Fishing for Genetic Links in Autism

This is the second installment in a 2-part series that addresses approaches to understanding the molecular underpinnings of autism.

In my January column (“Fishing Expeditions and Autism: A Big Catch for Genetic Research?” Psychiatric Times, January 2009, page 12), I described the great difficulties researchers face characterizing the genetic basis of the disease. Complexities range from trying to establish a stable diagnostic profile to making sense of the few isolated mutations that show clear associations (either with disease or syndrome variants).

Using the metaphor of a fishing net, I discussed 2 overall research strategies that geneticists commonly use to catch these elusive sequences of interest. One strategy is to cast nets that act like large purse seiners to collect many sequences in a single (and usually quite expensive) effort. The other strategy is akin to dropping a single fishing line into the genetic waters to see if anything “bites.” In Part I, I described one particularly successful strategy that snagged a large number of useful sequences.

Here, the focus narrows. I will not describe the isolation of many sequences, but rather only one. Our “catch” is called MeCP2, a gene whose mutations can give rise to a wide spectrum of related postnatal neurodevelopmental disorders—including autism spectrum disorders. I will now examine some background regions about gene regulation, move to the biological functions of MeCP2, and then focus on studies in animal models that provide tantalizing hints about the origins of autistic behavior. My goal is to show that research progress in autism is a continuum of efforts, ranging from large projects with lots of identifiable sequences to small projects that focus on the properties of single genes.

**Gene typologies and their regulation***

There is a lot of heavy-duty molecular biology behind MeCP2. Getting the clearest view requires us to review 4 pieces of background information. Feel free to skip to the section “MeCP2 and Rett syndrome” if Class II genes and CpG islands are working parts of your vocabulary.

Gene expression. As you recall from your undergraduate days, genes are broken down into 3 classes. Class I genes encode the information necessary to make ribosomal RNAs. Class II genes encode the information to make mRNA, and these genes are in the distinct minority (only about 2% of activatable sequences). Class III genes encode transfer RNAs.

Class II genes can be broken down into 2 functional parts. The first part includes the nucleotides that are necessary to encode the protein, which are called the “structural sequence.” The second part, which often lies in front of the gene, is called the “promoter.” Promoters act like tiny on-off switches that either allow or block the manufacture of the cognate message.

How is that message made? The enzyme complex that creates the message is called “RNA polymerase II.” Not all genes are transcriptionally active at the same time, and some never become activated at all (eg, neurons do not do the same job as, say, skeletal muscles, and have very different activation profiles). Understanding how the RNA polymerase II complex knows which gene to turn on in a complex cellular environment has been a focus of intense investigation for decades.

There can be many mechanisms to help this complex decide which gene to turn on. RNA polymerase II enzyme can be shown which genes it needs to turn on and which are supposed to be left alone. There are escort-like proteins that physically bind to the RNA polymerase II and guide the complex to its proper genetic destination. Another mechanism involves proteins that bind to the gene to be transcribed rather than to the RNA polymerase II. These proteins then act like homing beacons, guiding a wandering RNA polymerase II to its proper nucleotide destination. There are also repressor proteins that work in a similar but opposite fashion. They render a gene that could be activated into a repressed, transcriptionally inert state.

**Histone protein complexes.** Histones are groups of proteins around which DNA wraps, like twine around a ball. There are many of these wound-up balls along a chromosome that function like physical barriers. If the DNA that harbors a class II gene is wrapped around the histone complex, RNA polymerase II can have a very difficult time binding to it, and the gene is rendered silent. If the histone complex is bulldozed out of the way, the class II gene becomes available for activation.

Chromatin is a combination of DNA and histones: this mass of molecules can form surprisingly complex, higher-ordered structures. The structures are so specific that antibodies capable of binding to them can be created. The structures can then be individually isolated intact through a protocol called “chromatin immunoprecipitation.”

**Methyl groups and methylation reactions.** Methylation reactions involve adding a methyl group (CH3) to certain nucleotides in the double helix. If the methylation occurs on or near a class II gene, it can be rendered inactive.

How does this work? The methyl groups studding along the length of a gene often attract repressor proteins that perform the actual silencing function. Such repressor proteins are attracted to a given segment of DNA via sequences on the DNA called “CpGis” (sometimes referred to as CpG islands), which are short regions of DNA enriched for cytosine and guanine nucleotides (the “p” refers to the phosphodiester bond between cytosine and guanine). CpG islands often occur in clusters at the promoter sites within the class II gene. CREB.

The last piece of background information is the biology of the CREB protein (which stands for the tongue-twisting name “cyclic AMP response element binding”). The characterization of CREB function is one of the great research achievements in all of molecular neurobiology. The reason? CREB is involved in activating the genes that take part in learning, and it does so in virtually every animal ever tested (including humans). It specifically activates gene sequences involved in establishing memory formation by binding to their promoter regions and transcriptionally activating the gene. CREB is a classic activator. Understanding the genes to which CREB normally binds has resulted in the isolation of many sequences involved in human learning. Understanding CREB biology remains a subject of intense interest and plays a powerful role in our story.

**MeCP2 and Rett syndrome**

With this admittedly lengthy background information, we return to the biology of the MeCP2 protein and its role in autistic spectrum disorders.

The MeCP2 gene was first characterized by researchers who were interested in Rett syndrome. An X-linked dominant disease, Rett syndrome affects 1 in 10,000 persons and is found mainly in females. Symptoms usually present within the first 6 to 18 months of life and include motor and speech difficulties, seizures, increasing cognitive impairment, and growth retardation. About half of those affected eventually become nonambulatory and many have chronic GI disorders.

Several types of mutations have been characterized, from deletions and insertions to subtle point mutations (changes in single base pairs). Most germane to our story, MeCP2 mutations were eventually associated with certain autism spectrum disorders. While this association hardly represents the overarching genetic explanation for even a single autistic category, the finding was important. The function of MeCP2 is well known, and has been established in animal models that carry MeCP2 mutations. Having such a well-characterized ally could be useful in the understanding of autism.

MeCP2 is expressed in all the body’s cells, but some of its highest concentration of activities is in the CNS. The hypothalamus, at least in laboratory animals, is particularly robust. The MeCP2 binds to DNA that has been previously methylated. Indeed, MeCP2 literally means “methyl CpG binding protein.” This binding functions as a gene-silencing mechanism and, until recently, MeCP2 was considered to be a canonical repressor protein. As we shall see, this job description turned out to be overly simplistic.

**Animal studies and the 40,000-foot view**

As mentioned, a number of animal models have been used to investigate the molecular biology of MeCP2. Animals were genetically engineered to not express MeCP2 (“knock-out animals”) or to overexpress it. By examining gene expression profiles in spe...
cific tissues, it is possible to evaluate the “40,000-foot” view of the regulatory effects of MeCP2. The researchers did so by examining tissues in the hypothalamus.

The studies showed that MeCP2 did not just affect the expression of 1 or 2 genes but rather thousands. In some cases, the overexpression of MeCP2 turned off gene sequences in the hypothalamus. This is something you might predict, given MeCP2’s more traditional role as a repressor. However, in other cases, the overexpression actually turned on certain genes, which was not expected.

How can an overexpressed represor actually turn on a gene? The explanation turned out to be more conceptual than biological. MeCP2 was not always acting as a repressor. Sometimes it was an activator. Indeed, it was shown to be an activator 85% of the time!

The same expected/unexpected results were observed in the MeCP2-deficient animals (knock-outs). The absence of MeCP2 clearly caused some genes to overexpress (something you might predict if their “repressor” was missing). But in some cases, certain genes were actually turned off in backgrounds without MeCP2. This would only make sense if MeCP2 activated these genes.

Commonalities
The 40,000-foot view gives us a lot of information about general expression patterns, but it does not identify individual sequences. Researchers next sought to narrow down the choices by asking which genes are being turned on, which ones are being turned off, and what, if anything, does this have to do with cognition?

The researchers used 2 overall thrusts to answer these questions. The first thrust had to do with trying to discern any common patterns of activation and repression in the thousands of genes MeCP2 appeared to be regulating. The second involved naming the specific proteins to which MeCP2 might be binding. The researchers hit pay dirt on both accounts.

MeCP2 appeared to activate any gene whose promoters were greatly enriched in CpG islands. MeCP2 also seemed to repress genes whose promoters were not greatly enriched in CpG islands. A counterintuitive finding was also made regarding the number of methyl groups in the activating promoters. Close examination of these promoters revealed that the CpG islands were not heavily methylated. The odd combination (lots of chances to methylate but few actual methyl groups) seemed to “trick” the MeCP2 into activating the gene.

Close examination of the genes that MeCP2 repressed also revealed a counterintuitive finding. These sequences had heavily methylated islands, even though there were not many islands in total. That odd combination seemed to convince MeCP2 to turn off the gene. Research continues on many such genes in an attempt to understand the specific actions of this talented protein’s global reach (Figure).

But what about cognition? Rett syndrome and autistic spectrum disorders are phenomena that affect the ability to process specific kinds of information. Were any molecular interactions observed in these experiments that might give hints to these effects on cognitive behaviors? The answer to this question was yes, and it was the most exciting part of the work.

Narrowing the choices
The MeCP2 protein was certainly binding things in island-specific ways. But to what molecules, among hundreds of thousands possible, is it associating? Could the characterization of its partners lead to a greater understanding of the role of MeCP2 in Rett syndrome and autism?

It is possible to answer at least the binding parts of these questions in 2 ways. First, you can use a mass spectrometer to identify proteins associated with your target (as long as you are clever about stabilizing native molecular associations in specific tissues in your sample). Mass spectrometry can identify molecules based on their mass-to-charge ratios (in essence, you chemically fragment your sample into ions and then calculate the mass-to-charge ratio by passing them through a series of electric and magnetic fields). The researchers uncovered a whopper—MeCP2 was associating with CREB. A molecule whose mutations were known to be involved in autistic behaviors was actually binding to a molecule like CREB, known to be involved universally with information processing!

Researchers confirmed this using the chromatin immunoprecipitation protocol. This technique (described previously) involves isolating the histone-DNA-regulatory protein combinations by using antibodies that are capable of binding to ordered chromatin structures. The researchers were able to show that CREB was involved in half the promoters to which MeCP2 was binding. Specifically, CREB was associated with MeCP2 at the promoters where MeCP2 was an activator. Remarkably, CREB was completely absent at promoters that MeCP2 was known to be repressing. In other words, MeCP2 was specifically associating with molecules that normally associated with learning and was activating the genes to which both bind! Not only did this confirm the data from the mass spectrometry, it also extended and refined the results (see Figure).

Conclusions
Clearly, MeCP2 interacts with a

Figure

MeCP2 and neurodevelopmental disorders

From Rett syndrome to autism spectrum disorders, mutations in the MeCP2 gene are involved in a wide variety of neurodevelopmental disorders. As shown below, the protein is involved in both the activation and repression of specific gene sequences. Depending on its target structure, MeCP2 can associate with proteins involved in learning.

Two choices
MeCP2 protein can act either as an activator or repressor of target gene sequences. Which choice it makes depends in part on the structure of the target.

Implications
These structural conditions profoundly affect the binding properties of CREB and may even influence cognition.

MeCP2

CREB

Promoter

Structural gene

Numerous CpG (C) islands carrying many CpG islands in their promoters, and...

Lightly methylated (M)
...the islands, although capable of being fully methylated, are only lightly methylated.

MeCP2

Promoter

Structural gene

Few CpG (C)
genes with promoters carrying few CpG islands in their promoters, and...

Heavily methylated (H)
...the islands are heavily methylated.

MeCP2

Promoter

Structural gene

The target structure allows CREB and MeCP2 to bind to the promoter

The target structure does not allow CREB and MeCP2 to bind to the promoter

(Please see Autism, page 42)
broad swath of the transcription regulatory machinery available inside brain cells, both to activate and repress. It specifically interacts with CREB, a protein whose role in learning is preserved over a broad phylogenetic range.

Like all good data, however, these raise as many questions as they answer. At what level are these interactions occurring? Do they interact directly with RNA polymerase II at the promoter, thus exerting more local effects? Do they help in remodeling the higher-order chromatin structure, thus exerting more global effects? Does it do both? Most important, how do the deficits associated with Rett syndrome and autism come from the mutations observed in MeCP2? These questions will represent the next generation of experiments.

A number of caveats must be mentioned, of course. First, these findings have been shown in laboratory animals, not in humans, and the normal grumpy cautions are in order (these are somewhat assuaged because of the remarkable phylogenetic conservation of CREB-mediating learning responses). Second, these data come from the examination of the hypothalamus. But what about medial temporal lobe structures and other brain regions, such as the forebrain?

None of these questions tarnish the data. They simply contextualize the first—and, in my opinion, most remarkable—association between a known learning protein and a mutation involved in a known cognitive pathology.

And with such boundaries, we have come full circle. I hope that after reading both columns, you can see more clearly the edge of our understanding regarding autism and autism-related disorders. The purse seiner approach described in Part 1, combined with the narrower fishing pole approach described here, represents some truly amazing progress in the attempts to understand these baffling diseases.

It is a heck of a time to be dropping lines in these waters!

Dr Medina is a developmental molecular biologist and private consultant, with research interests in the genetics of psychiatric disorders.