

A machine learning approach to identify functional human phosphosites

December 11 2019



Artist's impression of phosphosite and machine learning. Credit: Spencer Phillips

Researchers at the EMBL's European Bioinformatics Institute (EMBL-EBI) have created the largest reference phosphoproteome to date of almost 120,000 human phosphosites. To identify those most likely to be critical, they used a machine learning approach capable of ranking them according to functional importance.

Proteins are the core molecular [machines](#) of the cell that can be

regulated by [protein](#) modifications, akin to molecular switches. Protein phosphorylation is one such molecular switch, that can alter the structural conformation of a protein, causing it to become activated, deactivated or modifying its function. Despite decades of work the total number of these modifications and which ones are truly critical for life remains a mystery.

This research, published in *Nature Biotechnology*, creates a freely-accessible resource that can be used by researchers to better understand which proteins are phosphorylated and which phosphosites have functional relevance. Access to this data has significant implications to accelerate the progression of research into many different biological processes and diseases.

Machine learning and data sharing

"This new resource would not have been possible if scientists around the world didn't share their [research data](#) and results," says Pedro Beltrao, Group Leader at the EMBL-EBI. "It would take a single machine over 500 consecutive days to run all the mass spectrometry experiments used to create this database. By applying [machine learning](#) to this huge dataset, we created a [scoring system](#) that will hopefully help researchers to determine which lesser-known phosphosites to explore next."

The researchers at EMBL-EBI curated over 100 publicly available phospho-enriched human datasets containing over 6,000 mass-spectrometry experiments from EMBL-EBI's PRoteomics IDentifications (PRIDE) database. This large-scale project has generated the biggest open access reference phosphoproteome database to date.

Functional human phosphosites

To identify the phosphosites most critical to human cells, machine learning was used to integrate diverse annotations for each site such as the degree of conservation. The phosphosite functional score generated in this study has [enormous potential](#) to help other scientists uncover more about their proteins of interest. It can be used to rank known phosphosites to distinguish those which are functionally relevant for molecular processes and disease.

For example, the researchers were able to demonstrate the practicality of their functional score model by identifying two high-scoring phosphosites which play a role in regulating neuronal differentiation.

"The functional score model created from this study can be used to uncover an abundance of new, functional phosphosites that may play crucial roles in disease," says David Ochoa, Project Coordinator at Open Targets. "We already know of several groups who are using the scoring model, so we would like to encourage researchers everywhere to explore the resource and make use of it."

More information: David Ochoa et al, The functional landscape of the human phosphoproteome, *Nature Biotechnology* (2019). [DOI: 10.1038/s41587-019-0344-3](#)

Provided by European Molecular Biology Laboratory

Citation: A machine learning approach to identify functional human phosphosites (2019, December 11) retrieved 17 April 2024 from <https://phys.org/news/2019-12-machine-approach-functional-human-phosphosites.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.