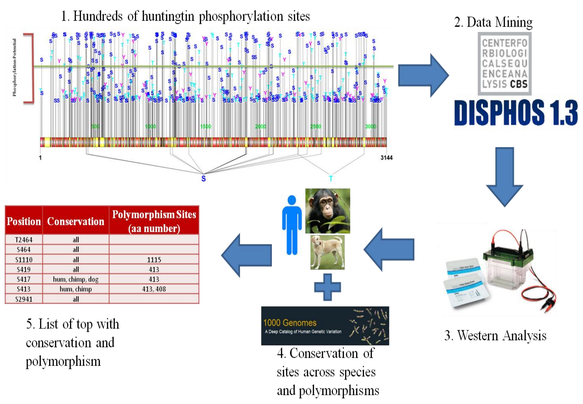
INTRODUCTION

Huntington's Disease (HD) is a neurodegenerative disease caused by a mutation in the huntingtin gene (HTT) [1-2]. Symptoms of HD include motor dysfunction and cognitive decline such as chorea or loss of motor ability, memory loss, and other psychological abnormalities [2]. The mutation, an abnormal CAG expansion (polyQ), encodes an extended polyglutamine track in the huntingtin protein (Htt), which results in HD [1, 3-4]. Although HD's root cause has been identified as a mutation in a single gene coding for a mutated version of the protein for over 20 years, HD remains incurable and fatal.

Phosphorylation is a type of posttranslational modification (PTM) that is critical in regulating a protein's function by changing the shape of the targeted protein [5-6]. Studies have shown that certain PTMs such as phosphorylation have significant positive influence on clinical phenotypes of HD by modifying mutant Htt [7-11]. Therefore, phosphorylation may be an effective clinical treatment aimed at the root of the disease because it has the potential to modify the mutant protein's structure so that the mutant protein will function more like a normal huntingtin protein. The first critical step is identifying a polyQ length dependent phosphorylation site, a site functionally important in HD pathogenesis. After identifying a site, conventional drugs such as kinase inhibitors may regulate phosphorylation patterns for that site in order to produce clinically desirable results. Such treatments for other diseases already exist as proofs of concept. A drug called Gleeves or imatinib targets a specific kinase (a phosphorylating molecule) that is one of factors causing leukemia [12]. Studies have shown that this drug has effectively demonstrated lasting positive effects as a treatment for leukemia simply by regulating a phosphorylation site [12].

Despite the potential benefits of finding a PTM site important in HD pathogenesis, it is currently a time consuming and costly procedure involving the use of mass spectrometry. Although there have been a few phosphorylation sites identified as being critical to HD, these sites were found through random selection of phosphorylation sites, not from a systematic approach [7, 13-14]. In our study, we first utilized online prediction programs to generate a list of all the potential phosphorylation sites in the huntingtin protein. We then narrowed down our list with a logical procedure of using conservation across species and the presence of polymorphisms within phosphorylation sites to develop a list of sites clinically worth testing. Finally, we verified the accuracy of our procedure by experimentally verifying the polyQ dependency of certain sites from our list. Our procedure of using prediction programs, conservation across species, and polymorphisms is an efficient method of selecting clinically beneficial phosphorylation sites for any protein-related disease without the need of expensive pre-screening tests such as mass spectrometry.

PROCEDURE



Position	Conservation	Polymorphism Sites (aa number)
T2464	all	
S464	all	
S1110	all	1115
S419	all	413
S417	hum, chimp, dog	413
S413	hum, chimp	413, 408
S2941	all	

Figure 2: Procedure Schematic
General overview of procedure sequence.

Using Prediction Programs to Predict Phosphorylation Sites
Obtain Homo Sapiens huntingtin protein sequence from NCBI and input sequence into prediction programs NetPhos 2.0 (Center for Biological Sequence) and DISPHOS 1.3 (Molecular Kinetics) to generate predicted phosphorylation sites.

Selecting 30 Candidate Sites
Sum the scores of each phosphorylation site from each of the two programs. Rank according to the summed scores and take the top thirty sites that have the highest summed predicted phosphorylation score.

Conservation
Obtain human, chimpanzee, and dog huntingtin protein sequences f¬rom NCBI and use Cluster (European Bioinformatics Institute) program to align the sequences. For each of the 30 phosphorylation sites, record the species that they are conserved across.

Polymorphism
Using data from the 1000 Genome Project, record the single nucleotide polymorphisms (SNP) in the huntingtin gene. For each polymorphism in the coding region, use a genetic code conversion table to record whether each polymorphism actually changes the amino acid in translation.

Western Blot Analysis
Perform western blot analysis on 6 phosphorylation sites from the 30 candidates to experimentally validate whether they are indeed actual phosphorylation sites and if so, whether they are polyQ dependent. Run normal and huntingtin knockdown lymphoblast lysates on 10% SDS PAGE (polyacrylamide gel electrophoresis) gel and transfer to nitrocellulose membrane. Probe membranes with phospho-specific antibodies for the 6 phosphorylation sites and perform the rest of the western blot analysis according to the protocol from Bio-Rad Inc. Probe a replica membrane with the pan-huntingtin antibody to confirm the specificity of the phospho-specific antibodies. To test the polyQ dependency, perform western blot analysis on two HD patient lymphoblast lysates (CAG 82/21, CAG 82/15) with same phospho-specific antibodies. Quantify the mutant and wild type huntingtin signal of the two HD patient lymphoblast lysates using ImageJ 1.47 software.

Programs Used
1. DISPHOS: http://www.dabi.temple.edu/disphos/
2. NetPhos ver. 2.0: http://www.cbs.dtu.dk/services/NetPhos/
3. ImageJ ver. 1.47: http://rsbweb.nih.gov/ij/

RESULTS

1 Selecting Candidates and Top 30 Ranking Accuracy Validation

1. Data mine using DISPHOS and NetPhos
2. Figure 3 shows top 30 potential phosphorylation sites ranked by summed score.
3. Figure 4 shows western analysis of six chosen sites, which reveals that the difference in ranking within top 30 candidates is negligible.

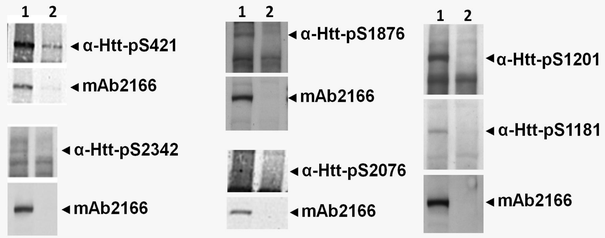


Figure 4: Western Analysis for Phosphorylation Check
The six selected phosphorylation site antibodies were tested to reveal whether they accurately targeted Htt. 1=Normal Cells. 2=Normal-Htt KD Cells with HTT-shRNA.

Rank	Position	Summed Score
1	S459	1.932
2	S457	1.912
3	S461	1.868
4	S421	1.811
5	S1218	1.808
6	S1201	1.766
7	S1215	1.724
8	S2657	1.683
9	S1876	1.67
10	T2464	1.655
11	S464	1.654
12	S1110	1.649
13	S419	1.647
14	S417	1.639
15	S413	1.631

Rank (cont)	Position (cont)	Summed Score (cont)
16	S2941	1.628
17	S471	1.626
18	S2076	1.617
19	S2074	1.603
20	S1181	1.596
21	S2653	1.593
22	S1106	1.567
23	S558	1.556
24	S2080	1.552
25	S2936	1.539
26	S2489	1.532
27	S540	1.531
28	S1113	1.524
29	S561	1.52
30	S2342	1.502

Figure 3: Top 30 Potential Phosphorylation Sites
The top 30 phosphorylation sites ranked by summed score (sum of site's score from the two prediction programs). Red sites indicate sites selected for western analysis.

2 Picking Phosphorylation Sites Worth Testing Through Conservation Across Species and Polymorphisms

1. Huntingtin protein sequences of humans, chimpanzees, and dogs aligned in Cluster program as shown in Figure 5.
2. Figure 6 reveals polymorphisms found by data mining 1000 Genome Project database. Presence of polymorphisms may indicate functional importance due to Figure 7.
3. Figure 8 reveals top 30, unranked, with conservation patterns and polymorphism sites within +/- 7 amino acids of the phosphorylation site as a kinase motif [15].
4. Figure 8 reveals one of our tested sites, S2076, has the most potential to be polyQ length dependent since it is conserved across all species and has two polymorphism sites.

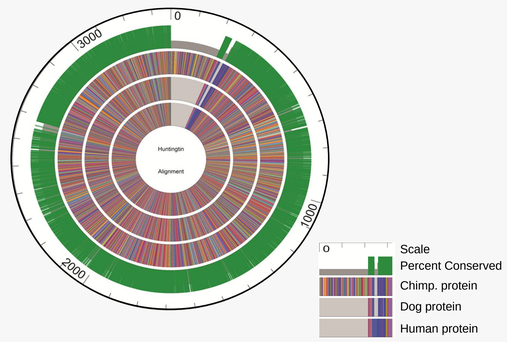


Figure 5: Huntingtin Alignment for Three Species to Test Conservation of Phosphorylation Sites Across Species

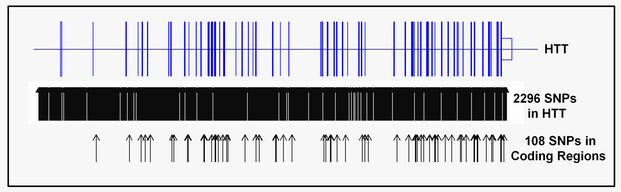


Figure 6: Polymorphisms Analyzed Using 1000 Genome Project Data
Data mining 1000 Genome Project yielded 2296 polymorphisms total in huntingin gene, but only 108 polymorphisms in coding regions of the gene. Black arrows mark locations of each polymorphism. Vertical blue lines indicate 67 exons of HTT.

Position	Conservation Across Species	Polymorphism Sites (aa number)
S2076	all	2074, 2081
S419	all	413
S417	all	413
S540	all	539
S558	all	551
S1110	all	1115
S1113	all	1115
S2342	hum, chimp	2337
S2074	hum, chimp	2074, 2081
S2080	hum, chimp	2074, 2081
S413	hum, chimp	413, 408
S459	all	
S457	all	
S461	all	
S421	all	

Position	Conservation Across Species	Polymorphism Sites (aa number)
S1201	all	
S1215	all	
S2657	all	
T2464	all	
S464	all	
S2941	all	
S471	all	
S1181	all	
S2653	all	
S1106	all	
S2936	all	
S2489	all	
S561	all	
S1218	hum, chimp	
S1876	hum, chimp	

Figure 8: 30 Phosphorylation Sites with Species Conservation and Polymorphism
Top 30 phosphorylation sites from Figure 3 without rank, but with conservation across 3 species (Human, Chimpanzee, Dog) and polymorphism patterns.

3 Phosphorylation Sites S2076, S1181, and S2342 Are PolyQ Length Dependent

1. Figure 9a reveals successful western analysis with clearly visible bands in proper locations.
2. Figure 9b reveals phosphorylation sites S2076, S2342, and S1181 demonstrated polyQ length dependency in phosphorylation patterns (different between mutant and normal huntingtin protein).
3. Comparing Figure 9b with Figure 8 reveals our procedure is effective in finding sites worth testing.

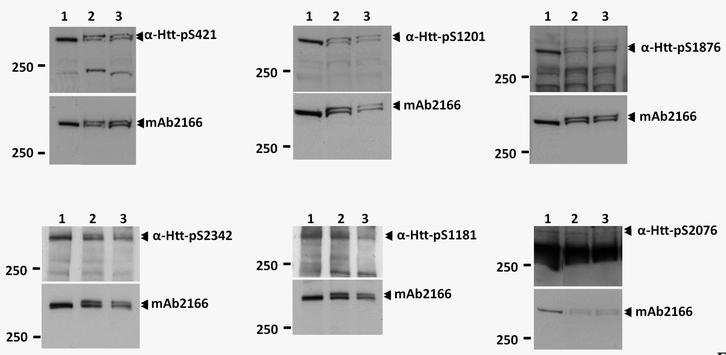


Figure 9a: Western Analysis of 6 Phosphorylation Sites for PolyQ Length Dependency
Empirical analysis of six phospho-specific antibodies with Htt general antibody (mAb2166) as control. 1=Normal-HttQ20/17. 2=HD Patient-HttQ82/21. 3=HD Patient-HttQ82/15.

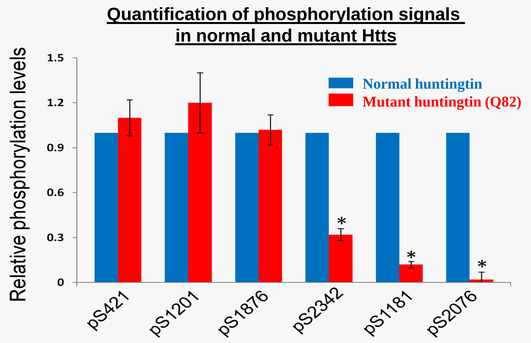


Figure 9b: Graph of PolyQ Length Dependency for 6 phosphorylation sites
*Each phosphorylation signal was normalized by mAb2166 signal. Student T-test conducted to relatively compare mutant phosphorylation levels to normal. P value: *<.001 (n =2)*

BACKGROUND

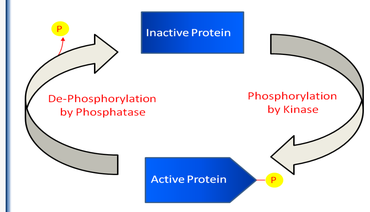


Figure 1: Phosphorylation Background
Phosphorylation attaches negatively charged phosphate molecule which alters protein shape and consequently its functions.

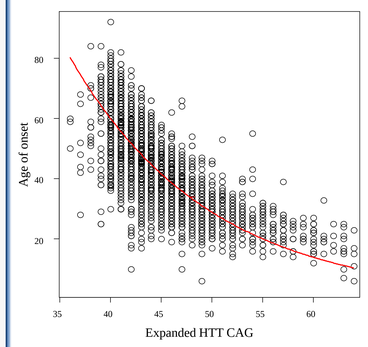


Figure 7: Variations in Age of Onset According to PolyQ Length
General inverse relation between age of onset and polyQ length. But individual points (patients) reveal considerable variations in age of onset for a given polyQ length suggesting involvement of factors other than poly-Q length HD pathogenesis [16].

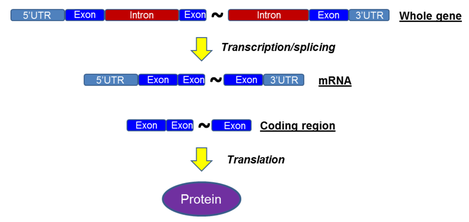


Figure 11: Protein Synthesis-From Gene to Protein

SUMMARY

1. First Western Analysis not only demonstrates negligibility of rankings within top 30, but also verifies prediction programs' ability to predict actual phosphorylation sites.
2. Phosphorylation sites S2076, S1181, and S2342 are poly-Q length dependent; therefore they are potential candidate target sites for clinical studies.
3. Comparison of poly-Q dependency and conservation/polymorphism result reveals all five sites (excluding S1181) follows expected pattern, supporting accuracy and usefulness of our procedure for finding additional sites in the future.
4. S2076 is our best candidate for future clinical studies according to our analysis method and poly-Q length dependency verification.

REFERENCES

1. The Huntington's Disease Collaborative Research Group, A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, 1993. 72: p. 971-83.
2. Vonsattel, J.P. and M. DiFiglia, Huntington disease. *J Neuropathol Exp Neurol*, 1998. 57: p. 369-84.
3. Gusella, J.F., et al., A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, 1983. 306: p. 234-8.
4. Jou, Y.S. and R.M. Myers, Evidence from antibody studies that the CAG repeat in the Huntington disease gene is expressed in the protein. *Hum Mol Genet*, 1995. 4: p. 465-9.
5. Burnett, G. and E.P. Kennedy, The enzymatic phosphorylation of proteins. *J Biol Chem*, 1954. 211: p. 969-80.
6. Chang, C. and R.C. Stewart, The two-component system. Regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiol*, 1998. 117: p. 723-31.
7. Ehrnhoefer, D.E., L. Sutton, and M.R. Hayden, Small changes, big impact: posttranslational modifications and function of huntingtin in Huntington disease. *Neuroscientist*, 2011. 17: p. 475-92.
8. Humbert, S., et al., The IGF-1/Akt pathway is neuroprotective in Huntington's disease and involves Huntingtin phosphorylation by Akt. *Dev Cell*, 2002. 2: p. 831-7.
9. Gu, X., et al., Serines 13 and 16 are critical determinants of full-length human mutant huntingtin induced disease pathogenesis in HD mice. *Neuron*, 2009. 64: p. 828-40.
10. Gines, S., et al., Enhanced Akt signaling is an early pro-survival response that reflects N-methyl-D-aspartate receptor activation in Huntington's disease knock-in striatal cells. *J Biol Chem*, 2003. 278: p. 50514-22.
11. Wilkinson, K.A., Y. Nakamura, and J.M. Henley, Targets and consequences of protein SUMOylation in neurons. *Brain Res Rev*, 2010. 64: p. 195-212.
12. Druker, B.J., et al., Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med*, 2006. 355: p. 2408-17.
13. Dong, G., et al., Mass spectrometric identification of novel posttranslational modification sites in Huntingtin. *Proteomics*, 2012. 12: p. 2060-4.
14. Schilling, B., et al., Huntingtin phosphorylation sites mapped by mass spectrometry. Modulation of cleavage and toxicity. *J Biol Chem*, 2006. 281: p. 23686-97.
15. Kennely, P.J. and Krebs, E.G., Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Bio Chem*., 1991. 266: p. 15555-58.
16. Lee, J., et al., CAG repeat expansion in Huntington disease determine age at onset in a fully dominant fashion. *Neurology*, 2012. 78: p. 690-95.