Database Detectives: Exploring Public Genomic Databases

## Overview

This activity guides students through using publicly-available genomics databases. Students start with four *Drosophila* genes and learn how to find information about these genes, as well as information about homologous genes in mice, zebrafish, and humans. Finally, they use human disease databases to identify the human health consequences of mutations to the *Drosophila* homologs.

This activity is loosely based on the game Clue. Students need to match four *Drosophila* genes to the correct experimental data and background research.

**Duration:** Coming soon!

## Learning Objectives

1. Explore commonly-used public genomic databases to become familiar with the biological information on them.
2. Identify genomic information for four different *Drosophila* genes.
3. Compare how gene location and expression differs between species for homologous genes.
4. Determine the associations between *Drosophila* gene homologs and human diseases.

## Materials

* Students will need:
  + An internet connection for this activity. Students will use these websites: [*FlyBase*](https://flybase.org/), [*Mouse Genome Informatics*](https://informatics.jax.org/), [*ZFIN (Zebrafish Information Network*](https://zfin.org/) , [*NCBI*](https://www.ncbi.nlm.nih.gov/datasets/gene/), [*OMIM*](https://omim.org/), and [*MalaCards*](https://www.malacards.org/)
* Download the student activity as:
  + [Web page](https://genomicseducation.org/module/database_detectives_student_guide.html)
  + [Word (docx)](https://github.com/fhdsl/GEMs/raw/main/docs/docx/module/database_detectices_student_guide.docx)
  + [Google Doc](https://docs.google.com/document/d/1mlyqdhbFCfYPEdiFKX9KH2FCy34EL8Eg/edit?usp=sharing&ouid=101727846555185296348&rtpof=true&sd=true)
* Google Slides presentation available for borrowing images [here](https://docs.google.com/presentation/d/1fy3FwkSLETnRUdf8huddq2n4qcb6eOyzwVg05dbudII/edit?slide=id.g2c8182c8825_0_50#slide=id.g2c8182c8825_0_50).
* Answer key available [here](https://docs.google.com/document/d/1QLoxy045YQydpZxQfaDCVLSRSi4fU00ceLBDNblAdsA/edit?tab=t.0). Please message Ava Hoffman (ahoffma2 at fredhutch dot org) to get access.

## Scientific Topics

In addition to working with public genomic databases, students are exposed to the following concepts:

* Curated Databases: databases where information is screened, validated, and organized by experts to ensure the data are accurate and reliable. Information from these databases is typically standardized and organized in a consistent way so it is easily accessible and searchable.
* Homologs and Orthologs: genes or proteins across species that are similar to each other due to having a common evolutionary origin. An ortholog is a special type of homolog, in which both genes *also* have similar functions within different species.
* Model Organisms: species used by researchers to study specific biological processes and diseases in an effort to understand how how the processes and diseases work in humans.
* Gene Ontology: a bioinformatics initiative focused on standardizing the vocabulary and annotations researchers use to describe genes, proteins, and data about gene and protein function.
* Genome Assembly: a version of the genome map.

## Outline

* Part 1: Using FlyBase
* Part 2: Identifying Homologs; Using Mouse Genome Informations, ZFIN, and GeneCards
* Part 3: Using OMIM and MalaCards
* Part 4: Solving the Data Mystery