My name is Ada Hamosh, and I'm going to be speaking to you about genetic resources on the web. I'm going to focus most of this lecture on two resources — Online Mendelian Inheritance of Man and Gene Test, but I'll also tell you about some educational resources as we go along.

OMIM, or Online Mendelian Inheritance of Man, and the first line is the URL, is a national database that's funded by the National Human Gene and Resources Institute and authored and maintained here at Johns Hopkins. It's fully integrated into the NCBI entree suite, and I'll tell you much more about that later, and it's the world's oldest and most comprehensive genetics resource. I have to tell you up front that I run it, so I'm biased. This is the home page for OMIM.

In the beginning, there was Mendel, and then there was McKusick, and Dr. McKusick started keeping notes to himself in the late 1950's about the relationship between genes and disorders, though little was known about genes at the time, of course, and the first published issue was in 1966 with 1,400 entries, and the last book version was in 1998 and went to three volumes as you can see in this photo.

We are currently in the exponential phase of gene and disorder discovery, particularly the relationship between the two and, as of this moment, there are 15,195 entries within OMIM. We catalog both disorders and genes and, as of this moment, there are 2,458 loci with one or more known disorders and 3,493 disorders that have been mapped by a variety of different mechanisms which you see. There are, at the moment, 1,553 genes with at least one known point mutation, causing either a disease or cancer, and there are 2,418 disorders that map back to these 1,553 genes. OMIM also catalogues mutations, but not all mutations and, as of this moment, there are 12, 144 mutations catalogued within OMIM. There are about 14 people who work full-time on OMIM, and then a variety of other writers around the world who author on OMIM. OMIM is entirely literature based, and most people write based on the journal, and some people write for OMIM based on particular fields that they cover for us.

This is the OMIM home page, and every MIM entry has a unique six-digit identifier, and you can find out what that is by looking at numbering system which is circled in red there. The 100,000 series were all autosomal dominant conditions, and those entries were created before May 15, 1994, and the 200,000 series are the autosomal recessive conditions. As of May 15, 1994, we decided that many disorders, that were both recessive and dominant, were caused by mutations in the same gene and, for that reason, now all autosomal entries created after May 15, 1994, have a 600,000. The 300,000 series is for X-linked conditions, 400,000 for Y-linked, and the mitochondrial loci phenotypes have a 500,000 series.

From the OMIM home page, if you would like to know how many entries we're up to, you can click on Statistics, which will take you to this page, and gives you the answer for today, and it's also divided into the types of entries and whether they're X-linked, Y-linked or autosomal or mitochondrial. So, an asterisk means that it's an established gene; a number sign,

or pound sign, means it's a phenotype description and the description of a gene resides elsewhere; and an entry with no prefix means either that the genetic basis of that condition is unknown, or its independence from another entry is not clear.

If you go to the month, you will get the most recent entry, and that's today's, and that's if you want to know whether there's anything new or exciting in the gene that you're studying, or whether there's anything new or exciting going on in general.

You can look at OMIM's summary of the gene map. It's not every known gene; it's the genes that are kept within OMIM, by clicking on the search, the gene map. In this case, I've typed in CFTR because it happens to be my favorite gene, and this is the result that you get. So, the first is a cytogenetic location for CFTR; the second field is the official gene symbol and other aliases; the next is the title for that gene; the next is the link for the MIM, to go to the MIM entry regarding the gene; the next are disorders associated with mutations in that gene; and the numbers there will take you directly to the MIM description of those disorders. Next is a common field which can have a variety of different information in it.

Next is the method of mapping, which I will not discuss here. And, finally, is whether there is any mouse homology that's known. If you click on the cytogenetic location, it will take you to the NCBI map viewer of the genome, which gives you much, much more detail about the map location. You can zoom in or out; you can find out, you can make different maps that you see depending....finally, you can search simply the morbid map by alphabetical order. So, the morbid map is the link of disorders and locations within the genome and, in this case, what I've done is I've typed in long QT, and I can see that there are a variety of different long QT syndromes that the MIM number next to the name of the syndrome is the description of the syndrome, the phenotype. The MIM number under the title OMIM is the MIM number, the entry that describes the gene. Okay. Finally, you can always type in any search you want right along the white bar, but if you wanted to do a detailed or limited search, you can click on search OMIM, and that will take you to this page, and then you can limit your search in whatever way you like.

Now, let's say you were seeing a patient who has pulmonic stenosis and you look at this patient and you see that they also have down slanting fissures and a web neck and you're wondering what syndrome could this be, besides just the pulmonic stenosis , and so, you type that in, right across the white bar. You don't have to do anything fancy — just type. And what you will get is a long search. But the first thing is Noonan Syndrome , and you vaguely remember something about Noonan Syndrome from medical school, so you want to find out a little bit more about Noonan Syndrome. You click on that link, and that takes you to this entry. This entry actually goes on for several pages, and you have this patient in front of you, and you don't have all the time in the world, so if you click on Clinical Synopsis, it'll give you the clinical synopsis of that disorder. So, in this case, it gives you, it always starts with inheritance and growth, and then head and neck, and all the way down the body. And so here, if you page through, you can see that pulmonic stenosis is a feature of this condition, so it fits, and you continue to page through the clinical synopsis and, at the end, you will find that

there's a molecular basis. And, so there's actually, the cause of this, at least for some patients, is known.

Now, go back to the home for this entry, Noonan Syndrome, and one of the most important things you must always do when you're using the web to do any research is to know how timely the information is that you're looking at. So, you could see the dates for the previous entry that it was, 2002 is the last time that we added anything to the clinical synopsis, and that was based on knowing the molecular basis. That's why that was added at that time. But if you go to this, to the contributor field for any MIM entry, you'll see the date that any thing was last added, and that's extremely important to do with any web based search. Okay.

Now, so you read about Noonan Syndrome. You think, wow, this patient really could have Noonan Syndrome. What do we know about the gene and mutations in the gene? So, you can click on that, that link right there, and that will take you to the gene entry. As you read about the gene entry, you can say well, you know, again, this patient's in front of me, I don't have time. Let me find out what we know about the mutations, and you can click on View List of allelic variance. And you'll get this list, and you see it goes on, and it goes on beyond what I've captured here. And you say, you know, I don't have time. I just need to know if there's a clinical test available. So, that you can actually answer immediately by going out to a different site, but you can go to that site and to the relevant page of that site, putting you directly within OMIM. So, if you click on Gene Test, you'll go to the Gene Test Search for Noonan Syndrome, and that you can see on the bottom of the page, it says Noonan Syndrome. It has a testing button, a research button, and a reviews button. The testing button will take you to the clinical testing that's available. Research takes you to research testing that's available, and the reviews will take you to review, and I'll show you all of those in a moment.

So, if you click on the testing, we return to this screen, and this a very busy and extremely useful new screen that Gene Test puts up. I have nothing to do with Gene Test, so I can rave about them independently. This tells you all the different laboratories that are doing genetic testing, and it tells you the type of testing that they're doing. And you can see, depending on whether you want to do prenatal testing, or whether you want a comprehensive full gene analysis or specific mutations if they're common, you can pick. If you click on the hot link within the, just below the name of the lab, you'll get gene test information about that lab — whether they're certified , who does the testing, who the contact people are, and then if you click on the link at the top, that'll actually take you to the home page for that specific lab, if that's your interest. Okay.

Now, let me tell you a little bit more about Gene Test. So, this is gene test URL; it's actually much easier than OMIM's to remember. They are also funded by NIH, through NHGRI. They are also funded by the National Library of Medicine, the Department of Energy, and through a HRSA grant. They are maintained at the University of Washington and they're directed by Roberta Pagon .

This is the Gene Test home page, and there are several important things to know about this home page. Number one is, if you look at that box in the green, you will see that there are currently 1,036 diseases for which there are 688 that there's clinical testing available, so being a clinical geneticist is really much more complicated than it used to be. In addition, there are 348 research tests, and when they say research only, that's today. Tomorrow, it'll be fewer, or there will be more research tests and more clinical tests. So, it's something that changes constantly.

Now, if you click on the Gene Reviews, which is where that red arrow points, that will take you to this page, and I've typed in Noonan here to find out about what kind, whether they have a gene review for Noonan. So, a gene review is an overview of a topic. This is what that search captures, and now I'm going to go to the review, and the gene review is an overview of a topic; it always has a structured, it has a structure that all reviews have to adhere to. It's written from the perspective of differential diagnosis and molecular genetics. And there's some discussion of treatment, but not very much in any of these genetic resources that I'm talking about because the rate of change of treatment is so fast that most web-based resources don't discuss treatment, for genetics anyway.

So, this is the home page for this, and you can see that there's always a structure. You can go wherever you want. There's a printable copy. Then, there are highlighted words. If you click on those highlighted words, you get the definition. Okay? And you can actually go to the full glossary which is very detailed and excellent if you aren't familiar with these terms. It's a very good place to get definitions of these terms. I always disable the glossary, which is something that I can do. And you can, too, if you're comfortable with them.

If you click on the Clinic Directory, that will take you to a list of clinics available not only around the United States, but around the world where genetic evaluation and testing are available. And this can be very useful because, of course, genetic diseases run in families, and family members don't all live in the same small community and, so, if you've diagnosed Noonan Syndrome in your patient, and they have three adult children that are distributed across the United States, you can find out what other clinic might see them. So, in this case, what I've done is I've gone to this site, and you can see you can search by state, by the kind of services available, or specific specialty clinics, and I just typed in Maryland, and it gives me a map of Maryland, so I can go to which city the family's closest to, and then click on different clinics that are available.

Now the final thing they have is a site called Educational Materials within this site, and that is a very nice place to find out detailed information about the principle of genetic testing. It's very important to know what a positive test result means and what a negative test result means. Now, of course, we all know what that means clinically, in clinical medicine but, for genetic testing, and particularly for molecular genetic testing, those can be very tricky pieces of information. A negative result might not mean that the patient doesn't have it; it might not mean that it's not testable in the rest of the family; it really depends on what you've tested and what the validity is of that test. And, so it's very important to understand that before you

offer genetic testing to your patients. And this is a very nice site for information about that if you need more information.

Now, I'm going to go back to OMIM, and I'm going to navigate within OMIM and, from OMIM, to other sites, so that you can see that, again, I'm biased, I run it, but it's a unique site in terms of the richness of its links and its ability to get you out to other relevant data quickly and easily. So, again, we're back at OMIM's home page and, in this case, I have, well, I'm going to take you from OMIM's home page to the Genetic Alliance, which is just on the blue bar, and I've circled that.

The Genetic Alliance is an umbrella organization of 600 different support groups, and they add more on a regular basis, and it's a very, very good site to find information about rare disorders, particularly written in lay language that your patients can understand. So, it's a very useful site to know about, and you can search by the name of the disease or by specific treatments or features of the disorder. Okay.

Now, I'm going to go back to within OMIM, and I'm at the home page, and I've typed in long QT because I've decided I just need to know more about long QT. I get back 28 answers. And, one of the other caveats we know about OMIM versus Gene Test is, while Gene Test have these very nice gene reviews, which are overviews, OMIM tends to split. So, if there are six different long QT syndromes caused by mutations of six different genes, obviously a long QT is a long QT is a long QT is a long QT. So, clinically, they're indistinguishable. However, because they are caused by mutations in different genes, it may mean that eventually, not at this moment, but eventually, the treatment will be different. And, so, as an editorial decision within OMIM, we decide to split, so that we always have the ability to, it's easier to start out split than to split later after you've already lumped. So we tend not to lump. So that's why you'll find all this. Now, the up side is that it's very easy to find out the details of a specific question you have. However, if you're a novice coming in, and you're a third year medical student and someone has a long QT, and suddenly you come back with 28 answers, it's a little scary. So, if there is a gene review, I would recommend you look at it but, unfortunately, for instance, there is no gene review on long QT, yet. There probably will be soon.

So, now, I've gone within long QT, and I clicked on long QT 1. Though, usually, if there's a sequence number, so there's long QT 1 through 6, the number 1 entry will usually have the most overview type information because it's the first one that was created; it's where we tend to talk about the others relative to that one. In this case, I want to find out more about the gene, and the gene that's mutated in long QT 1 is KCM Q1, and so I can click on that link, okay, and now I want to show you some features from within a MIM entry. So, if I click on that hot link, it will take me to the reference section, which gives me the full citation for that reference, and if I click on the PubMed ID, I go directly to the abstract. If I click... now I've gone back to this entry, and if I click on that light bulb at the end of a paragraph, and there's a light bulb at the end of every paragraph, that does a proximity search for all the key words. It does a PubMed search on all the key words in the foregoing paragraph. It's called a

neighboring feature. So, that's what I get back when I click on that. And that can be very useful if you don't....you know, if you put long QT into PubMed, you'd get hundreds of thousands of entries, and it would be overwhelming and almost impossible to find. Okay.

Now, I'm going to go along the blue bar from within a MIM entry. Okay. So I'm back in KCM Q1, and now I'm going to circle, I'm going to click on Locus Link. Okay. So, Locus Link is an important site to know about because, if you can't find something in OMIM, please look in Locus Link. Locus Link is NCBI's place to keep all genes. And they keep them even if there's very, very little information known about the gene, even if it's just a sequence that looks like it's likely to be a gene, and no one knows anything functional about it. That will exist in Locus Link. So, if you can't find something in OMIM, look for it in Locus Link. OMIM attempts to get every disorder and every gene of known function and the relationship between genes and disorders, but we are not attempting to get genes of unknown function yet, because we can't do everything. Okay.

Now, this is the Locus Link return to that search, and if I click on that number, the highlighted number, I will get to the full Locus Link page and, as you can see from within Locus Link, all those top buttons and colors are other links you can go to from within Locus Link. And I won't go into the details of what all of those are other than to tell you that you can get to many of them directly from an OMIM page, but I do want to highlight the purple variance one. That takes you to this page, which is the DB SNP, or the single nucleotide polymorphism data base which lists every SNP within that region for that gene. So, if you're trying to set up an association study and to figure out which are your best SNPs to use, this is the site to go to first. It's a very, very rich site on its own.

Okay, back to the entry I'm in. Below Locus Link, if you click on the nomenclature button, that will take you to the official gene symbol for this gene, which is what you should use if you're writing about it. If you're not writing about it, you can call it any favorite name you have. But if you're writing about it, you should use the official gene symbol. Okay.

The next line down takes you to the reference sequence that's curated by NCBI and is the agreed-upon reference sequence for that gene. We'll return this page and, if you click on that link, that will take you to the actual reference sequence and the associated annotations with that reference sequence. If we go back to this page, I won't go through the rest of the links, but the Gen Bank takes you to every Gen Bank sequence about that gene that's been entered. The protein is the protein sequence, and unique gene tells you about related genes in different species. That if there are sites within other data bases, we will actually link directly to the relevant page from within the OMIM page. So, HGMD stands for the Human Gene Mutation Database. It is located at Cardiff and it's funded, in part, by Celera. So, by definition, there's a nine-month delay. It attempts to catalog every published gene, not all variances — those are maintained in DB SNP and other sites, which I'll talk about in a minute — but it attempts to catalog every gene, and it is, again, with a nine-month delay. And you can go directly to the relevant page within HGMD from within an OMIM entry. Of course, you can search with an HGMD itself as well. If you want to see what mutations we catalog

within OMIM, you can click on the view list of allelic variance, which will take you to the allelic variants that we list, and you can click on the specific one and find out how we described it, where it's published, etc., and, of course, going to the relevant citation.

In OMIM, we do not attempt to catalog all allelic variants. We attempt to get the first five in a gene, anything that's associated with unique phenotypes. So, for instance, if there are two different phenotypes associated with mutations in one gene, in this example, it would be Long QT 1, we would, of course, make sure we had both of those among our list of allelic variance. Any mutation that's frequent in a population, or anything that has an interesting mechanism of mutation, but we're certainly not trying to catalog all allelic variants.

Now, I'm going to give you an example of CFTR and, in addition to HGMD which I just talked to you about, and that takes you to the HGMD link, there are other links. The first one is the CFMDB, which is the Cystic Fibrosis Mutation Data Base. If there is a locus specific data base that's recognized by the community, and it's told us that what we found exists, OMIM will be linked directly to it. So this takes you to the CF Mutation Database which catalogs every single mutation published or unpublished within CFTR. And that's just an example of what they look like.

Finally, if there are cell lines that are maintained at the Coriell cell repository that have mutations in a gene so that you could have cells to study, that will link it within the specific entry. So, that's what CCR takes....it takes you to the Coriell repository and allows you to pick up whatever cell line you might be interested in. Okay.

The next two sets I'm going to talk about are really educational resources, not for clinical use or for data mining, but to learn more about genetics. NCHPEG stands for the National Coalition for Health Professional Education and Genetics, and it's URL is indicated there. They have over 100 member organizations that contribute to their goal, and they are again funded by NIH, HRSA and others. This is their site, and what's important about this site is in the middle, where you can find a clearing house of genetic education programs, so if you'd like to be doing genetics education as part of your work or within your school, you can see what's available there. And then the genetic search engine, GROW, Genetic Resources on the Web, is a group of organizations that provide genetic resources on the web, and it will allow you, the advantage of that over, say, Google Search, is that you know you're searching within sites that are appropriate member sites. They will give valid and accurate information. Okay.

Finally, if you'd like, if you've heard this and decided you don't know enough genetics and would like to take a course for yourself, again I have nothing to do with this site either, so I'm endorsing it but I don't benefit in any way from my endorsement. Genetics, in your Practice, is sponsored by the March of Dimes. It's a free online course in Principles of Genetics that is case-based instruction, and the cases you can go through can be tailored to what your own clinical interests are, so it could be a prenatal case, a pediatric case, or an adult case, as you wish, and the URL is listed at the bottom. And that's what the site looks like.

You have to register to use it, but it's a free site, and then you can take whichever course you're interested in.

And thank you very much for your attention.