Can sequence predict function?

ArticleInfo		
ArticleID	:	4938
ArticleDOI	:	10.1186/gb-spotlight-20040419-01
ArticleCitationID	:	spotlight-20040419-01
ArticleSequenceNumber	:	290
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2004–4–19 OnlineDate : 2004–4–19
ArticleCopyright	:	BioMed Central Ltd2004
ArticleGrants	:	
ArticleContext	:	130594411

Cathy Holding

Email: cathyholding@aol.com

A novel bioinformatics approach for classifying proteins according to similarity of function, rather than of sequence, is described in the April 12 Proceedings of the National Academy of Sciences. Albert Y. Lau and Daniel I. Chasman of Variagenics say that that their approach could be used to construct a database that would allow experimental confirmation of genomic sequences with unknown function. But other researchers questioned the practical applications of the work and suggested it was merely an extension of techniques currently used.

Standard methods of prediction of protein function from sequence rely on either an arbitrary standard - such as a cutoff point at a particular percentage sequence identity - or on analysis of annotations assigned by other experimentalists, Chasman told us.

"[Here], the idea is that if you have a bunch of sequences that you know are functionally related, but they have a few amino acids different, the operational definition tells you that you can substitute those alternative amino acids into each of the sequences and you won't alter the function," Chasman said.

The researchers applied Dirichlet mixtures to ordered sets of multiple sequence alignments - in this instance, a group of tumor suppressor genes. They chose BRCA1, BRCA2, and WT1 because the genes are mutations that encode for diseases only found in mammals. "We claim that the mammalian sequences are all functionally related and distinct from the nonmammalian sequences," Chasman said.

In addition, classifying nonsynonymous single nucleotide polymorphisms with clinical relevance in the human population showed that low-frequency alleles were significantly more likely to affect function than high-frequency alleles - corresponding with predictions from population genetics. "We predict overall about 30% of polymorphisms that were described are predicted to have an effect on function, so [in addition] we give an actual estimate of the overall fraction," he said.

Werner Braun, professor in the Department of Human Biological Chemistry and Genetics at University of Texas Medical Branch, said that although he felt that overall it was an interesting novel approach that seemed to be working, he was concerned about the practicality of it: "I think that it's definitely not yet proven that it will work in practice for finding really individual residues for function." Braun, who was not involved in the study, said that it remained to be seen if the method would be useful in practice for suggesting new experiments for testing functional annotations of proteins.

Chasman said that functional classification of proteins was a necessary first step toward experimental confirmation of protein function. "[A recent paper] in PLoS Biology makes a plea for the community to start to make databases where proteins are classified according to functional similarity, and then for a joint effort from the experimentalists to try and go and confirm it."

However, Richard J. Roberts, research director at New England Biolabs, said that Lau and Chasman's method seemed to be just an extension of the fairly typical approaches in which researchers have been using protein sequence motifs to try to make good guesses about function. "It is known that there are some proteins where a single amino acid change actually changes the substrate specificity. I think those kinds of instances would completely fool this particular method that the authors are proposing."

Roberts said that an experimental approach - and not one based solely on modeling - is paramount. "I think the bioinformatics is pretty easy," he said, "It's really trying to get the experimentalists hooked up and on board to the idea that they have a really incredibly valuable role to play in terms of trying to get these genomes annotated."

References

1. *Proceedings of the National Academy of Sciences USA*, [http://www.pnas.org/]

2. Variagenics (now known as Nuvelo), [http://www.nuvelo.com/]

3. Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology,

4. Werner Braun, [http://www.hbcg.utmb.edu/faculty/braun/]

5. Roberts RJ: Identifying protein function - a call for community action *PLoS Biology* 2004, 2:293-294., [http://www.plosbiology.org/plosonline/?request=get-document%26doi=10.1371/journal.pbio.0020042]

6. Richard J. Roberts, [http://www.nobel.se/medicine/laureates/1993/roberts-autobio.html]

7. Identifying property based sequence motifs in protein families and superfamilies: application to DNase-1 related endonucleases