

Collection of open reading frames at 80 percent of human protein-coding genes

February 25 2016

An international collaboration of organizations, including Dana-Farber Cancer Institute, has reached a milestone in creating a library of complete genetic blueprints for the thousands of different proteins in human cells. The collection - consisting of open-reading frames (ORFs), the portions of genes that code for full-length proteins - is an essential resource for scientists studying the basic mechanics of human cells and how those processes go awry in disease.

In a paper published today by *Nature Methods*, the ORFeome Collaboration (OC), a group of 13 academic, commercial, and governmental organizations, announced that its collection of ORF clones now comprises about 80 percent of all protein-coding genes in [human cells](#) - 17,154 in all, and counting. It is the largest human-gene DNA collection openly available to the worldwide research community.

"The OC ORF collection can be of enormous utility in a broad range of research applications," said the paper's senior author, David E. Hill, PhD, associate director of the Center for Cancer Systems Biology (CCSB) at Dana-Farber, one of the founding institutions of the OC. "To explore cell physiology in a comprehensive way, scientists need a resource that allows them to express virtually any cell protein of interest. The OC is a unique and valuable tool for that type of work."

Thousands of scientists have used OC-supplied ORF clones in their research since the collaboration began in 2005. Applications include large-scale mapping of protein-protein interactions; production of

recombinant human proteins; functional screening of specific proteins; development of disease-specific protein interaction networks; studies on the effect of knocking down or knocking out key proteins in cells, and other uses.

The clones are available from multiple OC distributors around the world at minimal cost, with no restrictions by the OC on their use. Information on the collaboration and on ordering clones is available at the OC website: <http://www.orfeomecollaboration.org/>.

"This website also has a searchable database where we provide rich annotation of clones and encoded proteins to enhance utilization in the community," said Stefan Wiemann, PhD, of the German Cancer Research Center (DKFZ), Heidelberg, Germany, the first author of the study.

Each ORF contains the protein-coding regions of a specific gene. The ORF clones are encased in plasmids, which are injected into bacteria and stored in freezers at the OC's multiple distribution sites. The clones are provided in the Gateway[®] vector format, which allows for easy transfer to a large variety of vectors for expressing the corresponding proteins using for example *Escherichia coli*, yeast, and mammalian cells, or even cell-free expression systems.

The OC grew out of informal discussions among researchers at human ORFeome conferences sponsored by the CCSB at Dana-Farber in the early 2000s. "Attendees from various institutions began discussing what they were doing in the area of generating and validating ORFs," Hill explained. "We began to think about how we could work together to produce the largest possible collection."

"Different members of the OC have performed different roles in its operation," Hill continued. "Some groups have worked on adding new

clones to the collection, some do DNA sequencing, or concentrate on quality control of the ORFs and archiving them for members. Some do informatics work, while others are mainly involved in distribution. Reaching the current milestone has required a concerted effort from a very diverse group of people and organizations. Everyone involved has made an important contribution - which has made this a very enjoyable and productive collaboration."

This phase represents the "end of the beginning" as OC members are continuing to work together to expand the human ORFeome as well as adding ORFeomes for other model organisms.

Provided by Dana-Farber Cancer Institute

Citation: Collection of open reading frames at 80 percent of human protein-coding genes (2016, February 25) retrieved 26 April 2024 from <https://medicalxpress.com/news/2016-02-percent-human-protein-coding-genes.html>

<p>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.</p>
--