. *et al.* 2008. Disease-specific induced pluripotent stem cells. *Cell* Aug 6. [Epub ahead of print].
52. Jost, A. 1947. Recherches sur la différenciation sexuelle de l'embryon de lapin. III. Rôle des gonades foetales dans la différenciation sexuelle somatique. *Arch. Anat. Microsc. Morph. Exp.* **36**: 271–315.
53. Sinclair, A.H. *et al.* 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**: 240–244.
54. Palmer, S.J. & P.S. Burgoyne. 1991. In situ analysis of fetal, prepuberal and adult XX—XY chimaeric mouse testes: Sertoli cells are predominantly, but not exclusively, XY. *Development*. **112**: 265–268.
55. Spurdle, A.B., S. Shankman & M. Ramsay. 1995. XX true hermaphroditism in southern African blacks: Exclusion of SRY sequences and uniparental disomy of the X chromosome. *Am. J. Med. Genet.* **55**: 53–56.
56. Guerra Junior, G. *et al.* 1998. True hermaphrodites in the southeastern region of Brazil: a different cytogenetic and gonadal profile. *J. Pediatr. Endocrinol. Metab.* **11**: 519–524.
57. Domenice, S. *et al.* 2001. Molecular analysis of SRY gene in Brazilian 46,XX sex reversed patients: Absence of SRY sequence in gonadal tissue. *Med. Sci. Monit.* **7**: 238–241.
58. Jordan, B.K. *et al.* 2002. Familial mutation in the testis-determining gene SRY shared by an XY female and her normal father. *J. Clin. Endocrinol. Metab.* **87**: 3428–3432.
59. Jeays-Ward, K., M. Dandonneau & A. Swain. 2004. Wnt4 is required for proper male as well as female sexual development. *Dev. Biol.* **276**: 431–440.
60. Qin, Y. & C.E. Bishop. 2005. Sox9 is sufficient for functional testis development producing fertile male mice in the absence of Sry. *Hum. Mol. Genet.* **14**: 1221–1229.
61. Kim, Y. *et al.* 2007. Fibroblast growth factor receptor 2 regulates proliferation and Sertoli differentiation during male sex determination. *Proc. Natl. Acad. Sci. USA* **104**: 16558–16563.
62. Temel, S.G. *et al.* 2007. Extended pedigree with multiple cases of XX sex reversal in the absence of SRY and of a mutation at the SOX9 locus. *Sex Dev.* **1**: 24–34.
63. Maciel-Guerra, A.T. *et al.* 2008. XX maleness and XX true hermaphroditism in SRY-negative monozygotic twins: Additional evidence for a common origin. *J. Clin. Endocrinol. Metab.* **93**: 339–343.
64. Parma, P. *et al.* 2006. R-spondin1 is essential in sex determination, skin differentiation and malignancy. *Nat. Genet.* **38**: 1304–1309.
65. Chassot, A.A. *et al.* 2008. Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary. *Hum. Mol. Genet.* **17**: 1264–1277.
66. Maatouk, D.M. *et al.* 2008. Stabilization of {beta}-catenin in XY gonads causes male-to-female sex reversal. *Hum. Mol. Genet.* Jul 9. [Epub ahead of print].
67. DiNapoli, L., J. Batchvarov & B. Capel. 2006. FGF9 promotes survival of germ cells in the fetal testis. *Development* **133**: 1519–1527.
68. Krob, G., A. Braun & U. Kuhnle. 1994. True hermaphroditism: Geographical distribution, clinical findings, chromosomes and gonadal histology. *Eur. J. Pediatr.* **153**: 2–10.
69. Wiersma, R. 2004. True hermaphroditism in southern Africa: The clinical picture. *Pediatr. Surg. Int.* **20**: 363–368.
70. Donahue, R.P. *et al.* 1968. Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc. Natl. Acad. Sci. USA* **61**: 949–955.
71. Race, R. & R. Sanger. 1975. *Blood Groups in Man*, 6th edition. Publ. Blackwell, Oxford.
72. Chaudhuri, A. *et al.* 1989. Purification and characterization of an erythrocyte membrane protein complex carrying Duffy blood group antigenicity. Possible receptor for Plasmodium vivax and Plasmodium knowlesi malaria parasite. *J. Biol. Chem.* **264**: 13770–13774.
73. Chaudhuri, A. *et al.* 1993. Cloning of glycoprotein D cDNA, which encodes the major subunit of the Duffy blood group system and the receptor for the Plasmodium vivax malaria parasite. *Proc. Natl. Acad. Sci. USA* **90**: 10793–10797.
74. Tournamille, C. *et al.* 1995. Molecular basis and PCR-DNA typing of the Fya/fyb blood group polymorphism. *Hum. Genet.* **95**: 407–410.
75. Tournamille, C. *et al.* 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat. Genet.* **10**: 224–228.
76. Yazdanbakhsh, K. *et al.* 2000. Molecular mechanisms that lead to reduced expression of Duffy antigens. *Transfusion* **40**: 310–320.

77. Miller, L.H. *et al.* 1976. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype FyFy. *N. Engl. J. Med.* **295**: 302–304.
78. Horuk, R. *et al.* 1993. A receptor for the malarial parasite *Plasmodium vivax*: The erythrocyte chemokine receptor. *Science* **261**: 1182–1184.
79. Chakera, A. *et al.* 2008. The Duffy antigen/receptor for chemokines exists in an oligomeric form in living cells and functionally antagonizes CCR5 signaling through hetero-oligomerization. *Mol. Pharmacol.* **73**: 1362–1370.
80. King, C.L. *et al.* 2008. Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage *Plasmodium vivax* infection. *Proc. Natl. Acad. Sci. USA* **105**: 8363–8368.
81. Beeson, J.G. & B.S. Crabb. 2007. Towards a vaccine against *Plasmodium vivax* malaria. *PLoS Med.* **4**: e350.
82. He, W. *et al.* 2008. Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. *Cell Host. Microbe.* **4**: 52–62.