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

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

[1]CHDI, New York, NY, 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 designed for Huntington's Disease (HD) researchers [1]. HDinHD presents a synthesized view of HD experimental data and highlights a federated set of visualization and analysis tools developed by HD scientists. Researchers can interactively explore an interconnected set of experimental data, visualize analytical results or download datasets to incorporate into their own local databases and computational pipelines.

The HDinHD team continues to monitor the literature and community omics repositories to identify emerging HD experimental data. New HD studies are curated and analyzed according to established vocabularies, methods and pipelines and integrated into the HDinHD environment. HDinHD seeks to be responsive to community needs. Input is welcomed and may be submitted either through the Feedback link on the HDinHD website or by contacting CHDI directly.

Downloads

Striatum Disease Signature	Manuscript describing generation of molecular disease signatures in HD mice and supplemental files detailing results [2].
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 GeM-HD results.
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 and behavioural data.
Curated HD Datasets	Independent slices of HD experimental data, including **HDSigDB**, that underlie **HD Explorer**.

HDSigDB Gene Sets

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** is derived from curation and analysis of HD and triplet-repeat expansion disease studies deposited in GEO, ArrayExpress, and PRIDE. Additional sources of gene sets include selected PubMed articles (publication-based) and DNA Damage Response pathways (CHDI reports).

In Summer 2023, we leveraged large-scale HD perturbation data to introduce a new type of gene set within **HDSigDB.** "Perturbation gene sets" are constructed by grouping a set of perturbed genes that share a readout – e.g. aggregation – and result – e.g. the treatment effect (perturbed HD vs HD) shows amelioration or exacerbation of a phenotype as compared to the genotype effect (HD vs WT).

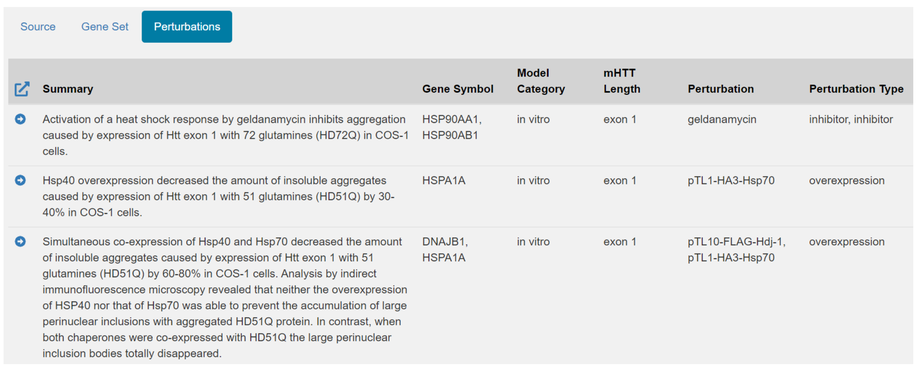
Gene Set Source	# of Gene Sets
CHDI report-based	38
Dataset analysis-based	3,350
Perturbation-based	465
Publication-based	1,238

To provide further granularity and interpretability, we broke down each group independently by 2 experimental factors related to the HD Model:

Does the experiment utilize an **in vitro**, **in vivo** or **ex vivo** HD model?

Does the HD Model contain a **fragment** or **full-length** mHTT construct?

Each perturbation gene set has a direct link to each perturbation experiment contributing to the perturbation gene set.



Summary	Gene Symbol	Model Category	mHTT Length	Perturbation	Perturbation Type
Activation of a heat shock response by geldanamycin inhibits aggregation caused by expression of Htt exon 1 with 72 glutamines (HD72Q) in COS-1 cells.	HSP90AA1, HSP90AB1	in vitro	exon 1	geldanamycin	inhibitor, inhibitor
Hsp40 overexpression decreased the amount of insoluble aggregates caused by expression of Htt exon 1 with 51 glutamines (HD51Q) by 30-40% in COS-1 cells.	HSPA1A	in vitro	exon 1	pTL1-HA3-Hsp70	overexpression
Simultaneous co-expression of Hsp40 and Hsp70 decreased the amount of insoluble aggregates caused by expression of Htt exon 1 with 51 glutamines (HD51Q) by 60-80% in COS-1 cells. Analysis by indirect immunofluorescence microscopy revealed that neither the overexpression of HSP40 nor that of Hsp70 was able to prevent the accumulation of large perinuclear inclusions with aggregated HD51Q protein. In contrast, when both chaperones were co-expressed with HD51Q the large perinuclear inclusion bodies totally disappeared.	DNAJB1, HSPA1A	in vitro	exon 1	pTL10-FLAG-Hdj-1, pTL1-HA3-Hsp70	overexpression

HDSigDB is available within HDinHD Downloads as lists of human and mouse gene symbols and Entrez gene IDs. It is also integrated within HD Explorer. **HDSigDB** is also included in the Enricher, a gene set enrichment analysis package from Ma'ayan Lab (Mt. Sinai) (https://maayanlab.cloud/Enrichr/) [3-5].

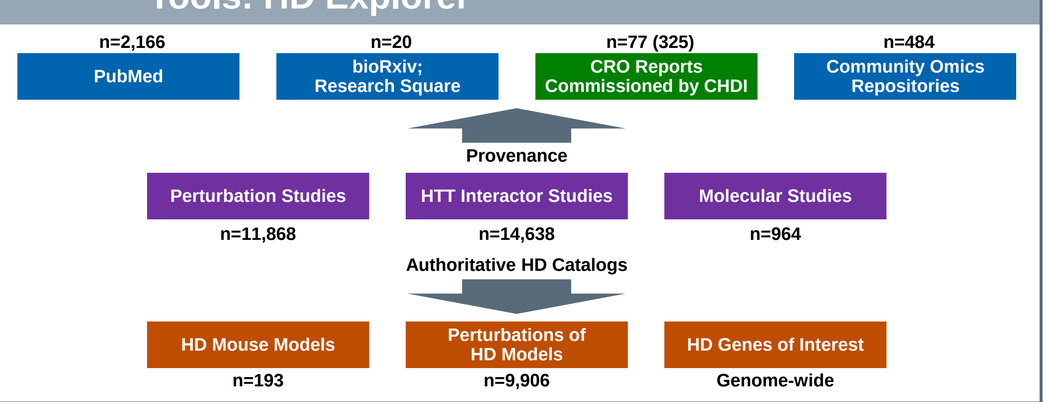
Federated Set of HD Tools Authored by the Community

CHDI	**HD Explorer**	Integrated network of HD experimental data curated and analyzed from the literature, community omics repositories and internal CHDI reports.
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.
CHDI	**ASViewer**	Visualization of Q-length and age-dependent gene and protein expression data from brain and peripheral tissues of the Mouse Allelic Series.
Khakh Lab (UCLA)	**Adult Astrocyte RNAseq Explorer**	Visualization tool providing Astrocyte gene expression profiles across brain regions and HD disease models.
Neri Lab (INSERM)	**Brain-C lab HD Knowledge base**	Browsable knowledgebase of integrated HD animal model data using precision machine-learning and 3D-visualization of RNA-seq data in brain structures of HD model mice.
Ma'ayan Lab (Mt. Sinai)	**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.
Yang Lab (UCLA)	**CoExMap Viewer**	Visualization tool allowing exploration of results and underlying data of a large-scale Weighted Gene Co-expression Network Analysis (WGCNA) of hundreds of samples from intact mouse striatum at 6-months of age as well as from gene set enrichment analysis of transcriptomic signatures of differentially expressed genes from 52 heterozygous HD knockout mice and wildtype controls [6].

Tools: HD Explorer

HD and HD-related Experimental Data Curated & Analyzed from Internal and External Sources.

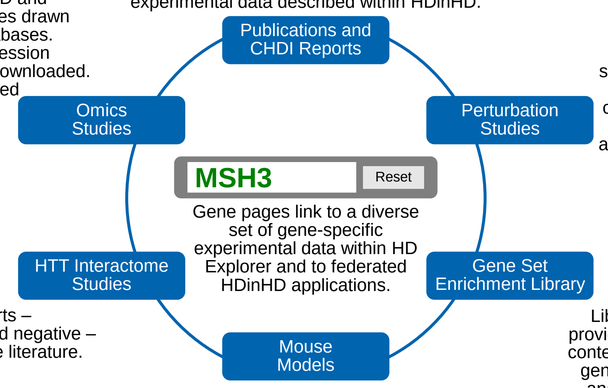
Shared HD Catalogs allow facile pivoting on Mouse Models, Treatments and Genes/Targets.



HD Explorer Entry Portals

Curated and analyzed datasets from HD and adjacent diseases drawn from public databases. Differential expression results can be downloaded. Links are provided to gene sets resulting from analyses.

HD-centric publications and reports linked to experimental data described within HDinHD.



Results of perturbations studies performed in the HD context described in publications and CHDI reports.

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. Gene sets have been derived from publications, CHDI internal reports and community omics repositories.

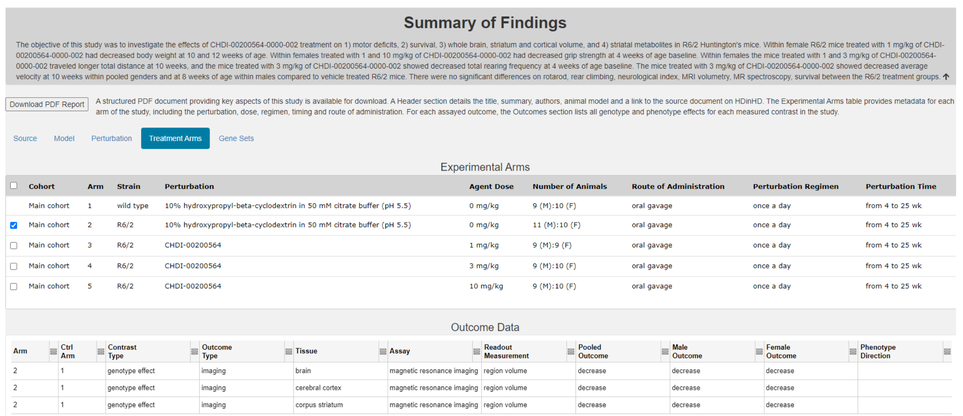
HD models and models containing modifications to a gene considered potentially relevant to HD. Individual models are linked to published experimental data generated using that mouse model. Links to JAX are provided where known.

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. 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. After entering a gene name, users can check on the expression of that gene/protein in the Mouse Allelic Series and visit sets of experimental studies performed on that gene within the HD context.

Component datasets are available for downloads to support labs who wish to incorporate batch data from HD Explorer as part of internal databases and analytical pipelines.

Perturbation Studies

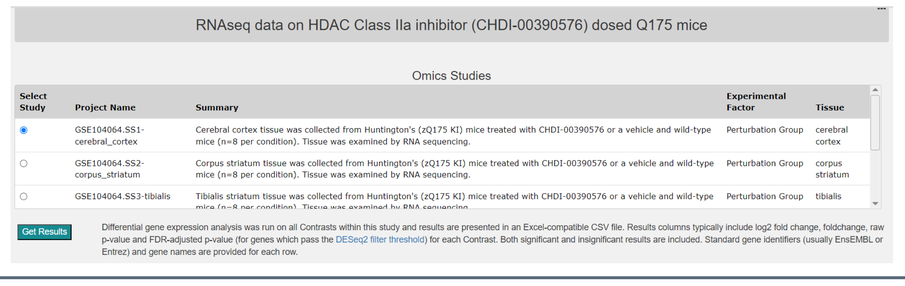
Outcomes are now presented at the arm-level for >4,250 studies drawn from >600 publications and reports. Drilldown pages for these studies now include PDF download files summarizing key experimental data from these studies, including arm-level metadata and genotype effects, chemically-induced effects and phenotype effects for all outcomes.



Omics Studies

Over 474 HD and HD-related (most often, neurodegenerative repeat-expansion diseases) studies identified largely from within public omics repositories have been curated using controlled vocabularies and ontologies to standardize metadata. The studies were divided into >960 datasets based on platform, tissue, etc. The bulk of these datasets were analyzed independently using a consistent and standard methodology. Full differential gene expression results are now available for direct download for 650 core platform datasets.

Platform	# of Studies
RNAseq	330
miRNAseq	26
ATACseq	2
microarray	263
proteomics	29



Acknowledgements

HDinHD is funded and developed by CHDI Foundation, Inc., a nonprofit biomedical research organization exclusively dedicated to collaboratively developing therapeutics that will substantially improve the lives of those affected by Huntington's disease. HDinHD was launched in 2015 in partnership with the laboratory of Giovanni Coppola (UCLA). Colleagues at Rancho BioSciences contributed data curation, data analysis, data modeling and software/data engineering support, and Bridlewood Consulting contributed solutions architecture, systems and software engineering support. CHDI thanks the investigators who have kindly contributed to HDinHD's federated set of community-developed tools.

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.
4. Xie Z et al, *Current Protocols* 2021, doi:10.1002/cpz1.90.
5. Wang N et al, *Neuron* 2022, 110(20):3318-3338.