HDinHD: A Data Portal for Huntington's Disease Research
https://www.hdinhd.org/

Viktoria Andreeva[3], Richard A Blevins[2], Irina Murasheva[3], Kabenla Armah[3], Chuck Borromeo[3], Ignacio Munoz-Sanjuan[1], Thomas F. Vogt[1], Liliana Menalled[1], David Howland[1], Dan Felsenfeld[1], Jian Chen[1], Jim Rosinski[1], Vahri Beaumont[1], Jeff Aaronson[1]

[1] CHDI Management/CHDI Foundation, Princeton, NJ, USA. [2] Bridlewood Consulting LLC, Flower Mound, TX, USA. [3] Rancho Biosciences LLC, San Diego, CA, USA.

Abstract

HDinHD (Huntington's Disease in High Definition; HDinHD.org) is an open online portal developed by CHDI and its partners to serve as a collaborative enabler for the Huntington's Disease (HD) research community [1]. CHDI Foundation is a not-for-profit seeking to accelerate the development of therapeutics for Huntington's disease. HDinHD presents a synthesized view of HD-related scientific data and tools through: a) actively sharing curation, analyses and modeling results with the community; b) highlighting and enhancing HD experimental data pulled from a cross-section of sources; and c) incorporating community data and tools into the federated portal. HDinHD's Downloads area serves as a reference library, providing access to results from computational analysis and modeling projects as well as to large compilations of curated HD experimental datasets. These resources are meant to be downloaded by HDinHD users to their local environment either for review or for integration into internal databases or computational pipelines. The Tools area provides access to a federated set of interactive browsing, searching and visualization tools operating over HD-related data. The HD Explorer tool, first released in April 2021, provides a single integrated framework where researchers can explore a wide range of diverse yet richly interconnected HD scientific data. CHDI and its partners curated and analyzed data from hundreds of HD studies spanning complementary experimental data types (e.g., perturbations studies, HTT protein-protein interaction studies, molecular studies) using standard, controlled vocabularies, and consistent methodologies. Data sources include published literature and 'omics studies, as well as dozens of previously unreleased CHDI in vivo pharmacological studies. HDinHD is a dynamic system that continues to incorporate additional community tools and complementary experimental data while expanding functionality and improving user experience.

HDinHD Home Page

HDinHD home page content includes RSS feeds from HDBuzz, HD related literature aggregated from PubMed, bioRxiv and medRxiv and an HD-related news feed. HDBuzz (https://en.hdbuzz.net) is a resource translating HD research news into plain language for patients and families [6]. Downloads, Tools, and New in HDinHD sections are available as independent tabs from the home page to registered users.
Prospective users may register for an account directly from the homepage to gain full access.

Downloads

Striatum Disease Signature	Manuscript describing generation of molecular disease signatures in HD mice and supplemental files detailing results.
Mouse Allelic Series	Raw, processed and analyzed molecular and behavioural data from the Mouse Allelic Series project.
GWAS Studies	Topic reports for genes implicated by early GWAS results generated by the GeM-HD consortium.
DNA Repair & Handling	Topic report plus visual and computable DNA repair pathways.
Causal Modeling Results	Simulation and other results from a series of causal models built from Mouse Allelic Series molecular & behavioural data.
Curated HD Datasets	Independent slices of HD experimental data underlying integrated HDinHD's HD Explorer Tool.

Downloads: Curated and Analyzed HD Datasets

The following **datasets** are available as independent downloads. Collectively, they are also available within HD Explorer (see Tools section), an interactive application that supports integrative mining of HD experimental data.

- **Htt Protein-Protein Interactors**
- **HD Gene Set Enrichment Library (HDSigDB)**
  To provide rich functional context for HD gene set enrichment analysis, we developed an HD-relevant gene set library called **HDSigDB**. The core of **HDSigDB** was derived from curation and analysis of HD and triplet-repeat expansion disease studies deposited in GEO and ArrayExpress. Additional sources of gene sets include PRIDE, selected PubMed articles and DNA Damage Response pathways.
- **HD Mouse Model Catalog**
- **Perturbation Studies**
- **HD Omics Studies**
- **Publications and Reports**

Federated Set of HD Tools Authored by the Community

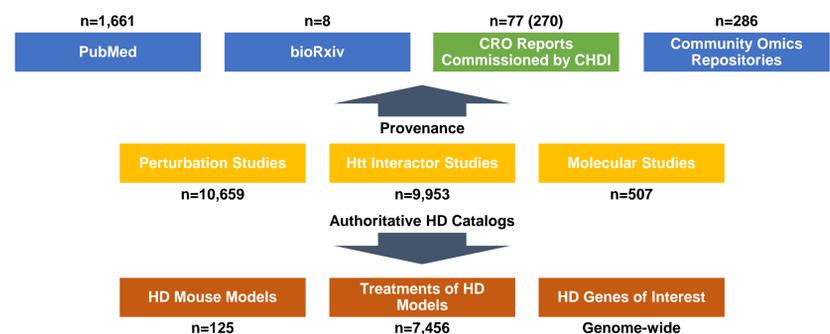|	
HD Explorer	Integrated network of HD experimental data curated and analyzed from the literature, community 'omics repositories and newly-released internal CHDI reports.
GeM MOA SNP Viewer	Summary findings from Huntington's disease genome-wide association studies that seek out genes influencing the pathogenesis and expression of Huntington's disease. (GeM-HD Consortium)
GeM Euro 9K	Visualization tools and summary results of a genome-wide association study to identify genetic modifiers of Huntington's disease. (GeM-HD Consortium)
ASViewer	Visualization of Q-length and age dependent gene and protein expression data from brain and peripheral tissues of the Mouse Allelic Series.
Adult Astrocyte RNAseq Explorer	Visualization tool providing Astrocyte gene expression profiles across brain regions and HD disease models. (Khakh Lab, UCLA)
HD Proteome Base	Proteomics query tool displaying differential expression data from brain and peripheral tissues of the Mouse Allelic Series, as well as baseline proteomic and phosphoproteomic data from the R6/2 mouse model. (Schaab, Evotec)
BioGemix Suite	Browsable knowledgebase of integrated HD animal model data using precision machine-learning and 3D-visualisation of RNA-seq data in brain structures of HD model mice. (Neri Lab, INSERM)
Enricher	Gene set enrichment analysis tool operating over a large, diverse collection of gene set libraries including HDSigDB, a gene set library containing HD and HD-related gene sets. (Ma'ayan Lab, Mt. Sinai)
REPAIR	Differential gene expression analysis results from >350 HD and HD-related datasets from GEO. (Coppola Lab, UCLA)

Tools: Enrichr

Enrichr (https://maayanlab.cloud/Enrichr/) [3-5] is now available as a federated application within HDinHD's Tools section. Enricher is a widely used gene set enrichment analysis package that leverages a comprehensive, centrally-managed collection of gene set libraries to provide rich functional context to user-submitted gene sets. As of December 2021, Enricher includes **HDSigDB**, a gene set library available within HDinHD Downloads and integrated within HD Explorer.

Tools: HD Explorer

HD and HD-related Experimental Data Curated & Analyzed from Internal and External Sources. Shared HD Catalogs allow facile pivoting on Mouse Models, Treatments and Genes/Targets.

HD Explorer Entry Portals

Contains >1,700 PubMed articles and CHDI reports referred to within the HDinHD ecosystem. Most publications are bi-directionally linked to experimental data detailed across HDinHD.

Includes >500 HD and HD-related datasets drawn from >280 GEO, PRIDE and ArrayExpress molecular studies. Datasets are curated and analyzed, and links are provided to gene sets resulting from these analyses.

Includes results of perturbations studies performed in the HD context that have been curated from >1200 PubMed articles and CHDI reports.

Contains >9,900 HTT (either mutant or WT) protein-protein interaction reports – both positive and negative – curated from the literature.

HDSigDB is a Gene Set Enrichment Library designed to provide rich functional context for HD-related gene set enrichment analysis. >2,800 gene sets have been derived from publications, CHDI internal reports and community 'omics repositories.

Provides a catalog of >120 HD Mouse models described in the literature. Individual mouse model lines are bi-directionally linked within HD Explorer to/from published experimental data generated using that mouse model.

Each box on the Explorer entry circle, as well as the central gene name search box, provide distinct portals into the integrated HD Explorer application. Once inside a portal, users can pivot via a rich set of semantic links to explore related HDinHD federated tools or adjacent HD Explorer data sections. For example, after entering a gene name, users check on the expression of that gene/protein in the Mouse Allelic Series and can visit sets of experimental studies performed on that gene within the HD context.
Component datasets are available on the HDinHD Downloads tab for labs who wish to incorporate HD Explorer batch data to facilitate internal data mining efforts.

Acknowledgements

References

1. Aaronson J et al, J Huntington's Dis 2021,10(3):405-412.
2. Obenauer J et al, bioRxiv, https://doi.org/10.1101/2022.02.04.479180.
3. Chen EY et al, BMC Bioinformatics 2013, 128(14).
4. Kuleshov MV et al, Nucleic Acids Research 2016, gkw377.
5. Xie Z et al, Current Protocols 2021, doi:10.1002/cpz1.90.
6. Wild E et al, Trends Mol Med 2012 Jan;18(1):1-3, doi: 10.1016/j.molmed.2011.09.003