Brendan Burke
Bioinformatics
6/4/2001

Applying Bioinfomatics Strategies in Huntingtons Disease Research

Introduction:

Traditionally genetics research has been based on correlative phenotypic evidence, the correlation of genotypic variables to phenotypic outcomes to formulate and test hypotheses about gene function. These traditional methods include penetrance correlations, concordance correlations and other applications of Mendelian genetics. However the new information offered by genome research provides a more efficient, genotypic approach to genetics research through the application of bioinformatics. While these new techniques provide the possibility of testing hypotheses prior to experimentation, the real value of bioinformatics is to help in the development of refined hypotheses will eventually be tested experimentally. In this paper I apply bioinformatics analysis to current hypotheses on the development of Huntingtons Disease in order to develop a refined hypothesis for experimentation.

Background:

Huntingtons Disease (HD) is the classic example of an autosomal dominant disorder in genetics research. Named after George Huntington, the Long Island doctor who originally identified the disorder, Huntingtons Disease has since been classified as a member of a distinct class of genetic disorders caused by a tandem repeat mutation of the CAG trinucleotide sequence. The mutant HD gene encodes an expansion of the polyglutamine tract in the HD protein, known as huntingtin. Mutant huntingtin includes a polyglutamine tract that is 36 -120 repeats long, as compared to 11 - 36 repeats in the

wild type protein. The mutant huntingtin protein is believed to be the cause of the neural degeneration associated with Huntingtons Disease.

Huntingtons Disease was the first disorder to be connected and mapped to a specific gene. In 1983 The Huntingtons Disease Collaborative Group, led by James Gusella, mapped the HD gene to within 2.5 million base pairs on the short arm of chromosome 4. In 1993 the research group localized the mutant HD gene to its specific locus. In doing so they made the discovery that the mutant HD gene contained a long repeat of the CAG trinucleotide. Attempting to establish a definite relationship between the HD gene and neural degeneration, the group bred mice in some of which the gene had been inactivated. However this resulted in embryonic death. Though unsuccessful in their attempt to prove a relationship, the group made significant discoveries establishing that HD gene was essential to organism survival. They also established that the HD gene played some role in early embryonic development, because in all cases of the gene's inactivation resulted in embryonic death.

Dr. Gusella's research precipitated many new discoveries about the HD gene and Huntingtons Disease. Researchers established a standard for gene stability, with a normal range of CAG repeats at 15 to 36, and an unstable expansion of CAG repeats at 36 to 120. Researchers additionally discovered that the length of the triplet repeat expansion is highly correlated with age of onset. These findings allowed doctors to diagnose Huntingtons patients before onset of observable symptoms, and also allow them to estimate at what age symptoms would begin to appear. Researchers also found a correlation between the length of the trinucleotide expansion and the rate of degeneration. It was found that individuals heterozygous for the gene showed faster deterioration from

the age of onset. Though much progress had been made, three major questions continued to baffle researchers: What was the function of the wild type huntingtin protein? How does mutant huntingtin cause neural degeneration? How does the HD gene, which is expressed all over the body, cause very targeted neural degeneration?

In November of 1995 another significant breakthrough was made in Huntingtons Disease research which offered some possible answers. A research team at Johns Hopkins University discovered the protein dubbed huntingtin associated protein (HAP1), which binds readily to the huntingtin protein. Several aspects of this protein correlated with previous Huntingtons Disease research and provided researchers with new hypotheses. First, was the localization the protein's expression. The fact that HAP-1 was found only in the brain offered an explanation as to how huntingtin, a protein present throughout the body, could only inflict damage on the brain. Furthermore HAP1 was found in especially high concentration in areas of the brain where Huntingtons Disease is known to inflict its greatest damage. Second, researchers also found that the tighter binding of HAP-1 to huntingtin could be correlated with an increased number of trinucleotide repeats. This finding is possibly connected with age of onset. The tight binding of huntingtin to HAP1 may cause early onset of Huntingtons Disease. Researchers attempted to establish a definite correlation between binding of HAP-1 to huntingtin and neurodegeneration. Just as previous experimentation established that the HD gene is essential for organism development, the inactivation of HAP-1 in mice again produced fatality during early stages of embryonic development.

Researchers have since isolated additional proteins proven to interact with huntingtin, dubbed HIP1, HIP2, etc. Of greatest interest was huntingtin interacting

protein 1 (HIP1.) HIP1, like the HAP1 protein, is isolated to the brain. However unlike HAP1, the strength of HIP1 binding with huntingtin seems to be a reverse correlation, with HIP1 interactions with huntingtin decreasing as polyglutamine tract length increases. Experimental deletion of the HIP1 in mice also resulted in embryonic death. Though no specific relationship can be drawn between these protein interactions and the neurodegeneration, it is significant that all of these proteins are vital to development, and have interactions affected by the polyglutamine repeat mutation. Further investigation of the role of the polyglutamine expansion in huntingtin is likely give us insight into the function of the protein itself and ultimately a means to prevent the effects of neurodegeneration in Huntingtons Disease.

The normal function of the polyglutamine repeat in wild-type huntingtin is still unknown. However clues to the significance of polyglutamine expansions can be found in their involvement in other genetic disorders. The class of diseases associated with polymorphic trinucleotide repeat mutations share distinct features atypical of other genetic disorders. The diseases are always autosomal dominant neurodegenerative disorders, with severity and onset of symptoms correlating directly with an increased number of repeats. Triplet repeat mutations occur in polymorphic C-G trinucleotides, which expand to become unstable. These repeats tend to increase in length over successive generations, causing increased severity in family lines carrying the disease. In all cases neurodegeneration is thought to be related to protein products, because the mutant gene is expressed in a mutant protein form. The fact that damage is isolated to the brain even, though the mutant gene is expressed in non-neural cells, is thought to indicate that the mutations affect interactions with proteins isolated to the brain.

The trinucleotide repeat class of diseases all have affects isolated to the brain, however each disease does have a distinct set of symptoms. Huntingtons Disease causes very distinct patterns of neural deterioration producing a three-fold set of symptoms disrupting motor, cognitive and emotional functions. Onset can occur from any age between childhood and old age, but typically symptoms typically appear between the ages of 35 and 50 years. Death occurs on average 15 – 20 years following onset of symptoms. The distinct similarities in this class of genetic diseases give strong indication that the expansion of the CAG trinucleotides is the cause of the symptoms of neurological degeneration. The distinction in the specific symptoms among these diseases, despite the similarity of the mutations, indicates that the protein probably affects protein interactions in specific areas of the brain.

The Old Hypothesis:

The exact mechanism of neurodegeneration in polyglutamine mutant diseases is still unknown. However by analyzing the correlations among the characteristics of the HD gene, the huntingtin protein, the various huntingtin interacting proteins and trinucleotide repeat mutations researchers have formulated hypotheses as to likely mechanisms. Possible mechanisms can be divided into two categories. The first category is gain of function mechanisms, in which the mutant form of the protein gains some function that is toxic to neurons. This would correlate with the HAP1-huntingtin interaction in which HAP1 binds more tightly to the mutant huntingtin. The second category is loss of function mechanisms, in which the mutation of huntingtin protein causes it to lose its original function and thus disrupts some vital cell process. This

would correlate with a case such as the HIP1-huntingtin interaction, which is decreased with mutant huntingtin.

Up until this point most researchers have supported hypotheses involving gain of function, as such hypotheses are supported by numerous characteristics of the disease. Most important is the dominant inheritance pattern of the HD disorder, meaning that the a single mutant copy can cause the disease, even with functionality of the other wild type HD gene. Additionally individuals who have two copies of the mutant HD gene and those with only one do not display significantly different severity of symptoms, whereas those missing one copy of the wild type HD gene do not develop symptoms at all. A common gain of function hypothesis is that expansion of the polyglutamine tract in a regulatory protein affects the interactions with other proteins, and thus disrupts regulatory pathway. Neural degeneration could be the result of incorrect neural messages being sent, and specific neural breakdown cause by the interaction of the huntingtin protein with various coregulatory proteins.

A New Hypothesis:

It is not as clear as it might seem that the cause of HD neural degeneration is a mutant gain of function, or at least not limited to gain of function. New research has called into question the likelihood of a gain of function mechanism in HD. Researchers have recently found that though the deletion of the HD gene results in embryonic death, mice in which the gene was inactivated after birth showed patterns of neurodegeneration similar to those of the huntingtin mutation itself. These findings contradict the assertion that dominant inheritance patterns in Huntingtons Disease indicate a gain of function

hypothesis, because the same effects neurodegeneration results when there is a mutant gene or no gene at all.

It is clear that in order to further develop hypotheses on the mechanism of neurodegeneration in HD it is necessary to gain additional insight into the nature of protein interactions among the huntingtin associated interacting proteins. We have a hypothesis that explains neurodegeneration as the result of tighter binding of mutant huntingtin to HAP1, yet we have evidence that indicates that there is a loss of gain of function because inactivation of the HD gene at later stages causes a similar approach. Thus we need to hypothesize a mechanism for autosomal dominant loss of function mutation.

This is where we can apply bioinfomatics analysis. Applying what we already know about the mapped regions and sequences of the human genome we can analyze the three proteins by comparing their homology to other proteins or sequences and using a statistical algorithm calculate the likelihood that similar structure indicates similar function. Highly homologous sequences are likely to have similar functions, giving us new insight into the function of these proteins. I analyzed huntingtin, HAP1 and HIP1 using eight different motif databases, however only returned significant results on one. The ExPASy Prosite database returned significant and correlating results for all three. It isolated a common structure called a leucine zipper present in each of the protein sequences.

Leucine zippers thought to be involved in involved in mediating interactions among proteins containing similar leucine structures. The structure contains periodic arrays of leucine residues existing in an alpha helical conformation. The structure is

found predominantly in regulatory proteins. This is similar to hypotheses on the function of the HAP1 protein, which is thought to be associated with gene regulation. This structure would give these proteins a common functional link and also a possible interacting link since leucine zippers can facilitate protein-protein interactions. These three proteins may be involved in an elaborate gene regulatory mechanism, associated with each other through by the leucine zippers and binding together in there regulatory processes by polyglutamine zipper. An increase in the expansion of the polyglutamine tract may cause active inhibition of huntingtin-HIP1 binding by the tighter huntingtin-HAP1 binding, in a regulatory feedback mechanism. *The huntingtin regulatory mechanism would be regulated for both loss of function in the case of the huntingtin-HIP1 interaction, and for gain of function in the case of the huntingtin-HAP1 interaction.* To build on the previous gain of function hypothesis, this hypothesis would account for both increased protein interaction and decreased interaction as a possible explanation for the dominant loss of function mechanism.

Additionally the leucine zipper provides another simpler hypothesis for a dominant loss of function mutations in HD. Leucine zippers are polar tracts and can initiate tight non-covalent binding between proteins. Additionally glutamine is a polar amino acid and thus the polyglutamine tracts can also be involved in protein binding. An expansion of the polyglutamine tract could initiate a tight protein-protein binding. Such an interaction would likely be nonproductive ones and interfere or inhibit the function involved proteins. *The leucine zipper and polyglutamine tracts may initiate a binding interact that would inhibit the function of the mutant huntingtin protein, and additionally bind to wild type huntingtin to form protein aggregates thus inhibiting wild type function*

as well. This interaction would explain the autosomal dominant inheritance pattern of this loss of function mutation as the mutation, inhibits functioning of both mutant and wild type proteins.

Future Experimentation:

Having developed a new refined hypothesis, we must now test it experimentally. The hypothesis applies the theories about polyglutamine binding mechanisms to the possible regulatory function of huntingtin, HAP1 and HIP1. In order to test this hypothesis we must introduce into an HD mutant gene carrier, a HD gene that has been cleaved to eliminate the CAG trinucleotide repeat and thus eliminate the polyglutamine tract from the wild-type protein. This would eliminate the protein binding interactions eliminate at least half of the proteins and thus under our hypothesis would result in slower or inhibit the neurodegeneration. This experiment would answer three important questions about the development of Huntingtons Disease. First, that there is a loss of function aspect to the disease mechanism. Second, that the huntingtin protein interactions are involved in a regulatory mechanisms. Third that the binding of HAP1 and HIP1 to mutant huntingtin are responsible for the genetic deterioration. Additionally, and this experiment would confirm the accuracy of the bioinformatics techniques used to develop this new refined hypothesis.

Conclusions:

Despite the fact that Huntingtons Disease was the first disorder correlated and mapped to a specific gene in 1993, Researchers still have not been able to determine the specific mechanism by which the HD mutation causes neurodegeneration. The likely reason is that the mechanism that causes degeneration is a complex, regulatory feedback

loop, with multiple protein interactions. This would explain the varied results from experimentation with the HD mutation, and explain how the mutation could be simultaneously autosomal dominant and involve a loss mechanism. Bioinformatics are vital here because they provide us with new means to develop and test hypotheses about gene and protein function. The use of bioinformatics helps researchers to develop refined hypotheses by bypassing certain aspects of lab experimentation. Additionally, as exemplified in this investigation of Huntingtons Disease, bioinformatics allows us to formulate hypotheses as to the function of proteins in complex mechanisms and formulate new hypotheses to explain experimental evidence. Bioinformatics is an valuable tool in the study of genetics, by giving us another method for analysis of genetic information, thus helping to improve the efficiency and accuracy of experimental results.

Ethical Questions:

Huntingtons Disease was the first disease mapped to a specific locus in the genome, and thus far the closest genetic correlation we have for any disease. Based on these factors we would expect that the Huntingtons Disease would be the first to be conquered by genetically developed treatments. However thus far this has not been the case, and the results that Huntingtons Disease research has produced over the last eight years since the protein was isolated have been mainly confirming or disproving hypotheses as to the proteins function. This information is very valuable in the process of an ultimate cure for the disorder, and bioinformatics will help bring greater efficiency to this process, however in following this process there may be applications of the current information that we overlook.

When researching terminal disease there is an obligation to produce results that will aid in treating the victims, and thus we have an obligation to pursue every possible venue for treatment. In particular in the case of Huntingtons Disease the only results research has produced to help potential victims of this hereditary disease are genetic tests to verify the carriers of the disease. However research has offered no options for treatment of the disorder and thus those who are tested gain nothing but the knowledge that they suffer from a terminal genetic disease. Furthermore the dominance and relative rare incidence of the Huntingtons Disease make creates an ethical dilemma when developing a public policy approach to treating this disorder. There post-testing options available to carriers of the disease, but genetic testing offers a pre-natal solution of selective abortion, to eliminate all carriers of the disease. Even if this were not a public policy, there are still public policy applications when it comes to regulating abortion, as some parents would opt to use screening and abortion to guarantee a healthy child. If there were some form of post-testing treatment options would weigh against such an option. Thus there is additional provides pressure in the case of Huntingtons Disease to develop treatments based on applications of current research.

In research on genetic disease there are two types of approach to any type of genetic analysis, one that seeks to understand then solve, and one that seeks out only that information that will produce a solution. Typical research approaches generally follow the first approach illustrated by the progression of HD research up to this point. However I think that this type of analysis, while the most prudent approach is certainly not the most efficient, and when conducting research on terminal illness the first priority must be helping the afflicted, and then gaining knowledge. Operating in this mindset there is one

relationship of vital importance that has not yet been pursued. All real research up to this point has focused on the mutation of the Huntington gene, its protein product and their specific function. Few researchers have ever sought to research the wild type Huntington gene and why in this form it does not cause brain deterioration.

The most recent discoveries in Huntingtons Disease research were such a breakthrough, that researchers never really stepped back and tried to interpret the data but instead continued to beat along the same path. Researchers found that a mutation causing an increase in CAG trinucleotides in the Huntington gene. Next they found a protein produced by this mutant Huntington gene called, huntingtin that bound to HAP-1 in the brain to cause neurological deterioration. Further increase in the number of CAGs in the HD mutation causes tighter binding of these two proteins and consequently cause earlier onset of Huntingtons disease and usually earlier fatality. Thus a pattern can be seen forming between damage to the brain and an increase in the number of CAG trinucleotides. As the number of CAG trinucleotides increases a huntingtin type with a greater amount of glutamine (the amino acid produced by the CAG trinucleotides) is produced which binds tighter to HAP-1 and causes greater damage to the brain.

In light of this apparent pattern, I offer the hypothesis that the smaller number of CAGs in the normal Huntington gene causes it to produce a protein with less glutamines, and thus binds to HAP-1 more loosely and causes no brain damage. If this protein could be produced and successfully injected into a HD patient, it could bind to HAP-1 and act as an inhibitory agent preventing the destructive interaction between the mutant huntingtin and HAP-1 thus preventing onset of the disease and slowing brain deterioration.

Though this option is a long shot based on current information I would encourage this type of approach to Huntingtons Disease research in light of the ethical questions we currently face in approaching treatment of the disorder.

References:

Pennachio, Len A. and Rubin, Edward M.; “Genomic Strategies to Identify Mammalian Regulatory Sequences;” *Nature*; February 2001, pp. 100 – 109

Emilien, G. “Impact of Genomics on Drug Discovery and Clinical Medicine;” *QJM*; July 2000 pp. 391 – 423

Ross, Chrostopher A.; Huntington Disease and the Related Disorder, Dentatorubral-Pallidoluysian Atrophy; September 1997 pp. 305 – 338

Kehoe, Patrick; “Age of onset in Huntingtons Disease: sex specific influence of apolipoprotein E genotype and normal CAG repeat length;” Journal of Medical Genetics; February 1999 pp.108 – 111

Li Xiao-Jiang; “A huntingtin-associated protein enriched in the brain with implications for pathology;” Nature; November 23, 1995 pp. 398 – 402

Li, Shi-Hua; “Association of HAP1 Isoforms with a Unique Cytoplasmic Structure;” Journal of NeuroChemistry; November 1998 pp. 2178 – 2185

Harding, AE; "The Gene for Huntingtons Disease;" British Medical Journal; August 4, 1994; p. 396

"Molecular Analysis and Clinical Correlation of HD Mutation;" Macmillan, JC and Associates; The Lancet; October 16, 1993

Gusella, James F; "Huntingtons Disease and Repeating Trinucleotides;" The New England Journal of Medicine; May 19, 1994

Blakeslee, Sandra; "Newfound Brain Protein May be 'Smoking Gun' in Huntingtons:" The New York Times; November 14, 1995; p. C3

Jackson, George R.; “Polyglutamine-Expanded Human Huntingtin Transgenes Induce Degeneration of Drosophila Photoreceptor Neurons;” Neuron, Sept 1998 pp. 633-642

Li Zhen; A putative Drosophila homologue of the Huntington’s Disease gene