Published in final edited form as:

J Child Psychol Psychiatry. 2009 January ; 50(1-2): 63–71. doi:10.1111/j.1469-7610.2008.01978.x.

The future of genetics in psychology and psychiatry: Microarrays, genome-wide association, and non-coding RNA

Robert Plomin¹ and Oliver S. P. Davis¹

¹ Institute of Psychiatry, UK

Abstract

Background—Much of what we thought we knew about genetics needs to be modified in light of recent discoveries. What are the implications of these advances for identifying genes responsible for the high heritability of many behavioural disorders and dimensions in childhood?

Methods—Although quantitative genetics such as twin studies will continue to yield important findings, nothing will advance the field as much as identifying the specific genes responsible for heritability. Advances in molecular genetics have been driven by technology, especially DNA microarrays the size of a postage stamp that can genotype a million DNA markers simultaneously. DNA microarrays have led to a dramatic shift in research towards genome-wide association (GWA) studies. The ultimate goal of GWA is to sequence each individual's entire genome, which has begun to happen.

Results—GWA studies suggest that for most complex traits and common disorders genetic effects are much smaller than previously considered: The largest effects account for only 1% of the variance of quantitative traits. This finding implies that hundreds of genes are responsible for the heritability of behavioural problems in childhood, and that it will be difficult to identify reliably these genes of small effect. Another discovery with far-reaching implications for future genetic research is the importance of non-coding RNA (DNA transcribed into RNA but not translated into amino acid sequences) which redefines what the word *gene* means. Non-coding RNA underlines the need for a genome-wide approach that is not limited to the 2% of DNA responsible for traditional genes that are translated into amino acid sequences.

Conclusions—The only safe prediction is that the fast pace of genetic discoveries will continue and will increasingly affect research in child psychology and psychiatry. DNA microarrays will make it possible to use hundreds of genes to predict genetic risk and to use these sets of genes in top-down behavioural genomic research that explores developmental change and continuity, multivariate heterogeneity and co-morbidity, and gene-environment interaction and correlation. A crucial question is whether the prediction of genetic risk will be sufficiently robust to translate into genetically based diagnoses, personalised treatments, and prevention programmes.

Keywords

Microarray; genome-wide association; quantitative trait loci; non-coding RNA; behavioural genomics

Celebrating the 50th volume of this journal by looking forward rather than backwards is a daunting prospect in relation to genetics because the pace of discovery is so fast. To appreciate this point it is useful to glance backwards over the past century. The word 'gene' was coined just 100 years ago when Mendel's laws were rediscovered. It took 50 years to determine that DNA was the mechanism of heredity. However, the next 50 years witnessed the breaking of the DNA code, the ability to transfer genes from one species to another, and the complete sequencing of the more than 3 billion nucleotide bases of the human genome, which has led to an explosion of discoveries. Of the many advances in the past 5 years, we

focus on two that will shape genetic research in child psychology and psychiatry in the future: non-coding RNA and genome-wide association.

This journal itself reflects the increasing influence of genetic research in child psychology and psychiatry in recent years. The number of papers published in *JCPP* on genetics was 2 in 1960–69, 2 in 1970–79, 9 in 1980–89, 45 in 1990–99, and 74 in 2000–07. A *JCPP* paper in 1987 was about ‘the new genetics’ of molecular genetics which involved what was then a new way of assessing DNA markers (restriction fragment length polymorphisms) and their possible use in linkage analysis (McGuffin, 1987). As another indication of the pace of change, these advances have been superseded in the subsequent 20 years, as described below. A watershed for the incorporation of genetics was a 2005 special section of *JCPP* on molecular genetics (Eley, 2005). A discussion paper in this section considered new developments in genetics (Plomin, 2005), which is another indication of the pace of change because the two focal topics of the present article were just emerging in 2005. Similarly, this preview which was submitted in January 2008 will no doubt be overtaken by events by the time it is published.

Although non-coding RNA and genome-wide association will be the focus of our preview of what’s to come in genetic research, we begin by mentioning the important role that quantitative genetics will continue to play in the future of genetic research in child psychology and psychiatry.

Quantitative Genetics

The future of genetics belongs to molecular genetics in part because quantitative genetic research has made the case that genetic influences are important for most behavioural disorders and dimensions. Molecular genetics is now able to identify some of the specific DNA variants responsible for this heritability, and nothing will advance the field more than this (see below). Nonetheless, quantitative genetics such as twin studies will continue to play an important role in genetic research because it gives us the ‘bottom line’ of genetic influence regardless of how many genetic variants affect behavioural differences or how small and complex their effects might be. Much remains to be learned even about this rudimentary question of *how much* genetics affects many disorders (Plomin, DeFries, McClearn, & McGuffin, 2008). However, the greatest impact of quantitative genetics will come from research that goes beyond this basic question to investigate *how* genes have their effect. One example is the analysis of developmental change and continuity: Genetic differences between people account for most stability from age to age whereas environmental factors account for change. If most genetic action is involved in age-to-age stability, longitudinally stable phenotypes will be useful targets for molecular genetic research.

A second example is multivariate genetic research which suggests that the genetic structure of common disorders differs substantially from current diagnostic classifications based on symptoms, both for psychopathology (Kendler, Prescott, Myers, & Neale, 2003) and learning disabilities (Plomin & Kovas, 2005). Genetic effects appear to be general, resulting in co-morbidity rather than heterogeneity, whereas environmental effects are specific to each disorder. For example, for psychopathology, a review of 23 twin studies and 12 family studies concludes that anxiety and depression are largely the same disorder genetically and what differentiates the disorders is non-shared environment (Middeldorp, Cath, Van Dyck, Boomsma, Middeldorp et al., 2005). Going beyond anxiety and depression, multivariate genetic research suggests two broad genetic categories of psychopathology: internalising and externalising disorders (Kendler et al., 2003). In other words, ‘generalist’ genes may affect several disorders within major domains such as internalising disorders, externalising

disorders and learning disabilities, even though the same multivariate genetic research provides evidence for some specific genetic effects. These results suggest that generalist genes might be good targets for molecular genetic research within these domains (Davis, Kovas, Harlaar, Busfield, McMillan et al., 2008).

A third and final example concerns the interface between nature and nurture (Rutter, 2006). Every quantitative genetic study is as much a study of the environment as it is a study of genetics, whereas molecular genetic research is intrinsically limited to genetics. Quantitative genetic research provides the best available evidence for the importance of the environment because heritabilities for common disorders are never 100 percent and are usually 50 percent or less. It is at least as important to identify the environmental causes of psychopathology as it is genetic causes. A major clue from quantitative genetic research is that the salient environmental influences are non-shared, that is not shared by children growing up in the same family; identifying these non-shared environmental influences in genetically sensitive designs remains for future research (Plomin, Asbury, & Dunn, 2001; Turkheimer & Waldron, 2000). Another surprising finding from quantitative genetics is that many environmental measures widely used in the behavioural sciences show genetic influence (Jaffee & Price, 2007; Kendler & Baker, 2007). This finding suggests that people create their own experiences in part for genetic reasons, known as *genotype-environment correlation* because it refers to experiences that are correlated with genetic propensities. In addition, the effects of the environment can depend on genetics and the effects of genetics can depend on the environment, called *genotype-environment interaction*, genetic sensitivity to environments. Investigating the developmental interface between nurture and nature requires the incorporation of specific measures of the environment to assess nurture and will be greatly enhanced by including specific genes to assess nature (Caspi, Sugden, Moffitt, Taylor, Craig et al., 2003).

Non-coding RNA

Quantitative genetic theory and methods have not changed much in the past few decades; it is immune to the rapid changes in our understanding of molecular genetics because quantitative genetics focuses on the net influence of genes and environment, regardless of the complex mechanisms by which genes and environment have their effects. One sign of the speed of changes in molecular genetics in the past few years is that it is difficult even to say what a gene is (Gerstein, Bruce, Rozowsky, Zheng, Du et al., 2007). For the past 50 years, the so-called *central dogma* of molecular biology has been that a gene is a sequence of DNA that is transcribed into messenger RNA (mRNA) which is then translated into amino acid sequences, the building blocks of protein. We now know that there are only about 24,000 of these traditional genes and their protein-coding regions constitute only 2% of the 3 billion bases of DNA. However, nearly all of the several thousand single-gene disorders involve mutations in these coding regions, which appears to support the central dogma that the major output of the genome is protein from these coding regions.

What about the other 98% of DNA? Until the past decade, it had been thought that it was 'junk' that had just hitched a ride evolutionarily. A discovery with far-reaching implications is that more than half of all human DNA is transcribed into RNA but this RNA is not mRNA that is translated into amino acid sequences (Kapranov, Cawley, Drenkow, Bekiranov, Strausberg et al., 2002; Cheng, Kapranov, Drenkow, Dike, Brubaker et al., 2005). It is called *non-coding RNA* (ncRNA) in the sense that the DNA is transcribed into RNA but the RNA is not translated into amino acid sequences. It seems unlikely that such a complicated mechanism would have evolved unless it served some function. Indeed, the proportion of ncRNA is correlated with the complexity of species, whereas the number of traditional genes is not. For example, the number of traditional genes is similar in humans (~20–

25,000) and nematodes (~19,000), but the proportion of ncRNA differs greatly: 57% and 33%, respectively (Frith, Pheasant, & Mattick, 2005). In addition, there is some evidence that ncRNA may have evolved more rapidly in primates than traditional protein-coding genes (Pang, Frith, & Mattick, 2006; Pollard, Salama, Lambert, Lambot, Coppens et al., 2006).

It is now clear that such ncRNA plays an important role in regulating the expression of protein-coding DNA, especially in the human species (Mattick, 2007). One type of non-coding RNA has been known for 30 years. Embedded in protein-coding genes are DNA sequences called *introns* that are transcribed into RNA but are spliced out before messenger RNA leaves the nucleus. The remaining parts of genes are spliced back together and are called *exons*. Exons exit the nucleus and are translated into amino acid sequences. Exons usually consist of only a few hundred base pairs, but introns vary greatly, from 50 to over 1,000,000 base pairs, and some genes have hundreds of introns. Only exons are translated into amino acid sequences that make up proteins. However, introns are not 'junk'. In many cases ncRNAs embedded in the introns of protein-coding genes regulate the transcription of the gene in which they reside and in some cases they also regulate other genes (Rodriguez, Griffiths-Jones, Ashurst, & Bradley, 2004).

About a quarter of the human genome involves introns. An exciting recent finding of great significance is that a further quarter of the human genome is transcribed to RNA but not translated into protein (Mattick, 2004), and much of the genome is transcribed from both strands (Birney, Stamatoyannopoulos, Dutta, Guigo, Gingeras et al., 2007). This other RNA, which can be anywhere in the genome (not just near protein-coding genes), may also include ncRNA responsible for the regulation of protein-coding genes (Kapranov, Cheng, Dike, Nix, Duttagupta et al., 2007). This has led the way to a new world of regulatory networks across the genome and to many new targets for potential sources of genetic variation (Mattick, 2005; Mattick & Makunin, 2006). This finding suggests the need for a much broader view of the word *gene*.

One class of non-coding RNA that has attracted much attention arises from both intronic and intergenic RNA and is called *microRNA*. These tiny RNAs are usually only 21 base pairs long even though the DNA coding for them can be up to 6,000 base pairs in length (Saini, Griffiths-Jones, & Enright, 2007). Although small, microRNAs play a big role in gene regulation, especially in the development of the nervous system (Kosik, 2006; Cao, Yeo, Muotri, Kuwabara, & Gage, 2006) and especially in primates (Berezikov, Thummmler, van Laake, Kondova, Bontrop et al., 2006). More than 500 microRNAs have been catalogued to date that regulate protein-coding genes by binding to and thus silencing mRNA (Lim, Lau, Garrett-Engele, Grimson, Schelter et al., 2005). Amazingly, these 500 microRNAs appear to regulate the expression of more than a third of all coding mRNAs (Lewis, Burge, & Bartel, 2005). Moreover, microRNAs are likely to be just the tip of the iceberg of ncRNA effects on gene regulation (Mendes Soares & Valcarcel, 2006). The list of novel mechanisms by which non-coding RNAs regulate gene expression is growing rapidly (Costa, 2005; Huttenhofer, Schattner, & Polacek, 2005; Vasudevan, Tong, & Steitz, 2007).

However, microRNAs are not the only type of ncRNA. Aside from the well-known tRNAs, snRNAs and rRNAs involved in protein synthesis, there are snoRNAs involved in RNA modification, rasiRNAs that direct heterochromatin formation and Piwi-interacting piRNAs. In fact, there is a large and growing taxonomy of ncRNA genes ranging from the tiny miRNA at 21 nucleotides to the X-inactivating *XIST* in humans at 17,000 nucleotides and the massive *Air* in mice at 108,000 nucleotides long (Satterlee, Barbee, Jin, Krichevsky, Salama et al., 2007). The collected families of ncRNA (607 as of October 2007) are available at *Rfam*, the RNA families database (www.sanger.ac.uk/Software/Rfam/)

(Griffiths-Jones, Moxon, Marshall, Khanna, Eddy et al., 2005). Several ncRNAs have been found to be enriched in brain tissue and have been shown to be involved in brain morphogenesis, neuronal cell fate and differentiation, transcription of neuron-specific genes and possibly even learning and memory (Vo, Klein, Varlamova, Keller, Yamamoto et al., 2005).

Like protein-coding genes, all of these ncRNA genes vary in sequence between people, variations which could contribute to complex nervous system disorders. Indeed, ncRNAs have already been implicated in glioma, PraderWilli syndrome, spinocerebellar ataxia, polyglutamine expansion diseases, neurodegenerative disorders and Fragile X syndrome (Mehler & Mattick, 2007). Tourette's syndrome provides an interesting example because it involves an association between the disorder and polymorphisms, not in a ncRNA, but in a site in the *SLITRK1* gene where a microRNA binds to regulate translation (Abelson, Kwan, O'Roak, Baek, Stillman et al., 2005). This opens up the possibility that there will be many more cases where a single polymorphism could create or destroy a ncRNA binding site, altering the regulation of the associated gene. Recent studies have suggested that this is likely to be the case (Saunders, Liang, & Li, 2007).

The importance of ncRNA for the future of genetic research is that we may have been looking in the wrong place in the genome to find genes responsible for the high heritability of many disorders and dimensions of behaviour in childhood. As mentioned earlier, almost all of the thousands of monogenic disorders involve the 2% of DNA in coding regions (exons) of traditional genes, as suggested by the central dogma of molecular genetics. Although monogenic effects are severe and rare, it was reasonable to assume that more minor variations in these same coding regions would contribute to the heritability of common disorders and quantitative traits. However, this logic may have been wrong: The reason why monogenic disorders are in coding regions may be the reason why genes contributing to common disorders and quantitative traits are *not* in coding regions. Functional changes in these evolutionarily conserved coding regions of genes often have drastic consequences --the monogenic disorders that we see may in fact be the most benign forms of mutations in these regions, mutations that are at least viable, whereas most spontaneous mutations in these regions are lethal. In contrast, the heritability of common disorders and quantitative traits seems more likely to be caused by variations in 'genes' such as ncRNA that have quantitative regulatory effects.

A concrete implication of ncRNA is that we need to look beyond the 2% of DNA that codes for amino acid sequences. We need to look at the entire genome because we now know how little we know about what genes are, how they work and where to find them. ncRNA may explain in part why progress has been slow in identifying linkages and associations with common disorders and complex dimensions (see below). When linkages with DNA markers are found, the next step is to look for the coding DNA nearby that could explain the linkage. A classic example is the first replicable linkage found with a common disorder, which was reading disability (Cardon, Smith, Fulker, Kimberling, Pennington et al., 1994). Ten years and many coding regions later, the culprit gene has still not been identified even though some suspects remain in the line-up (Galaburda et al., 2006; Paracchini, Scerri & Monaco, 2007). Turning to association studies, they have focused on 'candidate' genes using DNA markers in coding regions, especially 'functional' DNA markers that code for amino acid sequence differences. However, progress has been slow in identifying replicable associations. Moreover, when associations are found with DNA markers in 'intergenic' regions, the associations have been dismissed as invalid for that reason.

Looking beyond the 2% of coding DNA is likely to help resolve these problems. That is, ncRNA in the other 98% of the genome may be responsible for the linkages and associations

that have eluded gene-hunters. An immediate impact of ncRNA is that the search is no longer gene-centred but genome-wide.

Microarrays and Genome-wide Association

Conceptual advances, including ncRNA, coupled with advances in technology such as microarrays, have led to a revolution in molecular genetic research: genome-wide association. Instead of only examining the 2% of DNA in coding regions, we can now investigate a million DNA polymorphisms distributed throughout the genome. Ultimately, each individual's entire genome will be sequenced so that polymorphisms of any kind can be detected. Individual genome sequencing is already happening: The genomes of two of the main protagonists in the race to sequence the generic human genome have been sequenced (Venter, 2007; Watson, 2007). Moreover, an international consortium called the '1000 Genomes' project has begun genome sequencing of 1000 individuals at the cost of several hundred thousand dollars each (<http://www.1000genomes.org/>).

Since 2007, genome-wide association (GWA) has revolutionised attempts to find DNA variation responsible for the high heritability of many common disorders. The genome-wide feature of GWA means that it is hypothesis-free in the sense that it is not limited to DNA in coding regions, nor is it limited to ncRNA. This hypothesis-free aspect of GWA is important because new sources of DNA variation continue to be discovered. For example, in addition to single-nucleotide polymorphisms (SNPs) and simple-sequence repeat (SSR) polymorphisms, which have been widely used in gene hunting, it is now known that a major source of variation comes from structural variation in DNA such as deletions and duplications, called *copy-number variation* (Redon, Ishikawa, Fitch, Feuk, Perry et al., 2006). CNVs occur throughout the genome and vary widely in the population (Pinto, Marshall, Feuk, & Scherer, 2007); their potential importance has been highlighted in recent GWA reports (Estivill & Armengol, 2007). A survey of the human genome identified more than 3000 CNVs, 800 of which appeared at a frequency of at least 3 percent (Wong, deLeeuw, Dosanjh, Kimm, Cheng et al., 2007). Remarkably, CNVs result in the genomes of individuals differing by as much as 10 million base pairs. Given the speed of discoveries in molecular genetics following from the Human Genome Project, it is likely that many other surprises are in store, which underlines the need for a genome-wide approach.

The costs of individual genome sequencing are falling rapidly but are still a long way from the goal of a \$1000 genome (Service, 2006). As an interim step, microarrays have made possible parallel genotyping of a million SNPs, the most common type of DNA polymorphism, as well as quantifying a million non-polymorphic DNA sequences useful for detecting CNVs, for less than \$500. A microarray is a slide the size of a postage stamp that is dotted with short single-stranded DNA sequences called probes. It detects SNPs in the usual way: Fluorescently labelled single-stranded DNA from an individual is allowed to hybridise with a probe which will only happen if there is an exact match. The critical difference with microarray analysis is that an individual's entire genome is first chopped into small pieces, amplified (genome-wide amplification) and labelled with a fluorescent tag, so that the individual's genotypes can be determined simultaneously for millions of SNPs. Not only can microarrays be used to probe the entire genome, they can also be used to assess gene expression levels and to map ncRNA by indexing RNA transcripts, both protein-coding and non-coding (the transcriptome). An important direction for genetic research in the future will be to relate the genome to the transcriptome (Plomin & Schalkwyk, 2007).

In just a year's time, GWA studies have come to dominate the gene-hunting literature. The 'A' of GWA is association, which is simply a correlation between a particular allele and a trait in the population. The trait can be a qualitative trait like an ADHD diagnosis or a

quantitative trait like hyperactivity symptom scores. For example, one of the best documented associations involves a particular allele of the apolipoprotein E gene on chromosome 19 which is associated with late-onset Alzheimer's disease. In dozens of association studies, the frequency of allele 4 was found to be about 40 percent in individuals with Alzheimer's disease and about 15 percent in controls. Although association is a correlation between DNA sequence variation and a trait, this correlation implies causation because DNA sequence is not changed by behaviour or the environment.

Association studies have been reported for decades but now, rather than looking at a few 'candidate' genes, GWA can examine SNPs systematically across the genome. Before GWA, linkage analysis, which traces co-inheritance of a DNA marker and a trait within families, was able to scan the genome with only a few hundred DNA markers but could only detect genes of large effect size. Although linkage analysis was successful in identifying genes responsible for rare single-gene disorders, linkage analyses of common disorders and complex traits generally came up empty handed in terms of replicable linkages. This finding suggested that genetic influence on common disorders and complex traits may be due to multiple genes of small effect. Association is complementary to linkage in that it can detect much smaller effect sizes if sample sizes are large. However, unlike linkage, hundreds of thousands of DNA markers are needed to conduct association analyses across the genome. Such a project was inconceivable a few years ago because genotyping 500,000 SNPs in 1000 cases and 1000 controls would have required a billion genotypings, which would cost many millions of dollars. Now with microarrays such a study would cost a few hundred thousand dollars.

The dramatic shift to GWA research was triggered by an early GWA study which identified an association of large effect size for age-related macular degeneration (Klein, Zeiss, Chew, Tsai, Sackler et al., 2005). Although the GWA sample was very small by current standards (96 cases, 50 controls) and only 100,000 SNPs were genotyped, it identified an association with an odds ratio of 4.7 and a population attributable risk of about 50%. The frequency of the risk allele was 72% in cases and 36% in controls. The association was accepted for three related reasons: it was a huge effect, it was immediately replicated in three other independent studies, and it was in a gene (complement factor H) that made sense.

This stunning result created a surge of GWA studies using the new microarray technology that made GWA feasible. The first of these reported an association between SNPs in the interleukin-23 gene and inflammatory bowel disease using a 300,000-SNP microarray with about 500 cases and 500 controls (Duerr, Taylor, Brant, Rioux, Silverberg et al., 2006). This association was much smaller than the association found for age-related macular degeneration; the odds ratio was about 1.5 between cases and controls in two independent samples with about 500 individuals in each group. Nonetheless, the association has been replicated in several subsequent studies (Cho & Weaver, 2007).

GWA studies in other domains yielded many associations but of even smaller effect size which means that they will prove more difficult to replicate. The initial surge of GWA research culminated in the groundbreaking Wellcome Trust Case Control Consortium (WTCCC) which reported GWA results for 500,000 SNPs genotyped for seven common disorders each with 2000 cases and 3000 shared controls in collaboration with 50 research teams (WTCCC, 2007). More than 20 significant associations were reported across the disorders, including one for bipolar disorder, the only psychiatric disorder among the seven disorders.

However, the largest of these effect sizes is very small. For example, the significant WTCCC association for bipolar disorder was based on an allele frequency of 25% in cases

and 28% in controls, although effect sizes for some of the other disorders were greater. The most important implication of the many GWA studies to date is that there are few 'low-hanging fruit' for common disorders other than age-related macular degeneration. The largest effect sizes generally involve replicated relative risks of about 1.2 comparing cases and controls; for dimensional analyses, the largest associations explain about 1% of the variance in the population. For example, these are the effect sizes of a common variant in the *FTO* gene which has been significantly associated with obesity and body mass index in 14 studies (Frayling, Timpson, Weedon, Zeggini, Freathy et al., 2007). In our own GWA studies of reading and general cognitive ability across the normal distribution, our largest effect sizes are less than .5% of the variance (Butcher, Davis, Craig, & Plomin, 2008; Meaburn, Harlaar, Craig, Schalkwyk, & Plomin, 2007).

If the largest effect sizes are so small, hundreds of such associations will be needed to account for the heritability of a common disorder or quantitative trait. If these associations of very small effect size can be detected, it would not matter if trait-specific microarrays included hundreds or even thousands of SNPs in order to predict various aspects of a trait, its developmental changes, its interactions and correlations with the environment and all of these in different populations (Janssens, Aulchenko, Elefante, Borsboom, Steyerberg et al., 2006; Khoury, Little, Gwinn, & Ioannidis, 2007). Microarrays could be the new diagnostic manuals of the future, at least in relation to genetic risk.

The big question is whether we will be able to detect such small effects reliably. There is reason for optimism as the dust begins to settle from the initial GWA explosion. For example, GWA studies have so far used microarrays with one, three or five hundred thousand SNPs; new microarrays are now available with one million SNPs, thus providing greater coverage of the genome. In addition, these microarrays include one million non-polymorphic probes useful for detecting copy number variation mentioned earlier. Even for SNPs, the focus has been on common variants in which the minor allele has a frequency of at least 1%. Most SNPs are less common and such rare SNPs and other types of rare polymorphisms may cumulatively contribute to heritability, although obtaining power to detect rare polymorphisms will be challenging (Bearden, Glahn, Lee, Chiang, van Erp et al., 2008; Ji, Foo, O'Roak, Zhao, Larson et al., 2008).

Another promising direction for research is to focus on polymorphisms that are known to be functional. For example, a microarray is now in use that includes 15,000 SNPs that result in a change of an amino acid during translation. Functional polymorphisms are important because they add considerable power in GWA research by testing a direct association between the polymorphism and the trait rather than relying on indirect association via neighbouring polymorphisms. Eventually, individual sequencing will make it possible to examine polymorphisms of any kind. An obvious way to increase power to detect small effects is to increase sample sizes. However, rather than designing a single definitive study to detect associations of very small effect size, it might be necessary to use a series of replication studies to winnow the small kernels of grain from the chaff, side-stepping some of the daunting issues involved in correcting for multiple testing in a single study. Another way to increase power is to use quantitative genetic research to be more precise about how to carve nature at its joints, including gene-gene and gene-environment interactions.

What good will come from identifying genes if they have such small effect sizes? One answer is that it is possible to study pathways between each gene and behaviour. Even for genes with very small effect on behaviour, the road map is clearly marked for bottom-up analyses that begin with gene expression, although the roads quickly divide and become more difficult to follow to higher levels of analysis in the brain and the pathophysiology of behavioural disorders. Animal models will be increasingly important in bottom-up research

that traces pathways from genes through the brain because of the powerful techniques available to investigate the effects of gene expression on the brain. Although animal models of behaviour are challenging, animal models will be especially valuable when it comes to understand how genes work from the bottom up through the brain. Even if there are many genes of small effect, each gene is a tiny window that offers a peek at the brain's functioning. Moreover, looking through many of these windows simultaneously could lead to a systems approach to the brain because each of the many genes was identified on the basis of its association with a behavioural disorder. Similarly, sets of genes associated with behavioural disorders will also be useful in top-down analyses that begin with behaviour and investigate multivariate, developmental and genotype-environment interface issues, and translate these findings into gene-based diagnosis and personalised treatment as well as prediction and prevention of disorders (Harlaar, Butcher, Meaburn, Craig, & Plomin, 2005).

The speed of discovery in genetics is now so great that it would be impossible to predict what will happen in the next five years, let alone the next fifty years. Most geneticists would agree with Francis Collins (2006), the director of the US National Human Genome Research Institute and leader in the Human Genome Project, who expects that we will each have an electronic chip with our DNA sequence. Individual DNA chips will herald a revolution in personalized medicine in which treatment is individually tailored rather than one-size-fits-all. The greatest value of DNA lies in its ability to predict genetic risk which can lead to preventative interventions. That is, rather than treating problems after they occur, DNA will make it possible to predict problems and to intervene to prevent problems. Genetics can help to target children at genetic risk who are most likely to profit from interventions, which is important because successful prevention programmes usually require extensive and intensive, and thus expensive, interventions. Interventions could conceivably involve genetic engineering that alters DNA although so far gene therapy in the human species has proven difficult even for single-gene disorders. Behavioural engineering and environmental engineering are more like to pay off in preventing behavioural problems affected by many genes as well as many environmental factors (Goncalves, 2005).

What about the ethical issues raised by finding genes associated with child psychology and psychiatry? For example, could trait-specific microarrays be used to select 'designer babies' prenatally or to label children postnatally? The fear lurking in the shadows of these concerns is that finding genes associated with behaviour will limit our freedom and our free will. In large part, such fears involve misunderstandings about how genes affect complex traits (Rutter & Plomin, 1997). Moreover, knowledge alone does not account for societal and political decisions. Values are just as important in the decision-making process; decisions both good and bad can be made with or without knowledge. Depending on our values, genetics could be used to argue for devoting more resources to help genetically disadvantaged children. Indeed, genetics makes this view more palatable because it avoids assigning blame for behavioural problems solely to environmental failures of the school and family. We need to be cautious and to think about societal implications and ethical issues. But there is also much to celebrate here in terms of the increased potential for understanding child psychology and psychiatry.

Main Points of Review

- Quantitative genetic research such as twin studies will continue to make important developmental, multivariate and genotype-environment contributions to psychopathology. These issues, especially correlations and interactions between genes and environment, will be addressed with much greater precision when specific genes are identified in molecular genetic research.

- Non-coding RNA changes what we mean by the word *gene* and underlines the need for genome-wide approaches to gene hunting.
- Since 2007, genome-wide association has come to dominate gene hunting using microarrays that allow parallel genotyping of millions of DNA markers.
- The first wave of genome-wide association studies indicates that for common disorders and quantitative traits the largest associations are very small.
- The challenge is to identify sets of genes, each of small effect, that can be useful in predicting and preventing problems in childhood.

Acknowledgments

Preparation of this paper was supported in part by grants from the Medical Research Council (G050079), the Wellcome Trust (GR75492) and the U.S. National Institute of Child Health and Human Development (HD44454, HD49861, HD46167). OSPD is supported by an MRC studentship.

Abbreviations

GWA	Genome-wide association
QTL	quantitative trait locus
SNP	single-nucleotide polymorphism
ncRNA	non-coding RNA

Reference List

- Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, et al. Sequence variants in *SLITRK1* are associated with Tourette's syndrome. *Science*. 2005; 310:317–320. [PubMed: 16224024]
- Bearden CE, Glahn DC, Lee AD, Chiang MC, van Erp TG, Cannon TD, et al. Neural phenotypes of common and rare genetic variants. *Biological Psychiatry*. 2008 Advanced online publication. 10.1016/j.biopsych.2008.02.005
- Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, et al. Diversity of microRNAs in human and chimpanzee brain. *Nature Genetics*. 2006; 38:1375–1377. [PubMed: 17072315]
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*. 2007; 447:799–816. [PubMed: 17571346]
- Butcher LM, Davis OSP, Craig IW, Plomin R. Genome-wide quantitative trait locus association scan of general cognitive ability using pooled DNA and 500K single nucleotide polymorphism microarrays. *Genes, Brain and Behavior*. 2008; 7:435–446.
- Cao X, Yeo G, Muotri AR, Kuwabara T, Gage FH. Noncoding RNAs in the mammalian central nervous system. *Annual Review of Neuroscience*. 2006; 29:77–103.
- Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. Quantitative trait locus for reading disability on chromosome 6. *Science*. 1994; 266:276–279. [PubMed: 7939663]
- Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*. 2003; 301:386–389. [PubMed: 12869766]
- Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, et al. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*. 2005; 308:1149–1154. [PubMed: 15790807]

- Cho JH, Weaver CT. The genetics of inflammatory bowel disease. *Gastroenterology*. 2007; 133:1327–1339. [PubMed: 17919503]
- Costa FF. Non-coding RNAs: new players in eukaryotic biology. *Gene*. 2005; 357:83–94. [PubMed: 16111837]
- Davis OSP, Kovas Y, Harlaar N, Busfield P, McMillan A, Frances J, et al. Generalist genes and the internet generation: etiology of learning abilities by web testing at age 10. *Genes, Brain and Behavior*. 2008; 7:455–462.
- Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006; 314:1461–1463. [PubMed: 17068223]
- Eley TC. Special section on molecular genetics. *Journal of Child Psychology & Psychiatry*. 2005; 46:1029.
- Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genetics*. 2007; 3:1787–1799. [PubMed: 17953491]
- Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007; 316:889–894. [PubMed: 17434869]
- Frith MC, Pheasant M, Mattick JS. The amazing complexity of the human transcriptome. *European Journal of Human Genetics*. 2005; 13:894–897. [PubMed: 15970949]
- Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD. From genes to behavior in developmental dyslexia. *Nature Neuroscience*. 2006; 9:1213–1217.
- Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korb J, et al. What is a gene, post-ENCODE? History and updated definition. *Genome Research*. 2007; 17:669–681. [PubMed: 17567988]
- Goncalves MA. A concise peer into the background, initial thoughts and practices of human gene therapy. *BioEssays*. 2005; 27:506–517. [PubMed: 15832383]
- Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research*. 2005; 33:D121–D124. [PubMed: 15608160]
- Harlaar N, Butcher L, Meaburn E, Craig IW, Plomin R. A behavioural genomic analysis of DNA markers associated with general cognitive ability in 7-year-olds. *Journal of Child Psychology and Psychiatry*. 2005; 46:1097–1107. [PubMed: 16178934]
- Huttenhofer A, Schattner P, Polacek N. Non-coding RNAs: hope or hype? *Trends in Genetics*. 2005; 21:289–297. [PubMed: 15851066]
- Jaffee SR, Price TS. Gene-environment correlations: a review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*. 2007; 12:432–442. [PubMed: 17453060]
- Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. Predictive testing for complex diseases using multiple genes: fact or fiction? *Genetics in medicine*. 2006; 8:395–400. [PubMed: 16845271]
- Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nature Genetics*. 2008; 40:592–599. [PubMed: 18391953]
- Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, et al. Large-scale transcriptional activity in chromosomes 21 and 22. *Science*. 2002; 296:916–919. [PubMed: 11988577]
- Kapranov P, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*. 2007; 316:1484–1488. [PubMed: 17510325]
- Kendler KS, Baker JH. Genetic influences on measures of the environment: a systematic review. *Psychological Medicine*. 2007; 37:615–626. [PubMed: 17176502]
- Kendler KS, Prescott CA, Myers J, Neale MC. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry*. 2003; 60:929–937. [PubMed: 12963675]

- Khoury MJ, Little J, Gwinn M, Ioannidis JP. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *International Journal of Epidemiology*. 2007; 36:439–445. [PubMed: 17182636]
- Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005; 308:385–389. [PubMed: 15761122]
- Kosik KS. The neuronal microRNA system. *Nature Reviews Neuroscience*. 2006; 7:911–920.
- Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005; 120:15–20. [PubMed: 15652477]
- Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, Castle J, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005; 433:769–773. [PubMed: 15685193]
- Mattick JS. RNA regulation: a new genetics? *Nature Reviews Genetics*. 2004; 5:316–323.
- Mattick JS. The functional genomics of noncoding RNA. *Science*. 2005; 309:1527–1528. [PubMed: 16141063]
- Mattick JS. A new paradigm for developmental biology. *Journal of Experimental Biology*. 2007; 210:1526–1547. [PubMed: 17449818]
- Mattick JS, Makunin IV. Non-coding RNA. *Human Molecular Genetics*. 2006; 15:R17–R29. [PubMed: 16651366]
- McGuffin P. The new genetics and childhood psychiatric disorder. *Journal of Child Psychology and Psychiatry*. 1987; 28:215–222. [PubMed: 3473069]
- Meaburn EL, Harlaar N, Craig IW, Schalkwyk LC, Plomin R. Quantitative trait locus association scan of early reading disability and ability using pooled DNA and 100K SNP microarrays in a sample of 5760 children. *Molecular Psychiatry*. 2007 Advanced online publication. 10.1038/sj.mp.4002063
- Mehler MF, Mattick JS. Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiological Reviews*. 2007; 87:799–823. [PubMed: 17615389]
- Mendes Soares LM, Valcarcel J. The expanding transcriptome: the genome as the ‘Book of Sand’. *EMBO Journal*. 2006; 25:923–931. [PubMed: 16511566]
- Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI, Middeldorp CM, Cath DC, et al. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. *Psychological Medicine*. 2005; 35:611–624. [PubMed: 15918338]
- Pang KC, Frith MC, Mattick JS. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics*. 2006; 22:1–5. [PubMed: 16290135]
- Paracchini S, Scerri T, Monaco AP. The genetic lexicon of dyslexia. *Annual Review of Genomics and Human Genetics*. 2007; 8:57–79.
- Pinto D, Marshall C, Feuk L, Scherer SW. Copy-number variation in control population cohorts. *Human Molecular Genetics*. 2007; 16:R168–R173. [PubMed: 17911159]
- Plomin, R.; DeFries, JC.; McClearn, GE.; McGuffin, P. Behavioral genetics. 5. New York: Worth Publishers; 2008.
- Plomin R. Finding genes in child psychology and psychiatry: when are we going to be there? *Journal of Child Psychology and Psychiatry*. 2005; 46:1030–8. [PubMed: 16178925]
- Plomin R, Asbury K, Dunn J. Why are children in the same family so different? Nonshared environment a decade later. *Canadian Journal of Psychiatry*. 2001; 46:225–233.
- Plomin R, Kovas Y. Generalist genes and learning disabilities. *Psychological Bulletin*. 2005; 131:592–617. [PubMed: 16060804]
- Plomin R, Schalkwyk LC. Microarrays. *Developmental Science*. 2007; 10:19–23. [PubMed: 17181694]
- Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*. 2006; 443:167–172. [PubMed: 16915236]

- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006; 444:444–454. [PubMed: 17122850]
- Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. *Genome Research*. 2004; 14:1902–1910. [PubMed: 15364901]
- Rutter, M. *Genes and behavior: nature-nurture interplay explained*. Oxford: Blackwell Publishing; 2006.
- Rutter M, Plomin R. Opportunities for psychiatry from genetic findings. *British Journal of Psychiatry*. 1997; 171:209–219. [PubMed: 9337969]
- Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences USA*. 2007; 104:17719–17724.
- Satterlee JS, Barbee S, Jin P, Krichevsky A, Salama S, Schratt G, et al. Noncoding RNAs in the brain. *Journal of Neuroscience*. 2007; 27:11856–11859. [PubMed: 17978024]
- Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and microRNA target sites. *Proceedings of the National Academy of Sciences USA*. 2007; 104:3300–3305.
- Service RF. Gene sequencing. The race for the \$1000 genome. *Science*. 2006; 311:1544–1546. [PubMed: 16543431]
- Turkheimer E, Waldron M. Nonshared environment: a theoretical, methodological, and quantitative review. *Psychological Bulletin*. 2000; 126:78–108. [PubMed: 10668351]
- Vasudevan S, Tong Y, Steitz JA. Switching from repression to activation: microRNAs can up-regulate translation. *Science*. 2007; 318:1931–1934. [PubMed: 18048652]
- Venter, JC. *A life decoded: my genome: my life*. London: Allen Lane; 2007.
- Vo N, Klein ME, Varlamova O, Keller DM, Yamamoto T, Goodman RH, et al. A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. *Proceedings of the National Academy of Sciences USA*. 2005; 102:16426–16431.
- Watson, JD. *Avoid boring people: and other lessons from a life in Science*. Oxford: Oxford University Press; 2007.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
- Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, et al. A comprehensive analysis of common copy-number variations in the human genome. *American Journal of Human Genetics*. 2007; 80:91–104. [PubMed: 17160897]