

Complement Factor H
29-Apr-06 - 1:00 pm - 4:30 pm

Margaret Pericak-Vance: All right. I too would like to thank the organizers, and I'm especially thrilled that I get to talk about the history, and how this gene was found. I think as I go along in the talk you'll see why they switched the order, and I actually followed Jonathan in the presentations. Okay. Finding or dissecting genetics of AMD really posed a challenge, and I work on a couple other disorders besides macular degenerations since I'm a Genetic Epidemiologist including Alzheimer's where we had found the ApoE risk factor, and autism. And if you'd asked me five years ago, which of these three disorders we'd be more likely to be successful with, I would have ranked AMD as at the bottom, because I thought this was going to be an incredibly difficult disease to make any sense of genetically. Even though I did not put that on my grant applications, I remained optimistic. But anyways, what are some of those reasons? First of all the delayed oxidative of AMD, the inheritance of AMD is often not obvious. The symptoms occur, as we all know late in life. The patient's parents, siblings, often as siblings as well are often deceased. So you don't have a really good pedigree structure that you have with some other diseases. And then when you go to get a family history, when you ask did your mother or father, or whoever have AMD, historical accounts and the family history's are often inaccurate. The sibling risk ratio for AMD is approximately estimated between two and four. And to put that in perspective now, the sibling risk ratio for Alzheimer's is about the same, whereas the sibling risk ratio for something like autism is about fifty to seventy-five. So even though that sibling risk ratio isn't that high, it's high enough that you would expect to be able to find some genetic influences in your studies. Another thing about AMD, and this is why I thought it was going to be especially difficult is the known environmental influences will confound genetic analysis. So as my two predecessors have said, AMD is a complex disease where likely multiple genes and environmental factors contribute to its etiology. And these factors may work independently or as we're seeing in some of the more recent publications through complex interactions. So you would think with the fleur of the media, and everything that occurred that all the research on AMD genetics happened early last year. When in reality studies have been going on for quite awhile, and that's why I'd like to talk about some of the whole genome linkage screens that were going on before the major breakthrough with the CFH. Especially with respect to chromosome one, Klein et al, had published a single large pedigree that they showed linkage in, and they had a significant log score, and that log – in that large pedigree as in 1998. There are also several linkage screens in, again my predecessors had eluded to these different screens, and I think I have about eight listed there that looked at various types of the phenotypes of AMD. And these had occurred all before identification of the CFHG. And then right about the same time as the identification of the CFH, there was a medical analysis of the above linkage screens trying to show which regions were most consistent across the various studies. So with respect to chromosome one when you looked at these eight or nine linkage screens, chromosome one was consistent across studies, but even though there was this consistency, and everybody said we think there's a gene there. Our linkage evidence is really, really strong on chromosome one. There was still no consensus candidate. The other thing is chromosome one is only one of multiple linkage regions that people were looking at. It wasn't like it was the only peak on

these linkage screens, as you'll see in a minute. There were several peaks, and those in - the ones with the stars indicate the most interesting peaks. Chromosome one as you see is one of them. But another one that was of interest was chromosome ten. And even though we won't be talking about it, it's been alluded to in terms of the loc gene today. So there were several regions to look at, chromosome one being one of the main ones. So why were we able to be successful, and I think the question that Jonathan just asked sort of, and his answer alluded to this. First of all there were new technology and resources. There were these polymorphism called SNPs. They were abundant across the genome. There were new approaches. These approaches that allowed us to do high resolution follow up mapping of these linkage peaks. As well as hold genome association, and new analytical approaches. Methods that were allowed us to sift through this massive data that we were now able to generate from these families in order to locate the one or few genetic variants involved in AMD, and the hundreds of thousands of possibilities. And of course the recognition of our funding agencies that more funding support was needed in order to do studies of this particular magnitude. So I know you saw this slide before, and Jonathan and I tried to coordinate what we were presenting, and these are the four approaches that he talked about in terms of, in terms of finding the disease gene discovery and complex traits. And what I'm going to do now is talk about them, and put them in a little bit more detail in terms of the CFH story. And I realize that most of the authors of these papers are in the audience so hopefully I've been able to pull from them the most important points to convey to you guys. Okay so whole genome association, and this genome-wide association was one of the papers in Science by Klein et al., interesting enough they only used ninety-six cases and fifty controls from the Arid Study. They did not use the expanded five hundred thousand SNPs now that's available. They use one hundred thousand SNPs, which at that time was a huge number of markers. And the results ended up them finding one SNP non-coding that met conservative Bonferroni correction, and this was at 1q31, which was the linkage region that you've seen, you saw in the previous slides. They found the second non-coding SNP that wasn't quite as significant, but it was fairly significant and located nearby. So what they did next after they mapped this, and they had this very significant SNP, is they use the HapMap, one of the tools from the genome project to determine the patterns of linkage to equilibrium. And they found that both of these SNPs, the highly significant one that met Bonferroni correction, as well as the other SNP fell within the compliment factor HG. And this just shows you the raw, the data that came out of their association analysis. Here's the SNP that met Bonferroni correction, and that indeed was in the CFH complex. So they had the SNP. They realized that this wasn't itself the functional polymorphism, so what did they do to try to find it? Well, they performed some re-sequencing. They found the coding SNP, the now famous rs1061170 in exon 9 that resulted in the tyrosine-histidine change, and it's usually referred to as Y402H, that was significantly associated with AMD. They also included some immuno-fluorescence experiments showing that the CFH protein was present in human retina. When they looked at their data for this particular marker—the Y402H—when they looked at individuals who had at least one copy of the risk allele, their odds ratio was 4.6, and for individuals with two copies of the risk allele, it was 7.4. So their conclusion was: the involvement of CFH supports role of information in AMD, and, indeed, Y402H was the risk polymorphism involved. So this was the whole genome WGA approach.

On the other hand, the second approach that Dr. Haines talked about was the approach that we used in our joint study between Vanderbilt and Duke. And this was also published in the *Science* trilogy. What we had was a data set. We had a data set that we specifically designed that was family-based and case-controlled. So we had two independent data sets. The family-based were made of multiplexed families, as well as a set of discordant sib pairs. The case control was 500 singleton cases and 200 examined controls. We had previously, together with Mike Gordon, performed genomic screen through CIDR₁ and along with some of the regions of interest that everyone else had found, was our chromosome 1 data. And in our own particular subset of the data, we also had linkage to chromosome 1, so this seemed like a reasonable linkage area for us to follow up on.

So we took the chromosome 1 region of interest from the genome screen and we followed up using SNP association methods in these two independent data sets that we had. And this just shows you that when we did this, we genotyped – we started off, I think, by genotyping 40 SNPs over a larger region. We, then, kept adding SNPs into an iterative process and when we finally added the SNPs in and compared it in the two data sets, the family data set, as well as the case-controlled data set, it became obvious that, even though this entire region showed significant evidence for association, the only gene where the two data sets where everything converged, both the case control, which is the red, and the family-based converged, was in CFH. So out of these various complement genes in here, we focused on the CFH. And what we did is we screened—looked at the complete variation in 24 cases and 24 controls from our data set—and when we did this, we found this famous RS1061170, or Y402H, was very different in terms of the cases versus the controls.

So when we looked at this in the entire data set, the entire family-based, which was smaller than our case-controlled data set, we had a very significant P-value, and felt this, as the previous group, was indeed the functional polymorphism of interest. Similarly we found odds ratios of 2.45 if you had one risk allele, versus 3.3. This is for all AMD. When we looked at just the neovascular, it was higher.

Moving on so I don't run out of time, I'm going to talk a little bit about the functional approach, and this was published by Edwards and colleagues in the same *Science* trilogy. They zeroed in on the ARMD1 linkage region on chromosome 1 and they used a staged approach where they had two data sets: a discovery data set, as they called it, and then the replication set of cases and controls. And, as Jonathan had alluded to in his talk, they identified 24 non-synonymous coding SNPs across the region where ARMD1 locus—that locus originally reported by Klein in that large family--mapped. And they found significant association only with the genes in the RCA, so they were able to zero in very rapidly to the region that they wanted to look at. And this just shows you—this is from their paper—where even though there is significant association across the region, you really see the majority of the significant results are in CFH. It looks very similar, almost, to the plot we have in our paper.

So, they then went from their discovery sample to their replication sample, and they also find Y402H to be most consistent result. And their relative risk of having at least one C allele, with either the CC or the CT genotype, was 2.7.

Finally, looking at CFH as a candidate gene approach—this is published by Hageman in *PNAS*. Again, why they were interested in this is because the studies where it was involved in drusen, there's this gene, this kidney disease, MPGM-2, where there's a patient reported with a mutation in it, and that maps exactly to the region, CHF maps to the region of linkage in AMD, so they put all these together and it became an outstanding candidate for them to look at. They had several data sets. They screened for variants, and what was interesting—and I'll move along here—about their paper is they not only looked at Y402H by itself, they looked at haplotypes, and what they found that there was not only a risk haplotype involved at CFH and AMD, there were also protective haplotypes, and they were the first group that actually said, “now it's not as simple as we think; there's more going on here and we really need to explore this further.”

Some additional reports – there have been several additional reports in the past year. There have been some reports that looked at neovascular versus geographic atrophy forum showing that it was involved in risk in both. There have been reports on smoking and CFH where, even though they both continue to be consistent risk factors, there's really been no interaction between the two observed to date.

An interesting study that I just read and so I thought I'd include it, it was a study of CFH in Japan where they really did not see any association with Y402H in the cases and control data that they looked at; however, they did find two risk haplotypes were found, as well as a protective haplotype. So they found association with CFH, they just didn't find the association with Y402H, which was the presumed functional mutation. So they concluded that risk alleles and haplotypes may vary among the different ethnic groups, and so there may actually be more that we have to look at in terms of trying to determine how CFH is important. Is it on different backgrounds that make a difference?

Anyway...so, in conclusion. AMD and CFH really divine some new paradigm for dissecting complex traits. It really is the Alzheimer's of this decade and it uses everything in all the resources that we garnered from the human genome initiative in order to give us a story for successfully working on the etiology of genetics in AMD. And I'll stop there. Are there questions?

Q: In individuals with MPGN, were known to have drusen since at least 1987, and also a case was described for MPGN that had lack of factor H in 1986. Why did it take so long to look at factor H?

A: This is a question I've asked myself many times, as soon as I heard that story. I mean, I just think that, you know, we're talking about a kidney disease, which, I guess, many of us don't think of if we're thinking of eye diseases. I don't know why, because I agree that once you've heard the story and you knew about the linkage, it made a perfect example that you should look there. And I applaud the group that went the functional way and looked at that. I agree with you. I wish I had an answer. We should have found it ten years ago?

Thank you.