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

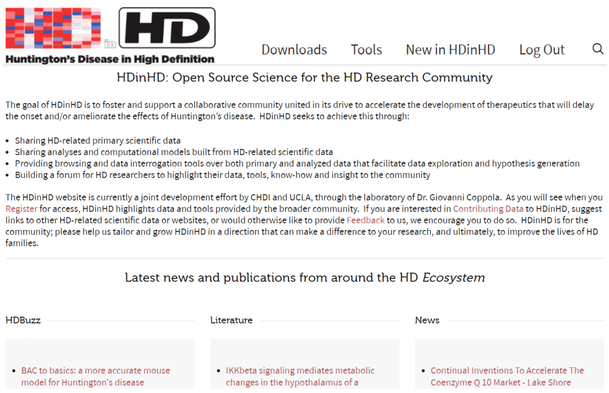
Viktoria Andreeva[3], Richard A Blevins[2], Irina Murasheva[3], Kabenla Armah[3], Chuck Borromeo[3], Ignacio Munoz-Sanjuan[1], Thomas F. Vogt[1], Liliana Menalled[1], David Howland[1], Dan Felsenfeld[1], Jian Chen[1], Jim Rosinski[1], Vahri Beaumont[1], Jeff Aaronson[1]

[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]. 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



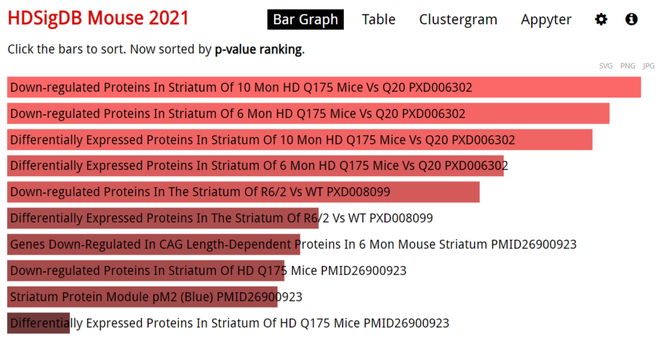
HDinHD home page content includes RSS feeds from HDBuzz, HD related literature aggregated from PubMed, bioRxiv and medRxiv and an HD-related news feed. Downloads, Tools, and New in HDinHD sections are available as independent tabs from the home page to registered users.
Prospective users may register for an account directly from the homepage to gain full access.

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-HD Consortium	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.
	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.
Schaab (Evotec)	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.
Neri Lab (INSERM)	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.
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.
Coppola Lab (UCLA)	REPAIR	Differential gene expression analysis results from >350 HD and HD-related datasets from GEO.

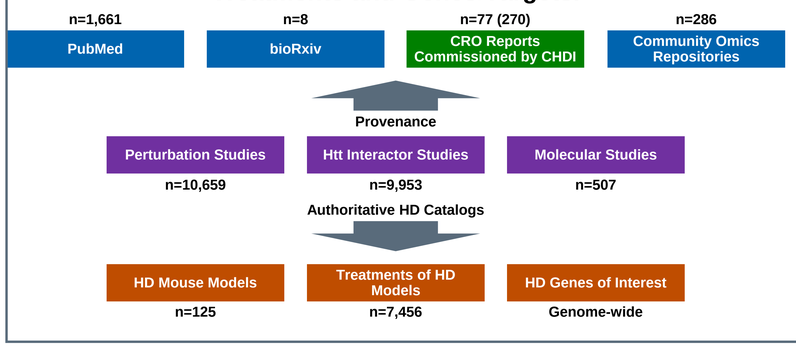
Tools: Enrichr

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.

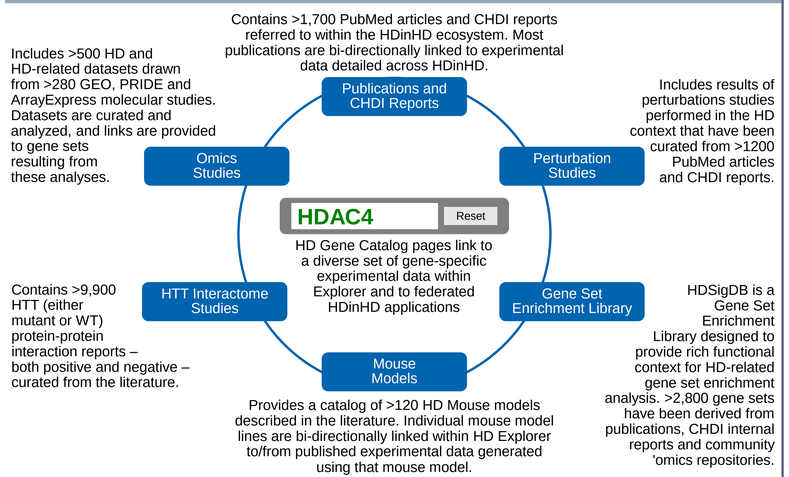


Tools: HD Explorer

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.



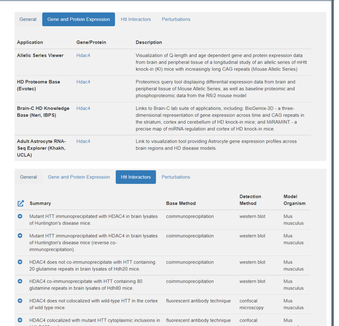
HD Explorer Entry Portals



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.

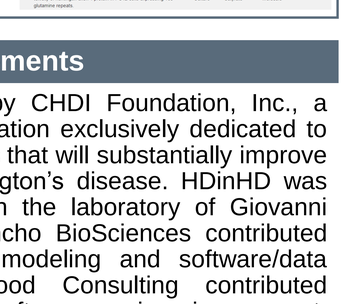
HD Explorer: Content and Navigation

HDAC4-specific links to HDinHD federated tools from HDAC4 gene page



Links to experiments reporting Htt Protein-Protein Interactions with HDAC4

Links to perturbation experiments involving HDAC4 within the HD context



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 GeM-HD results.
Clarivate Analytics	DNA Repair & Handling	Topic report plus visual and computable DNA repair pathways.
GNS Healthcare	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**

Downloads: Striatal Gene And Protein Disease Signatures

Striatal Gene and Protein Disease Signatures developed on Mouse Allelic Series data and validated in external datasets [2]

Str266R

Wt1	Pcdh20	Fam126a	Bcr	Gpr139	Ccm2	Oscar	Fam83d	Fgf3
Onecut1	Tmc3	Hes6	Anks1b	Stk32a	Inhba	Id4	Homer1	Abhd11os
Tnip3	Pcdhb22	Gpr149	Acy1	Ano3	Lzts3	Myh7	Ppp4r4	Plekhg4
Sfmbt2	Pcdhb3	Sh3yl1	Spock3	Pxdn	Wipf3	Rspo1	Pde10a	Ptprv
Rgs13	Asl	Syde2	Ppp1ca	Hpca	Slc26a10	Drd2	Rgs14	Bnipl
Tnfrsf13c	Rbm11	Nsun7	Ephx1	Kctd1	Zbtb46	Impg1	Arpp21	Il2rb
Crnde	Cbx4	Ccdc177	Ppp3ca	Lrrk2	Asb2	Piwil2	Dusp14	Vwa7
Ifnlr1	Nagk	Has1	Ctnnbp2	Itpr1	Npl	Gask1b	Coch	Plk5
Fgfr4	Greb1l	Psme1	Fmnl1	Sh2d5	Cd59a	Gpr6	Atp6v1c2	Spata21
Acy3	Pcdhb9	Cyp4x1	Plcxd1	D7Ertd443e	St8sia2	Ddx11	C4a	Ryr1
Smim24	Lrrn3	Brinp3	Sbsn	Gsg1l	S100a10	Lmo2	Arpp19	Ffar3
Tnnt2	Zfp7	Gba2	Dbpht2	Camk2n1	Drd1	Sfn	Htr1b	Phex
Klhl14	Pcdhb16	Sgk3	Ppp1r1b	Abcc12	Pipox	Rxrg	B3gnt2	Dgat2l6
Slc45a3	Smoc1	Ace	Atp2b1	Rgs9	Pcp4	Dnah1	Car12	Theg
Dsp	Vps37d	Cap1	Camkk2	Ptpn5	Zbtb18	Bank1	Epyc	Sec14l3
Ccdc87	Galns	Fbln5	Sec14l1	Ttll3	Fsbp	Rgs4	Cnr1	Tmem114
Vill	Pcdhb19	Akt2	Rnf207	Il17rc	Pde1b	Mas1	Ankrd35	Odf4
Runx2	Cdh18	Baiap2	Gipc2	Cntn5	Crocc	Acvrl1	Neto2	Mafa
Cbx8	N4bp2	Grm4	Car11	Rasgrp2	Itga5	Arhgef39	Scn4b	Slc4a11
Chdh	Gsto1	Zfyve28	Garem1	Shank3	Kdm4b	Crtnap3	Tnfsf4	Ddit4l
Pcdhb21	Dusp23	Wdr78	Fancb	Dpy19l3	Ago4	Nrep	Ccdc155	Wnt8b
Polr2a	Trpc7	Gpm6b	Jcad	Rgs7bp	Hrk	Kcnh4	Abi3bp	Tmprss6
Pcdhb2	Samd14	Clec12a	Tbc1d4	Adcy5	Rbp4	Hipk4	Myo5c	Myo7b
Rdh12	Cep164	Ppp2r2a	Rps6ka4	Tpm2	Acvr1c	Ssc5d	Lrrc10b	Gpx6
Htr2c	Rgs19	Gstm6	Atf6	Fam184b	Osbpl8	Itga9	Penk	Ifi27l2b
Dsg2	Hook2	Tbc1d8	Dusp18	Malat1	Nrn1l	Ppp1r16b	Shisa2	Sohlh1
Insyn2b	Ltk	Rerg	Tesc	Dock4	Ppp1r1a	Camk1g	Cyp2a5	
Pcdhb12	Zfp711	Hbegf	Chn1	Ddn	Slc39a2	Upb1	Clspn	
Vwa5b2	Gng3	Hebp1	Vrk1	Gpr83	Adora2a	Tcf7	Krt9	
Scn9a	Pcdhb5	Slmap	Kcnab1	Gabrd	Rhobtb2	Dmkn	Ptpn7	

Str115P

Chdh	Grin1	Camkk1	Inf2	Tbc1d8
Acy3	Pitpnm2	Cyld	Itpka	Matn4
Ahi1	Actn1	Rin1	Rem2	Arpp19
Macrod1	Grm5	Calcoco1	Bsg	Kcnip2
Armcx2	Cdkl5	Anks1b	Olfm2	Ano3
Nagk	Atp2a2	Cbr3	Jcad	Spata2l
Dis3	Prkcb	Fbxl16	Ankrd63	Rgs14
Dus3l	Crocc	Pde1b	Dlgap2	Sema7a
Fahd2	Bcr	Cacna2d3	Rasgrp2	Foxp1
Prepl	Baiap2	Synpo	Hpca	Arc
Gprasp1	Sorbs1	Pcp4	Them6	Rgs9
Rbm3	Kctd16	Homer1	Lrrtm1	Sh3tf2
Pfas	Ryr3	Shank3	Ngef	Pde10a
Pck2	Atp2b1	Apoe	Itpr1	Rasd2
Lmnb2	Phyhip	Sh2d5	Ppp1r16b	Tcf20
Psme1	Osbpl8	Gria3	Npl	Scn4b
Mri1	Dlgap3	Ppp4r4	Rgs7bp	Chrm4
Rap1gap	Cap1	Rcn1	Kcnj4	Drd1
Adcy5	Trim46	Ptpn5	Pde7b	Pex5l
Syngap1	Inpp5j	Ppp1r1b	Coch	
Htt	Mast3	Sec14l1	Rps6ka4	
Erc2	Cacnb2	Camkk2	Ntrk3	
Dab2ip	Phactr1	Shisa7	Mlf2	
Mink1	Grm1	Camk4	Wipf3	

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.
5. Xie Z et al, *Current Protocols* 2021, doi:10.1002/cpz1.90.