Rancho BioSciences – CHDI Areas of Collaboration

December 6th, 2022

J. Obenauer
J. Greene
V. Andreeva
L. Brovold

Adding New Value to Your Life Science Data

At Rancho BioSciences,
we unlock hidden value for you through our data curation services.

From clinical to genomic, assays to chemistry, **we create new efficiencies and workflows, so your data is well-organized,** easy to find and ready to be utilized at a moment's notice.

Our brains for hire have hundreds of combined years of experience in medical and pharma data.

We work with you to identify your data needs and we deliver projects beyond expectations, on budget and on time.

Rancho Services

DataCuration

Curation Projects

- Curating internal and public data sets
- Knowledge mining and building content
- F.A.I.R. and harmonization
- Ontologies, dictionaries, MetaData

DATASCIENCE

Data Science Projects

- R, Python and automation
- Data Governance and Data Models
- Catalogs
- Ontology implementation
- Business Analysis

DataManagement

Data Management Projects

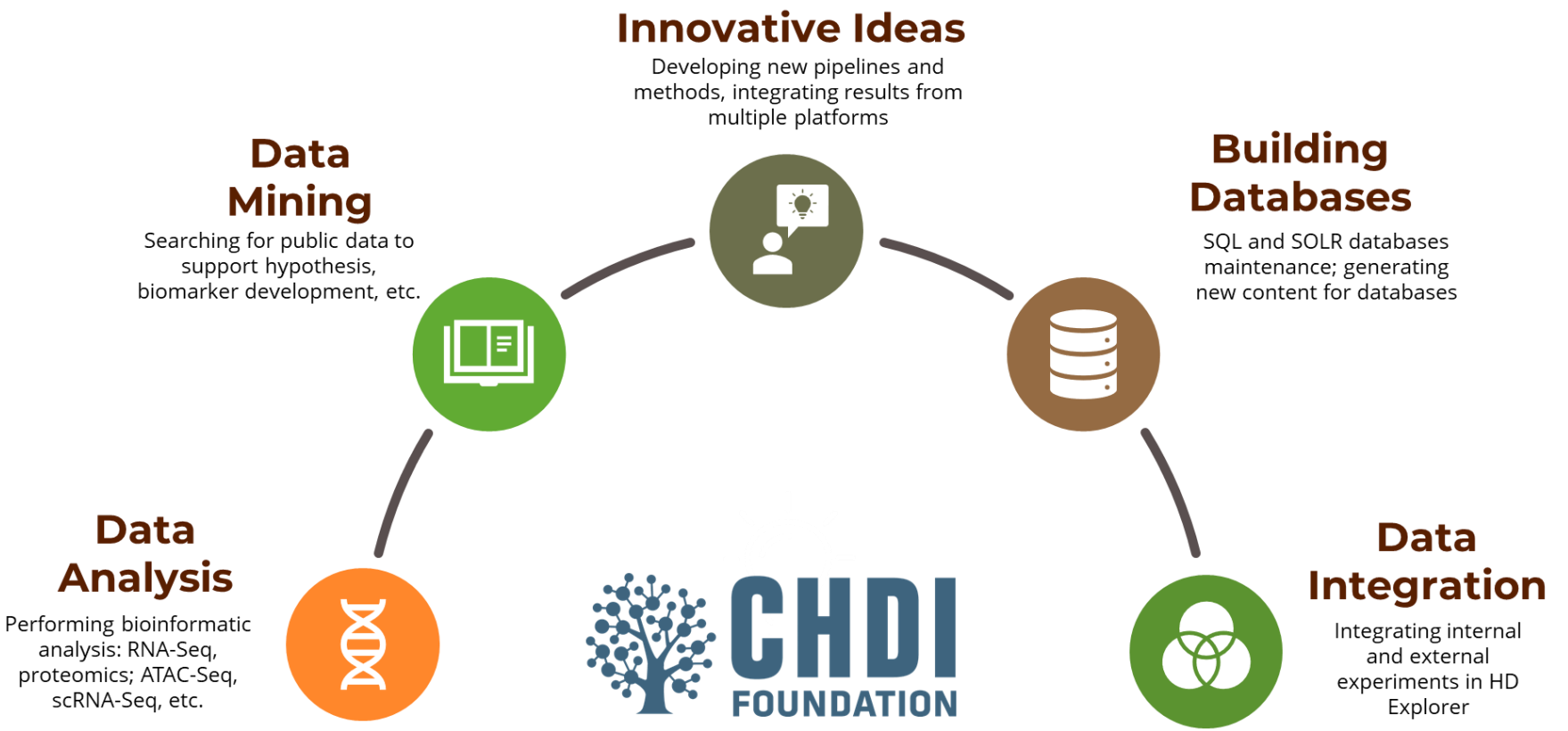
- Data Catalogs
- FAIR data management Strategies and policies
- Regulatory compliance
- Data migration

Bioinformatics

Bioinformatics Projects

- Custom workflows and pipelines
- DNA, RNA, ChIP Seq and single cell analysis
- Variants
- Pathways
- Precision and translational medicine

Overview


Innovative Ideas
Developing new pipelines and methods, integrating results from multiple platforms
Data Mining
Searching for public data to support hypothesis, biomarker development, etc.
Building Databases
SQL and SOLR databases maintenance; generating new content for databases
Data Analysis
Performing bioinformatic analysis: RNA-Seq, proteomics; ATAC-Seq, scRNA-Seq, etc.
CHDI FOUNDATION
Data Integration
Integrating internal and external experiments in HD Explorer

Huntington's disease (HD)

- Rare monogenic neurodegenerative disease for which there is no treatment
  - Movement disorder (like Parkinson's)
  - Cognitive disorder (like Alzheimer's)
  - Psychiatric disorder (irritability, depression, anxiety)
  - Symptoms usually appear at ages 30-40
- Caused by tri-nucleotide CAG repeat expansion in exon 1 of the HTT gene which encodes Huntingtin protein
  - Severity of disease increases with age and number of CAG repeats
  - Leads to form of Huntingtin protein with poly-glutamine (polyQ) repeats (nomenclature like 50CAG or Q50)
- Primary effects are seen as damage to medium spiny neurons in the striatum and cortex

Specific areas to be discussed

- Definition and characterization of disease signatures
- Measuring the reversal or modulation of disease phenotypes in omics studies
- Analysis of alternative splicing events in RNASeq data
- HDinHD resources (Huntington's Disease in High Definition; HDinHD.org)
- Neurodegenerative Data Consortium (NDC)

Rancho BioSciences

Definition and characterization of disease signatures

10 HD data sets: Q175, Q140, R6/2

		Q175 10M			Q175 10M			Q175 10M		
		Full Series Q175 10M		1725	Cohort1Time1 Q175 10M		3056	Cohort1Time2 Q175 10M		2379
Ensembl	Gene	Log2FoldChange	AdjPValue	Significant	Log2FoldChange	AdjPValue	Significant	Log2FoldChange	AdjPValue	Significant
ENSMUSG00000000001	Gnai3	0.024920491322331	0.829955833	0	0.0595341159696	0.605861448	0	0.0946829380189	0.310118823	0
ENSMUSG00000000003	Pbsn	NA	NA	0	NA	NA	0	NA	NA	0
ENSMUSG00000000028	Cdc45	0.591097671565424	7.69735E-05	1	0.6040247086827	0.001666882	1	0.3412297861401	0.018533206	1
ENSMUSG00000000031	H19	-0.75124598069033	0.824962472	0	-0.7440563359969	0.485452662	0	-0.7721351227887	0.365855698	0
ENSMUSG00000000037	Scml2	0.46141332210931	0.020611057	1	0.4151658545698	0.071334112	0	0.4589380810414	0.018560526	1
ENSMUSG00000000049	Apoh	0.033658626741548	0.984838985	0	0.5968653053518	0.629992	0	-0.1301443068195	0.98569626	0
ENSMUSG00000000056	Narf	-0.021307127222134	0.971935052	0	-0.0864228019766	0.426561208	0	-0.0134928105698	0.98569626	0
ENSMUSG00000000058	Cav2	-0.061177864907126	0.684835251	0	0.0552273717392	0.963474772	0	-0.1391799952339	0.483709614	0
ENSM										
ENSM										
ENSM										
ENSM										
ENSM										
ENSM										

Q140 10M			Q140 10M			Q140 6M			Q140 6M		
Full Series Q140 10M		1593	Miniseries Q140 10M		1994	Full Series Q140 6M		1625	Miniseries Q140 6M		984
Log2FoldChange	AdjPValue	Significant	Log2FoldChange	AdjPValue	Significant	Log2FoldChange	AdjPValue	Significant	Log2FoldChange	AdjPValue	Significant
0.0024017205654	0.993379453	0	0.05233436913924	0.50681714	0	-0.0283765074939	0.773129348	0	-0.025918364212	0.895698669	0
NA	NA	0	NA	NA	0	NA	NA	0	NA	NA	0
0.3767318295109	0.020268926	1	0.14561412240974	0.726305576	0	0.22778033694041	0.2340329	0	0.3395860401043	0.146184489	0
-1.1791763172808	0.583727634	0	-2.03087541484897	0.077780778	0	-0.1769057316728	0.988433712	0	-0.859332550541	0.777210335	0
0.3401507950641	0.097724393	0	0.33850250662965	0.346927529	0	0.27607615894653	0.192718336	0	0.4156110397479	0.033326654	1
-0.2997535430845	0.882421704	0	-0.39078459333283	0.934559655	0	-0.3733525311501	0.784503287	0	0.5770834842193	0.812586451	0
-0.0439187568667	0.544234802	0	-0.0772888689674	0.420371011	0	-0.1619447860916	0.000755249	0	-0.065948661656	0.312029359	0
-0.0743810341001	0.505731524	0	-0.11840672600155	0.46347213	0	-0.0935390413369	0.312083517	0	-0.087271728139	0.25447916	0
0.0275419212222	0.936434867										
-0.1217433491783	0.047481377										
0.2269425070049	0.001316888										
-0.1334443394532	0.566104273										
-0.1729710042855	0.936448271										
NA	NA										
-0.356727007342	0.01886428										

Q140 6M			R6/2 3M			R6/2 3M		
Cohort2Time1 Q140 6M		2563	HDAC R6/2 3M		6089	KMO R6/2 3M		5963
Log2FoldChange	AdjPValue	Significant	Log2FoldChange	AdjPValue	Significant	Log2FoldChange	AdjPValue	Significant
-0.074033075859	0.606678538	0	0.1789792152018	0.002303969	0	0.186323233310521	2.4938E-06	0
NA	NA	0	NA	NA	0	NA	NA	0
0.2267303798134	0.21714343	0	1.29577681943	1.81277E-27	1	1.34914547133467	6.30932E-25	1
0.4665478640297	0.966766922	0	-0.5246306165734	0.737163886	0	-0.39821578182257	0.762646348	0
0.444820892604	0.083907079	0	0.8635338440764	1.40221E-07	1	0.830828593937651	7.27329E-09	1
0.1573605678145	0.974958215	0	0.1667803651858	0.968351531	0	0.259056271174236	0.922133482	0
0.049317942974	0.87258265	0	-0.209058754279	2.23394E-06	0	-0.15856993283035	0.011434584	0
-0.3421953902535	0.001984258	1	-0.0614469489499	0.955614242	0	-0.48457741506705	8.77443E-07	1
-0.0030366024779	0.999755781	0	0.0107382753525	0.968351531	0	-0.05124713767136	0.66253573	0
-0.0229494467005	0.982855485	0	-0.3589621542979	7.52392E-13	1	-0.36734768000879	3.21472E-11	1
-0.0974860529765	0.307355813	0	-0.039861849225	0.744096196	0	-0.09588097957305	0.508486273	0
0.0358469373139	0.982855485	0	-0.0539866368592	0.945918381	0	-0.23125951612818	0.344974476	0
-0.2172635105556	0.684620917	0	0.0291559318114	0.968351531	0	-0.4060369960837	0.217972325	0
NA	NA	0	NA	NA	0	NA	NA	0
0.0546500734673	0.982855485	0	-1.4061166648641	9.05711E-09	1	-1.7416182993019	1.78186E-23	1

GEO IDs:
GSE65774 (Full Series)
GSE78274 (Miniseries)
GSE104086 (HDAC)
GSE105158 (KMO)

Overlap of 10 HD RNA-seq experiments

Experiment	Significant Genes	Group Overlap	All Overlap
Full Series Q175 10M	1,725	1,088	287
Cohort1Time1 Q175 10M	3,056		
Cohort1Time2 Q175 10M	2,379		
Full Series Q140 10M	1,593	658	
Miniseries Q140 10M	1,994		
Full Series Q140 6M	1,625	439	
Miniseries Q140 6M	984		
Cohort2Time1 Q140 6M	2,563		
HDAC R6/2 3M	6,089	4,551	
KMO R6/2 3M	5,963		

287 genes were significant in all 10 HD vs. wild type gene lists.

Genes were rejected as being poorly characterized if they were identified only by gene models (like Gm10406), Riken clones (A830036E02Rik), or GenBank accessions (AW49522). This left 266 genes to be validated.

180 (68%) of these genes are in module M2 of a published HD WGCNA analysis, and 54 others (20%) are in module M20. These were the two most CAG-length-dependent modules in that study. (Langfelder P et al., *Nat Neurosci* 2016, 19:623-633, PMID 26900923).

Str266 RNA signature

Str266R									
Wt1	Pcdhb12	Rgs19	Gpm6b	Car11	Ttll3	Pipox	Ddx11	Dusp14	Ptprv
Onecut1	Vwa5b2	Hook2	Clec12a	Garem1	Il17rc	Pcp4	Lmo2	Coch	Bnipl
Tnip3	Scn9a	Ltk	Ppp2r2a	Fancb	Cntn5	Zbtb18	Sfn	Atp6v1c2	Il2rb
Sfmbt2	Pcdh20	Zfp711	Gstm6	Jcad	Rasgrp2	Fsbp	Rxrg	C4a	Vwa7
Rgs13	Tmc3	Gng3	Tbc1d8	Tbc1d4	Shank3	Pde1b	Dnah1	Arpp19	Plk5
Tnfrsf13c	Pcdhb22	Pcdhb5	Rerg	Rps6ka4	Dpy19l3	Crocc	Bank1	Htr1b	Spata21
Crnde	Pcdhb3	Fam126a	Hbegf	Atf6	Rgs7bp	Itga5	Rgs4	B3gnt2	Ryr1
Ifnlr1	Asl	Hes6	Hebp1	Dusp18	Adcy5	Kdm4b	Mas1	Car12	Ffar3
Fgfr4	Rbm11	Gpr149	Slmap	Tesc	Tpm2	Ago4	Acvrl1	Epyc	Phex
Acy3	Cbx4	Sh3yl1	Bcr	Chn1	Fam184b	Hrk	Arhgef39	Cnr1	Dgat2l6
Smim24	Nagk	Syde2	Anks1b	Vrk1	Malat1	Rbp4	Cntnap3	Ankrd35	Theg
Tnnt2	Greb1l	Nsun7	Acy1	Kcnab1	Dock4	Acvr1c	Nrep	Neto2	Sec14l3
Klhl14	Pcdhb9	Ccdc177	Spock3	Gpr139	Ddn	Osbpl8	Kcnh4	Scn4b	Tmem114
Slc45a3	Lrrn3	Has1	Ppp1ca	Stk32a	Gpr83	Nrn1l	Hipk4	Tnfrsf4	Odf4
Dsp	Zfp7	Psme1	Ephx1	Ano3	Gabrd	Ppp1r1a	Ssc5d	Ccdc155	Mafa
Ccdc87	Pcdhb16	Cyp4x1	Ppp3ca	Pxdn	Ccm2	Slc39a2	Itga9	Abi3bp	Slc4a11
Vill	Smoc1	Brinp3	Cttnbp2	Hpca	Inhba	Adora2a	Ppp1r16b	Myo5c	Ddit4l
Runx2	Vps37d	Gba2	Fmnl1	Kctd1	Lzts3	Rhobtb2	Camk1g	Lrrc10b	Wnt8b
Cbx8	Galns	Sgk3	Plcxd1	Lrrk2	Wipf3	Oscar	Upb1	Penk	Tmprss6
Chdh	Pcdhb19	Ace	Sbsn	Itpr1	Slc26a10	Id4	Tcf7	Shisa2	Myo7b
Pcdhb21	Cdh18	Cap1	Dbpht2	Sh2d5	Zbtb46	Myh7	Dmkn	Cyp2a5	Gpx6
Polr2a	N4bp2	Fbln5	Ppp1r1b	D7Ertd443e	Asb2	Rspo1	Fam83d	Clspn	Ifi27l2b
Pcdhb2	Gsto1	Akt2	Atp2b1	Gsg1l	Npl	Drd2	Homer1	Krt9	Sohlh1
Rdh12	Dusp23	Baiap2	Camkk2	Camk2n1	Cd59a	Impg1	Ppp4r4	Ptpn7	
Htr2c	Trpc7	Grm4	Sec14l1	Abcc12	St8sia2	Piwil2	Pde10a	Fgf3	
Dsg2	Samd14	Zfyve28	Rnf207	Rgs9	S100a10	Gask1b	Rgs14	Abhd11os	
Insyn2b	Cep164	Wdr78	Gipc2	Ptpn5	Drd1	Gpr6	Arpp21	Plekhg4	

Recurrence in proteomics

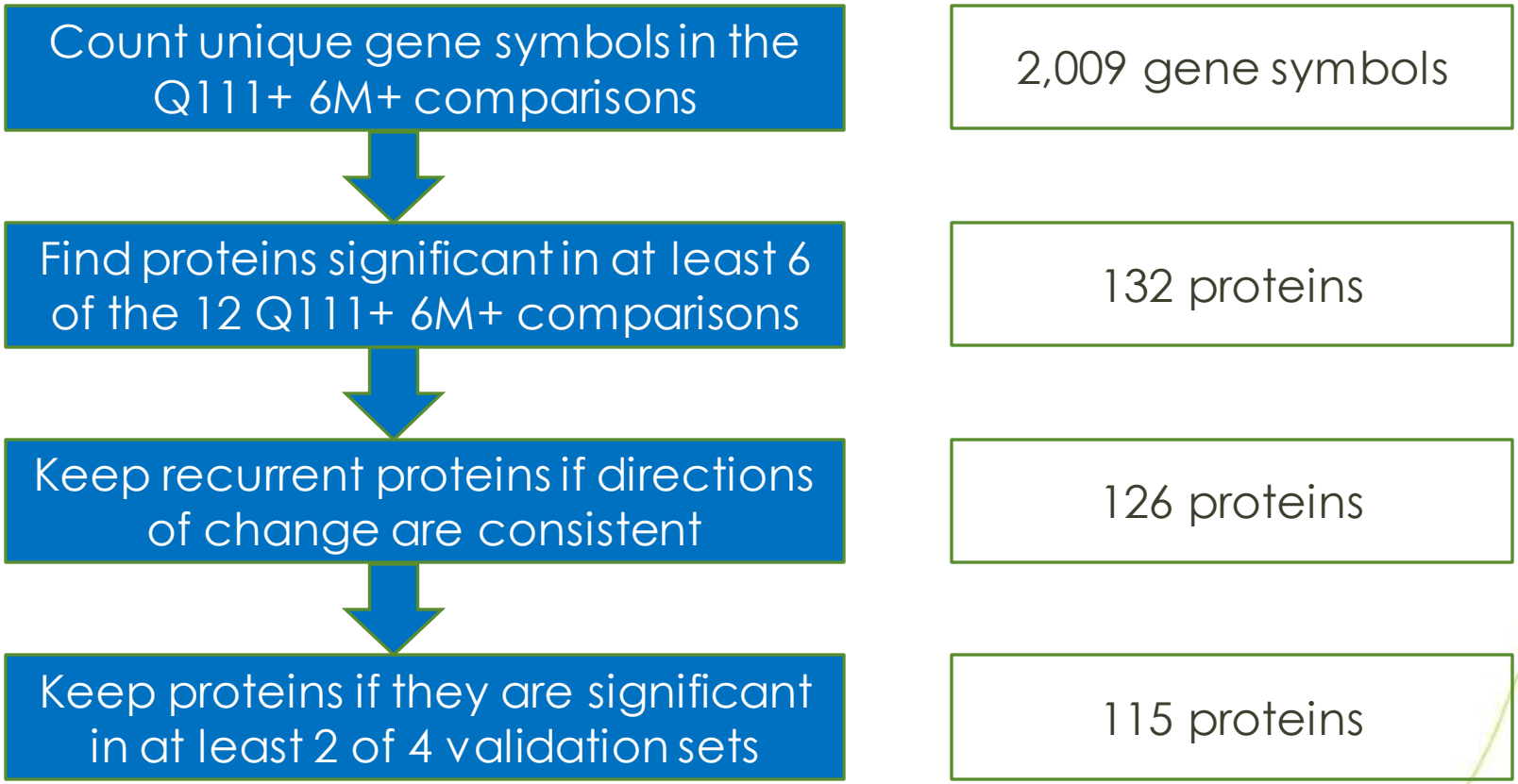
- 12 differential expression tests were used to make the candidate protein signature
  - Q111+ 6M+ set (Q111, Q140, Q175, at 6 or 10 months)
  - 6 Q-length comparisons (like Q175 10M vs. Q20 10M)
  - 6 age comparisons (like Q175 10M vs. Q175 2M)
  - All 12 comparisons are from the PXD006302 striatum data
  - Validation data: PXD013771 (R6/2 2M and R6/2 3M) and 2 unpublished cohorts (R6/2 3M, Q175 10M)

Gene	ID	Q111|6M|age	Q111|6M|qlen	Q175|6M|age	Q175|10M|qlen	Q111|10M|age	Q140|6M|age	Q175|10M|age	Q111|10M|qlen	Q140|10M|qlen
Tcf20	Q9EPQ8	-1.584639	-1.6488095112	-1.7863251263	-2.43350789106	-2.781370342	-1.553639727	-1.862505529	-2.7371520525	-2.3597191172
Rgs9	O54828	-1.40849769	-1.3951802267	-1.7179769484	-1.71922973157	-2.021198188	0	-1.752485715	-1.8139728517	-1.7949211449
Ppp1r16b	Q8VHQ3	-1.39813484	-1.405231261	-1.3795812224	-1.42165814886	-1.510462026	0	-1.486765743	-1.3101027976	-1.3454169598
Dlgap3	Q6PFD5	-1.21595582	-1.2046572976	-1.2308946541	-1.25291855901	-1.439124573	0	-1.295114238	-1.2375323476	-1.209052414
Rasgrp2	Q9QUG9	-1.29394668	-1.3542550546	-1.3309171917	-1.51476390964	-1.559924166	0	-1.453531683	-1.4936829167	-1.5339104292
Itpr1	P11881	0	-1.3418659049	-1.3773340547	-1.43539247961	-1.418228257	-1.305360025	-1.435994003	-1.4118002821	-1.3833426815
Rgs14	P97492	0	-1.3640921898	-1.3782754563	-1.63538675961	-1.526172294	-1.371093715	-1.513157637	-1.5408219004	-1.4240426942
Itpka	Q8									
Shank3	A0									
Arpp19	P5									

Q140|6M|qlen	Q140|10M|age	Q175|6M|qlen	MinFC	AvgFC	MaxFC	ExptCount	RecurCount	QlenCount	AgeCount	Direction
-2.0814040775	-1.862319831	-2.467719174	-1.55363973	-2.09659261	-2.78137034	12	12	6	6	-
-1.5577902089	-1.7621201054	-1.8601546314	-1.39518023	-1.70941159	-2.02119819	12	11	6	5	-
-1.3323635798	-1.4547489092	-1.5286325037	-1.3101028	-1.41573618	-1.5286325	12	11	6	5	-
-1.2664210849	-1.3087364362	-1.3719023438	-1.2046573	-1.27566452	-1.43912457	12	11	6	5	-
-1.4399342564	-1.4252740485	-1.516005011	-1.29394668	-1.4469223	-1.55992417	12	11	6	5	-
-1.4373342492	-1.3785517723	-1.5106995343	-1.30536003	-1.40326393	-1.51069953	12	11	6	5	-
-1.4667779432	-1.4898820983	-1.6672412543	-1.36409219	-1.48881309	-1.66724125	12	11	6	5	-
-1.393386537	-1.3131888623	-1.4707903685	-1.27773297	-1.37676975	-1.50510341	12	11	6	5	-
-1.3523597427	-1.4426319221	-1.5187918854	-1.24343271	-1.50127818	-2.15421033	12	11	6	5	-
-1.7112426575	-1.5465051431	-1.6377529897	-1.3528907	-1.60927354	-2.23058775	12	11	6	5	-

Ran
bio ciences

Protein signature steps



Str115 protein signature

Str115P									
Chdh	Pfas	Grin1	Ryr3	Camkk1	Shank3	Inf2	Ngef	Tbc1d8	Pde10a
Acy3	Pck2	Pitpnm2	Atp2b1	Cyld	Apoe	Itpka	Itpr1	Matn4	Rasd2
Ahi1	Lmnb2	Actn1	Phyhip	Rin1	Sh2d5	Rem2	Ppp1r16b	Arpp19	Tcf20
Macrod1	Psme1	Grm5	Osbpl8	Calcoco1	Gria3	Bsg	Npl	Kcnip2	Scn4b
Armcx2	Mri1	Cdkl5	Dlgap3	Anks1b	Ppp4r4	Olfm2	Rgs7bp	Ano3	Chrm4
Nagk	Rap1gap	Atp2a2	Cap1	Cbr3	Rcn1	Jcad	Kcnj4	Spata2l	Drd1
Dis3	Adcy5	Prkcb	Trim46	Fbxl16	Ptpn5	Ankrd63	Pde7b	Rgs14	Pex5l
Dus3l	Syngap1	Crocc	Inpp5j	Pde1b	Ppp1r1b	Dlgap2	Coch	Sema7a	
Fahd2	Htt	Bcr	Mast3	Cacna2d3	Sec14l1	Rasgrp2	Rps6ka4	Foxp1	
Prepl	Erc2	Baiap2	Cacnb2	Synpo	Camkk2	Hpca	Ntrk3	Arc	
Gprasp1	Dab2ip	Sorbs1	Phactr1	Pcp4	Shisa7	Them6	Mlf2	Rgs9	
Rbm3	Mink1	Kctd16	Grm1	Homer1	Camk4	Lrrtm1	Wipf3	Sh3rf2	

A preprint describing the Str266R and Str115P signatures is available on bioRxiv.org.

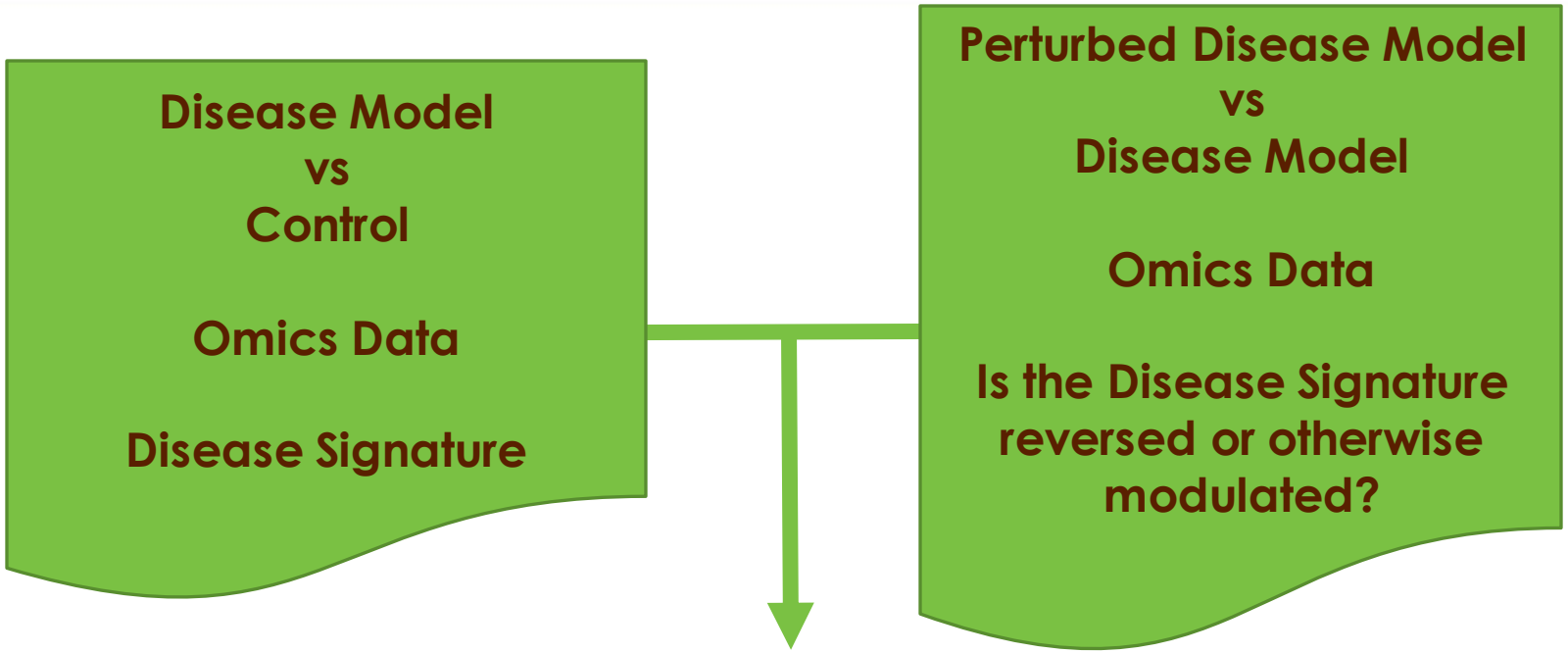
https://www.biorxiv.org/content/10.1101/2022.02.04.479180v2

Str40RP			
Acy3	Atp2b1	Ptpn5	Crocc
Chdh	Camkk2	Rasgrp2	Osbpl8
Nagk	Sec14l1	Shank3	Ppp1r16b
Psme1	Jcad	Rgs7bp	Homer1
Cap1	Rps6ka4	Adcy5	Ppp4r4
Baiap2	Ano3	Wipf3	Pde10a
Tbc1d8	Hpca	Npl	Rgs14
Bcr	Itpr1	Drd1	Coch
Anks1b	Sh2d5	Pcp4	Arpp19
Ppp1r1b	Rgs9	Pde1b	Scn4b

Rancho BioSciences

Reversal/modulation of disease phenotypes in omics studies

Reversal/modulation of disease phenotypes in omics studies



- Bayesian method developed (C. Hartl)
- Gene by gene classification of modulations with associated probabilities
- Roll up to summarize modulation effects within a given perturbation
- Comparison of modulations across studies

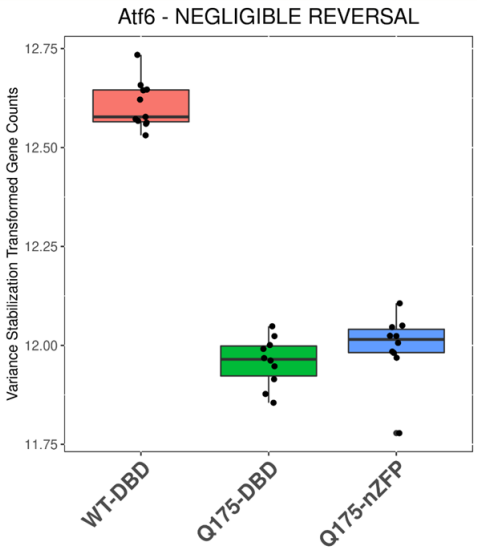
Reversal/modulation of disease phenotypes in omics studies

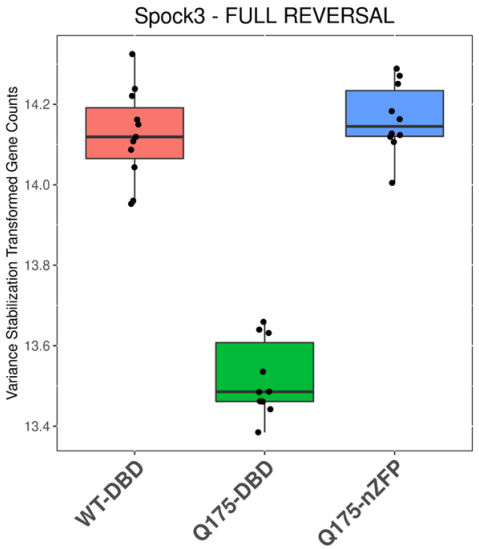
Examples for RNASeq phenotype modulations

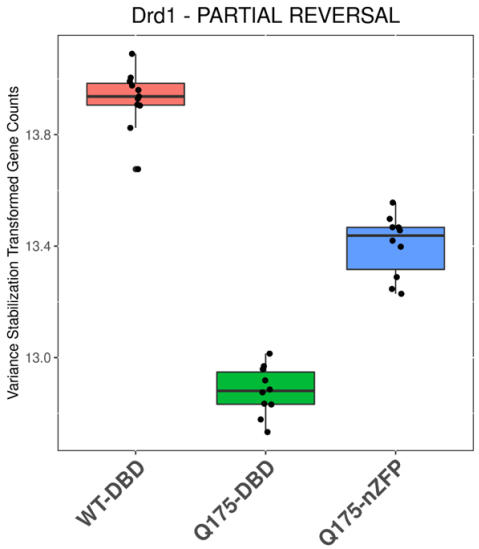
Q175 HD Model vs WT, Striatum

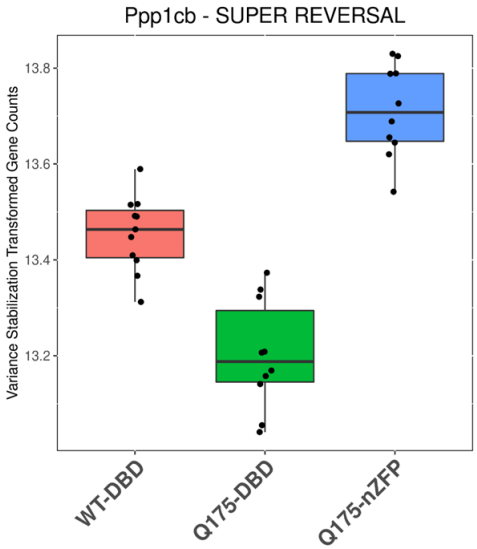
Treatment: CAG-specific ZFP to repress mutant HTT (Zeitler, et al, Nat Med, 2019, PMID: 31263285)

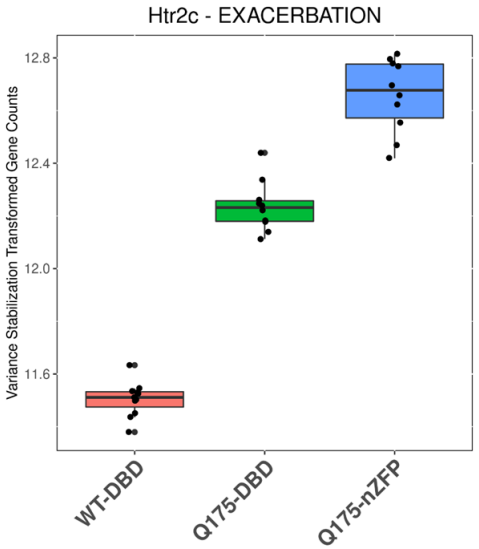
Treatment at 6M
Samples at 12M



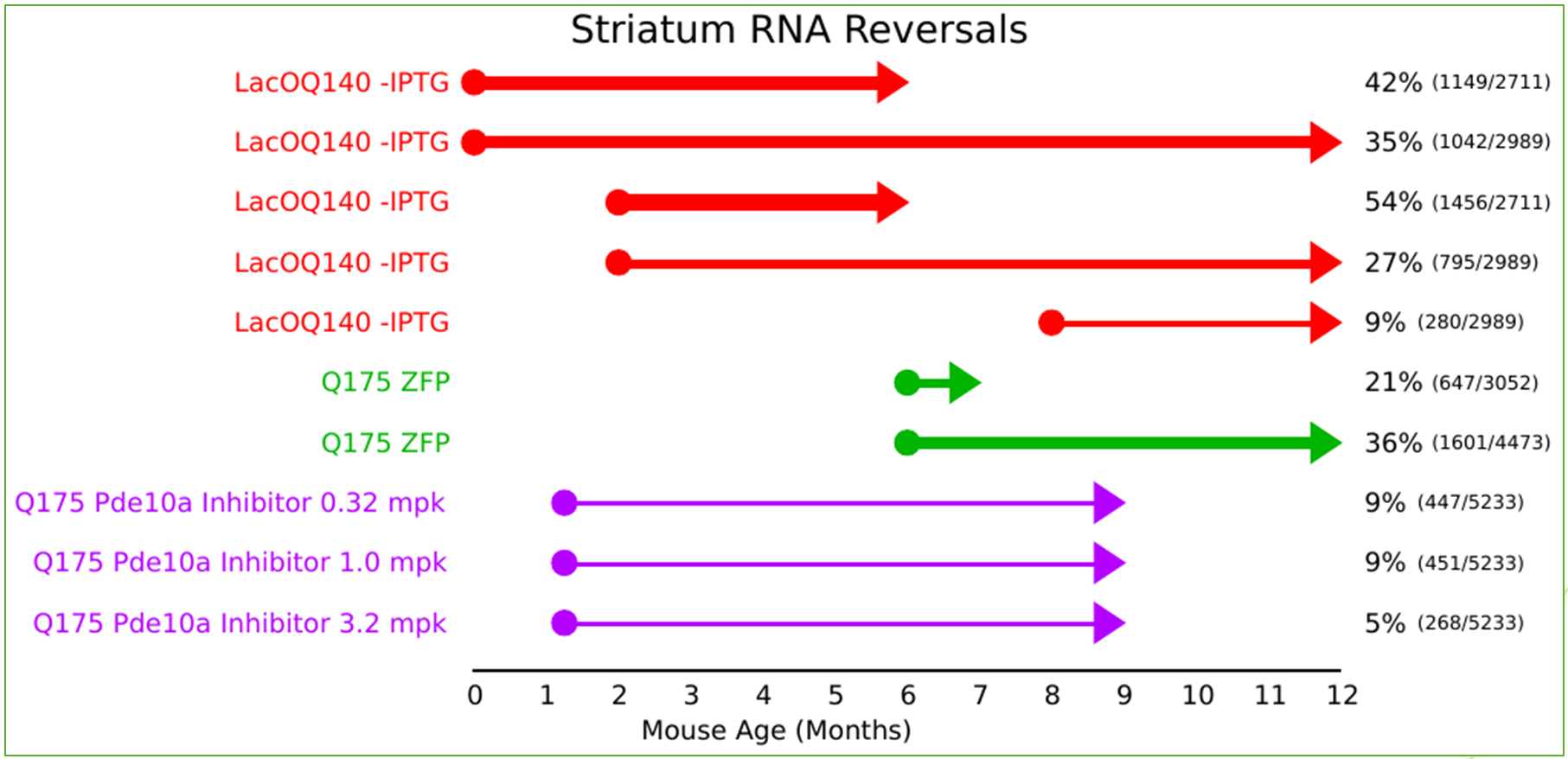








Reversal/modulation of disease phenotypes in omics studies



LacOQ140 results and phenotype reversal method: Marchionini, et al, JCI Insight, 2022, PMID: 36278490

Reversal/modulation of disease phenotypes in omics studies

Scenario	Study	HD Signature	Reversed	Pct. Reversed	Exacerbation	Pct. Exacerbated	Full Reversal	Partial Reversal	Negligible Reversal	Super Reversal
6[24h]6 zQ175 TrkbmAb STR	Q175 TrkmAb	3407	700	21	344	10	263	389	2363	48
6[4h]6 zQ175 TrkbmAb STR	Q175 TrkmAb	4079	762	19	342	8	303	360	2975	99
0[6]6 LacOQ140 -IPTG STR	LacQ140^{I}(*)	2711	1149	42	11	0	522	582	1551	45
0[12]12 LacOQ140 -IPTG STR	LacQ140^{I}(*)	2989	1042	35	4	0	248	786	1944	8
2[4]6 LacOQ140 -IPTG STR	LacQ140^{I}(*)	2711	1456	54	1	0	527	915	1254	14
2[10]12 LacOQ140 -IPTG STR	LacQ140^{I}(*)	2989	795	27	12	0	147	646	2182	2
8[4]12 LacOQ140 -IPTG STR	LacQ140^{I}(*)	2989	280	9	17	1	66	214	2692	0
6[1]7 Q175 nZFP STR	Q175 nZFP	3052	647	21	149	5	190	430	2256	27
6[6]12 Q175 nZFP STR	Q175 nZFP	4473	1601	36	232	5	566	954	2640	81
1[8]9 zQ175 Pde.Inh-0.32mpk STR	Q175 Pde10a Inhib	5233	447	9	216	4	81	364	4570	2
1[8]9 zQ175 Pde.Inh-1.0mpk STR	Q175 Pde10a Inhib	5232	451	9	270	5	119	323	4511	9
1[8]9 zQ175 Pde.Inh-3.2mpk STR	Q175 Pde10a Inhib	5231	268	5	139	3	81	151	4824	36

Rancho BioSciences

Analysis of alternative splicing events in RNASeq data

Analysis of alternative splicing events in RNASeq data

Small Molecule Splicing Modulators (SMSM) treatment of stem cell derived neurons

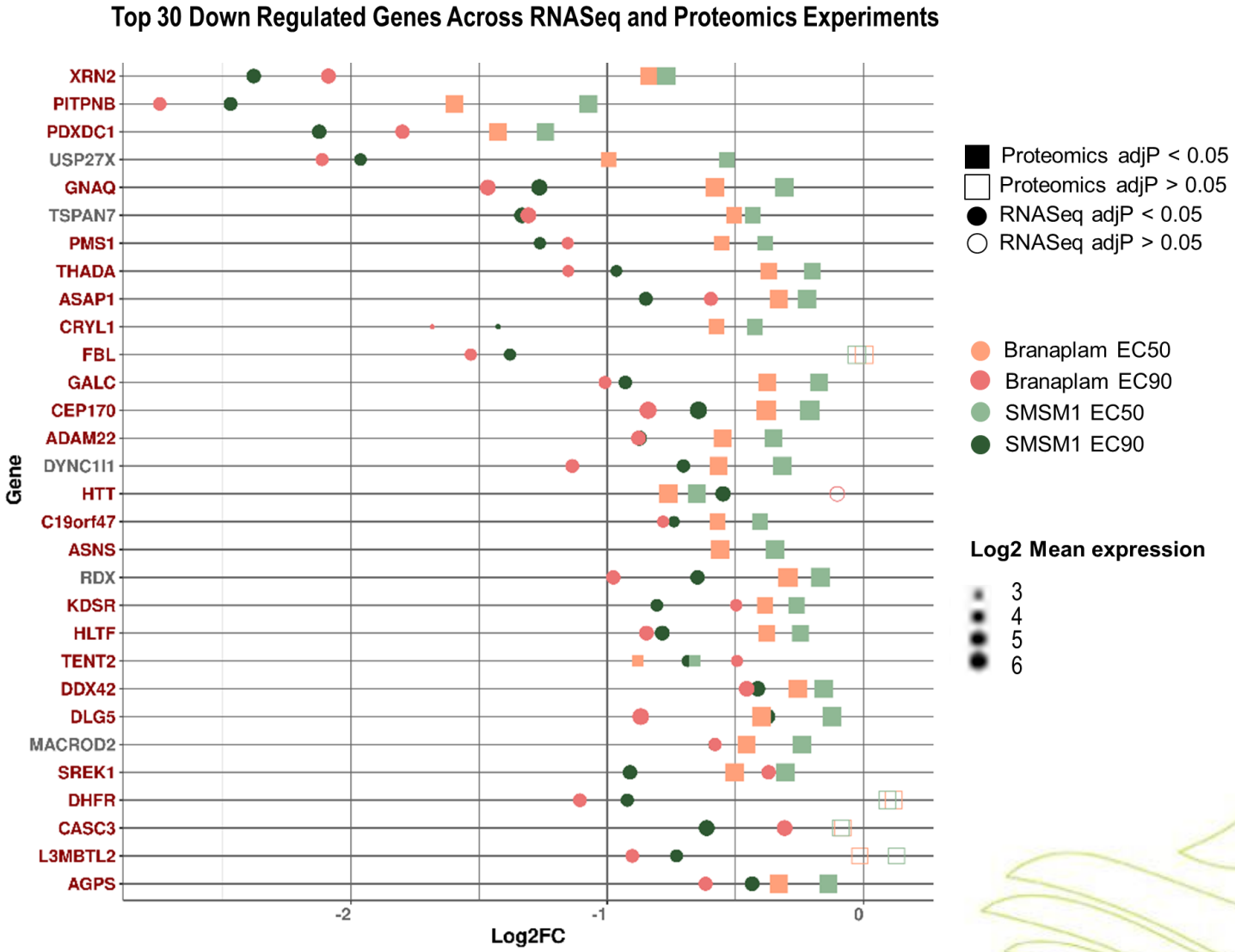
Inclusion of cryptic pseudo-exon in HTT transcript which lowers HTT/Huntingtin levels

RNASeq and Proteomics: which other genes and proteins are affected?

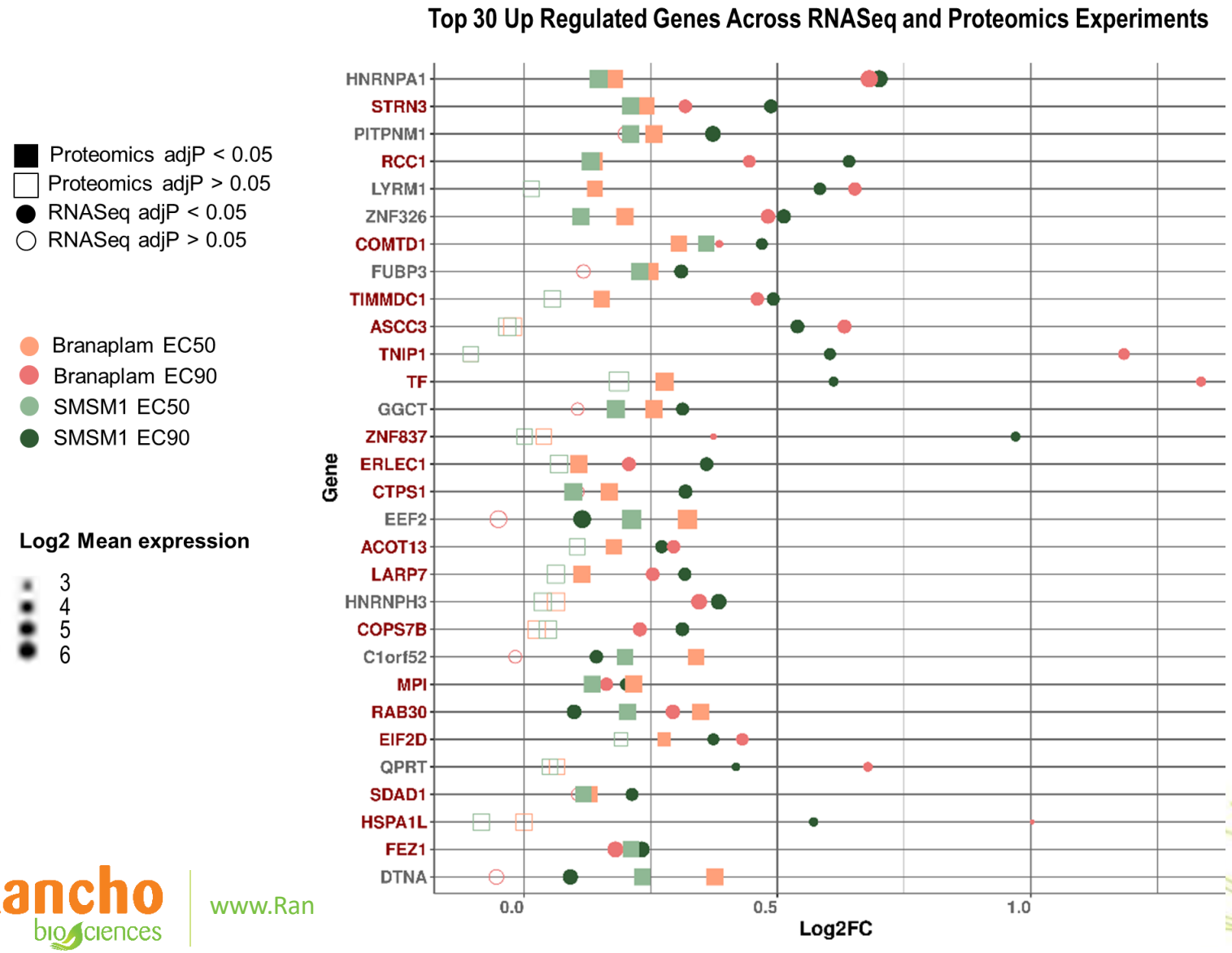
Differential Gene Expression
Differential Protein Expression
Differential Exon and Exon Junction Utilization
Transcript Isoform Switching and Consequences

Meta-analysis

Analysis of alternative splicing events in RNASeq data



Analysis of alternative splicing events in RNASeq data



Rancho BioSciences | www.Ran

Rancho biosciences

HDinHD (Huntington's Disease in High Definition; HDinHD.org

HDinHD overview

New in HDinHD Register Log In

HDinHD: Open Source Science for the HD Research Community

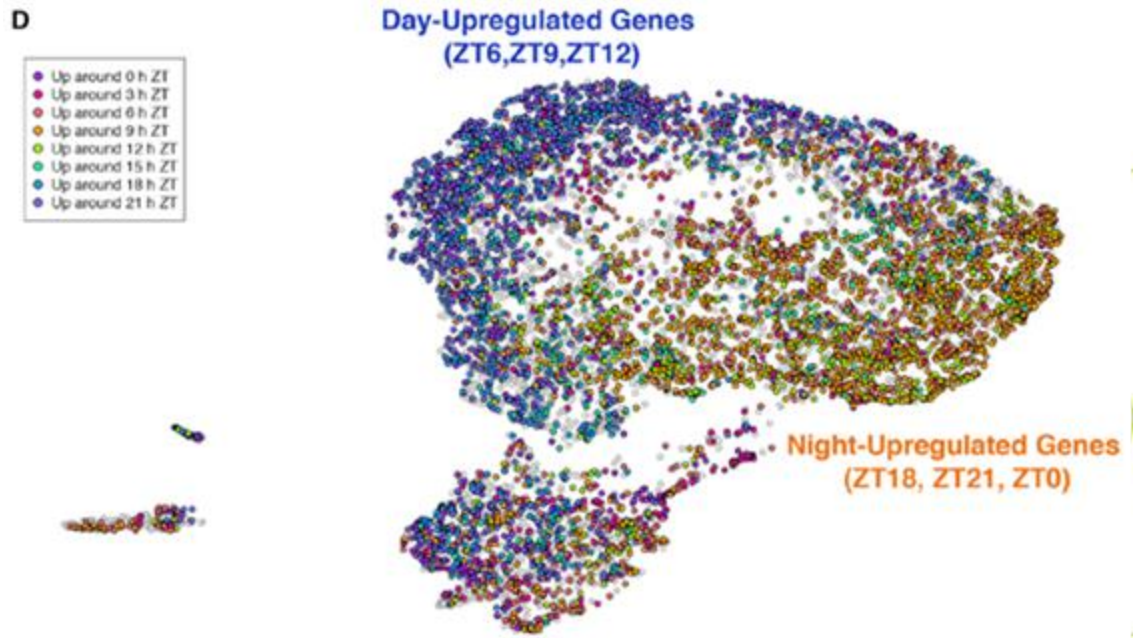
- HDinHD (Huntington's Disease in High Definition; HDinHD.org) is an open online portal developed by CHDI and its partners to serve as a collaborative enabler for the Huntington's Disease (HD) research community.
- HDinHD presents a synthesized view of HD-related scientific data and tools through:
  - 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;
  - c) incorporating community data and tools into the federated portal

HDinHD: federated tools

Tool	Description
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)
REPAIR	Differential gene expression analysis results from >350 HD and HD-related datasets from GEO. (Coppola Lab, UCLA)

HDinHD: federated tools (continued)

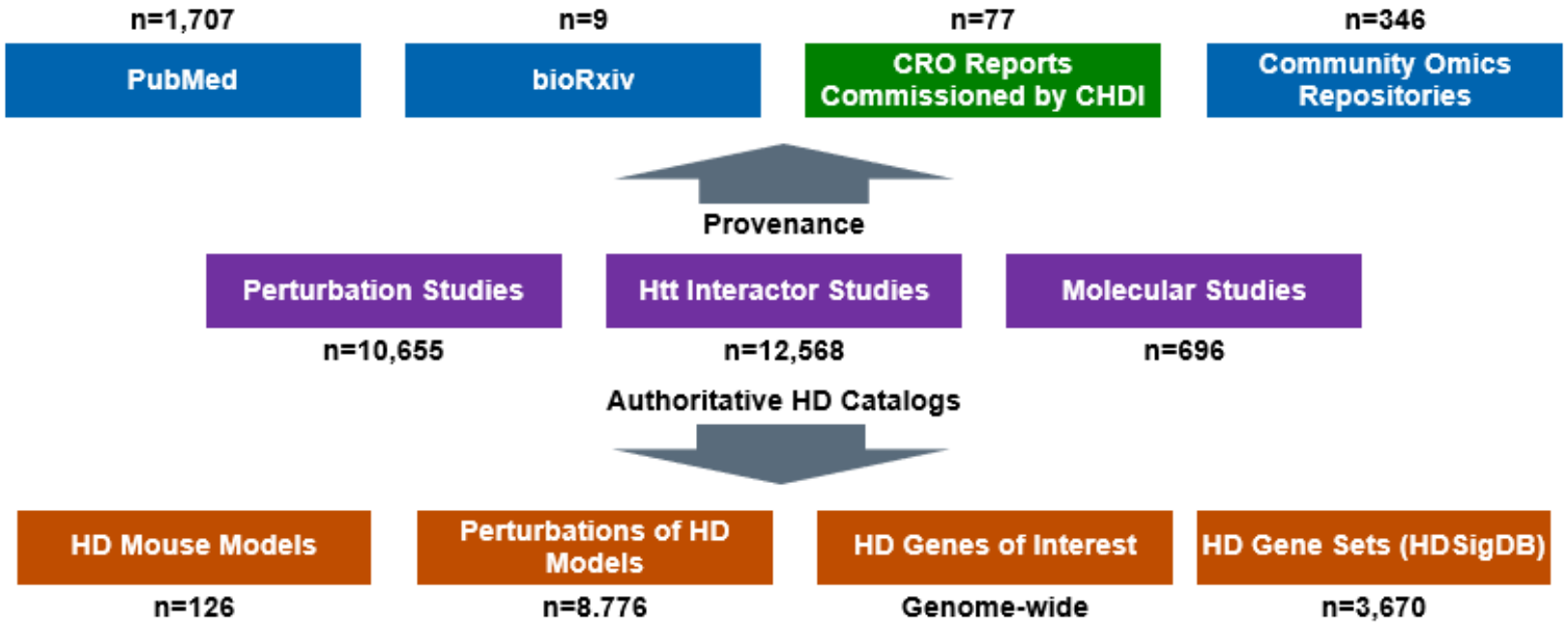
Tool	Description
Enricher	Gene set enrichment analysis tool operating over a large, diverse collection of gene set libraries including HDSigDB, a gene set library containing HD and HD-related gene sets. (Ma'ayan Lab, Mt. Sinai)
CoExMap Viewer	Visualization tool allowing exploration of results and underlying data of a large scale Weighted Gene Co-expression Network Analysis (WGCNA) of hundreds of samples from intact mouse striatum at 6-month 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. (Yang Lab, UCLA)



Wang N et all, Neuron 2022,110(20)

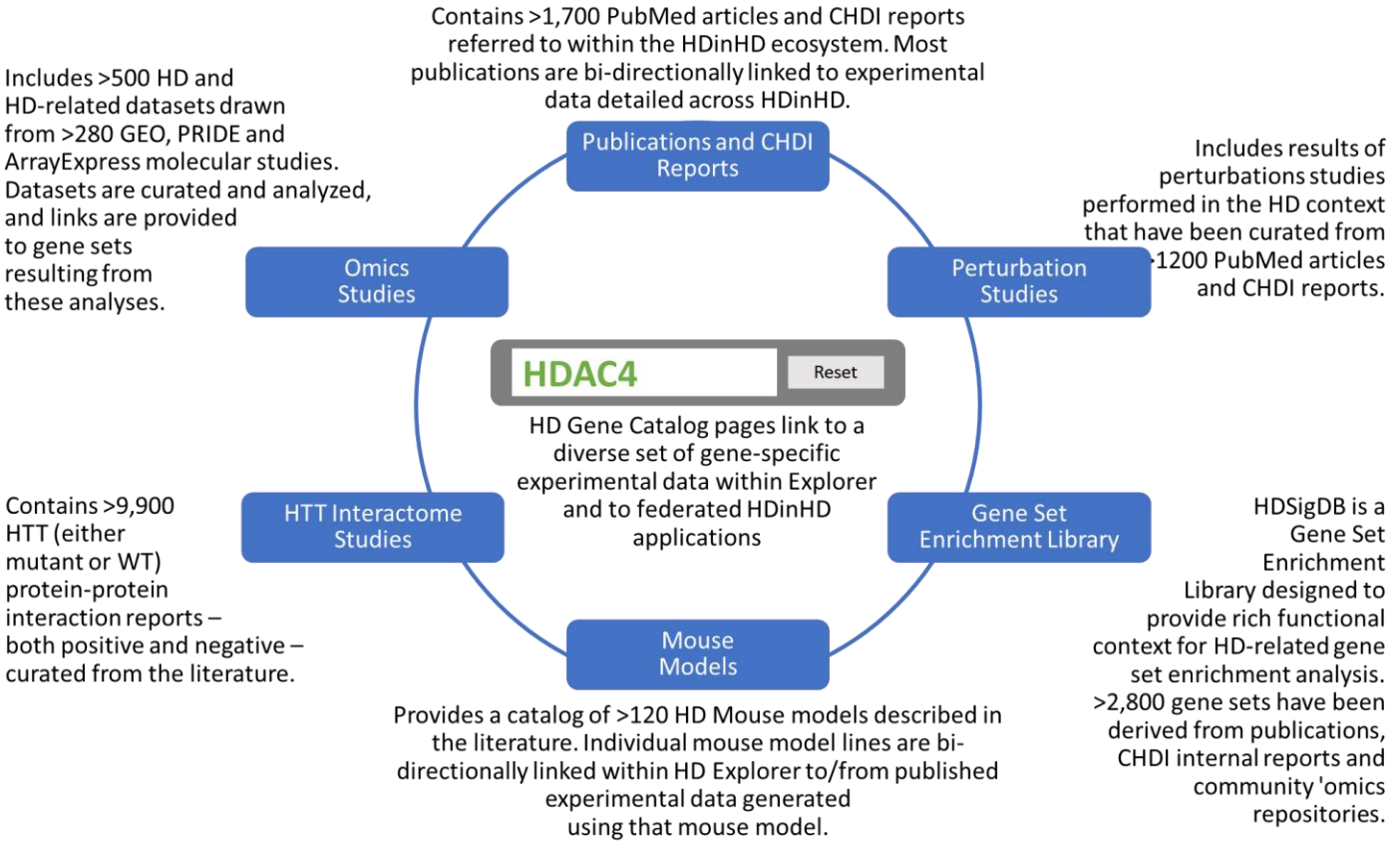
HDinHD: HD Explorer

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



HDinHD: HD Explorer

- 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.



HDinHD downloads

Datasets	Description
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: • **Htt Protein-Protein Interactors** • **HD Gene Set Enrichment Library (HDSigDB)** • **HD Mouse Model Catalog** • **Perturbation Studies** • **HD Omics Studies** • **Publications and Reports**

Acknowledgements

- CHDI
  - Jeff Aaronson
  - Jian Chen
  - Jim Rosinski
  - Vahri Beaumont
  - Deanna Marchionini
  - George Mcallister
  - Jonathan Bard
  - Ignacio Munoz-Sanjuan
  - Tom Vogt

- Rancho Biosciences
  - Christopher Hartl
  - Konstantin Bobkov
  - HDinHD team: Kabenla Armah, Rob Gill, Russell Harris, Suzanna Ralenko, Irina Murasheva and dozens of curators

New answers for Neurodegenerative diseases can be found in public domain data … Rancho can help.

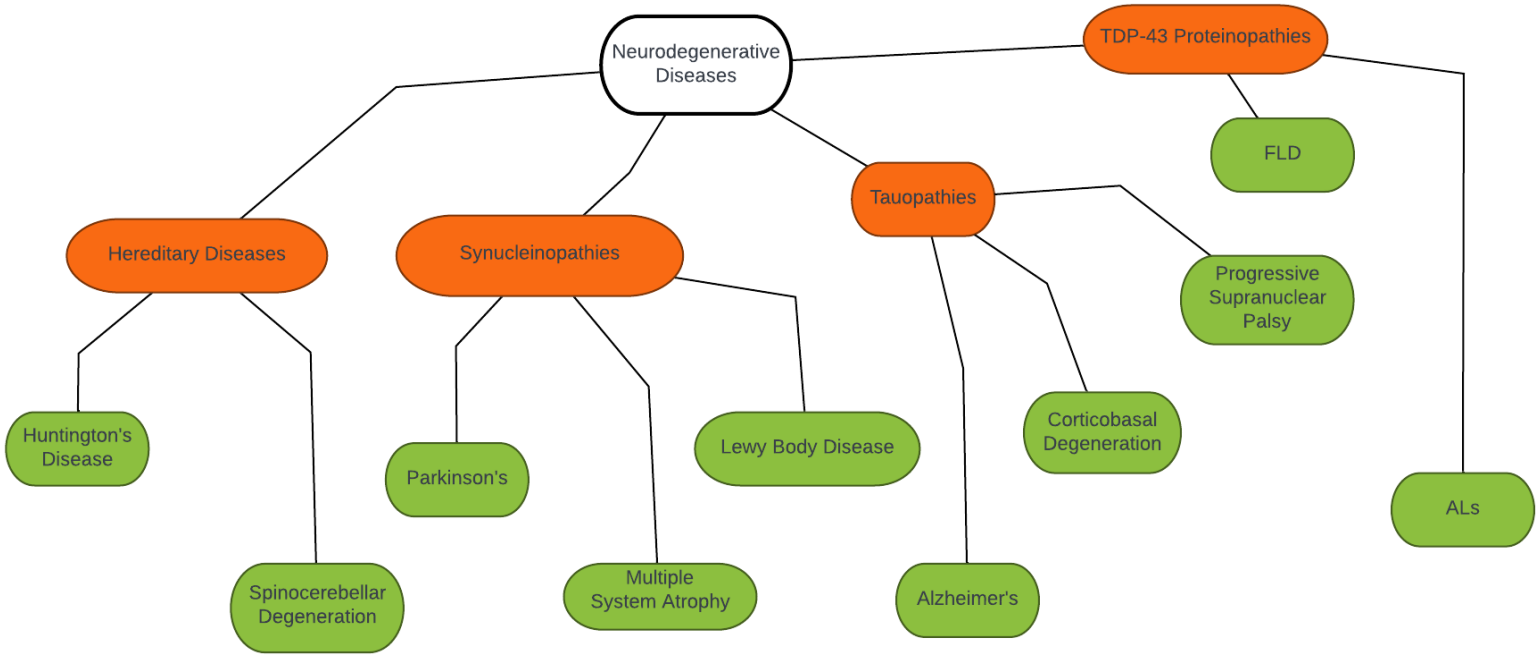
PROS

1. Availability of ever-increasing amounts of datasets of many kinds in the public domain
2. Pharma companies can **increase** the number of experiments available to its scientists **many fold** by combining with their private experiments
3. Many datasets are available at **no cost**!

CONS

1. **No common standards** exist for how many datasets are created, formatted, metadata captured (if at all)
2. It is **difficult and time consuming** for pharma companies to fully exploit these valuable sources of free available research data.
3. **80% effort** is spent on curating/wrangling/ integrating data versus only **20% on analysis**

Data Sets to Consider … 100s!



- **Tauopathies:** Alzheimer's, Corticobasal Degeneration sets: ADNI, AMP-AD, ***160+ genomic studies*** from EBI/GEO, GWAS repository (https://gwas.mrcieu.ac.uk/datasets/)
- **TDP-43 Proteinopathy**
  - FLD : ***13+ genomic studies*** in EBI/GWAS repo; Gene4MND;
  - ALS: ***54 studies*** in EBI/GWAS repo, ALS C.A.R.E; AnswerALS; ALSOD; project MINE
- **Huntington's**: 3 genomic studies in EBI/GWAS repo; mouse models at CHDI
- **Synucleinopathies**
  - Parkinson's: PPMI, BioFIND, FS1, FS-Too, DATATOP, LRKK2, PARS, DeNoPa, S4, Neurogenetik, ***>90 studies with genomic data*** in EBI/GWAS repository

Rancho Biosciences Can Help

Rancho has **both the expertise** in:

- **curating** neurodegenerative datasets for pharma, foundations and government institutions,
- world-class bioinformaticians and data scientists for **data mining & analysis.**

Using Big Data across multiple departments, and companies - spreads & shares the cost

Value and return on investment with a pre-competitive

Neurodegenerative Data Consortium (NDC)

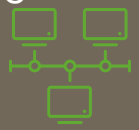
DATA SCIENCE

Aggregate public databases to make unique inferences across large datasets

01

BIOINFORMATICS

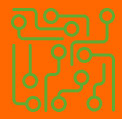
Analyzing large public databases to mine genomic data and develop ML predictive algorithms



02

CURATION

Curate and harmonize datasets. Extract literature information using NLP and curators.



03

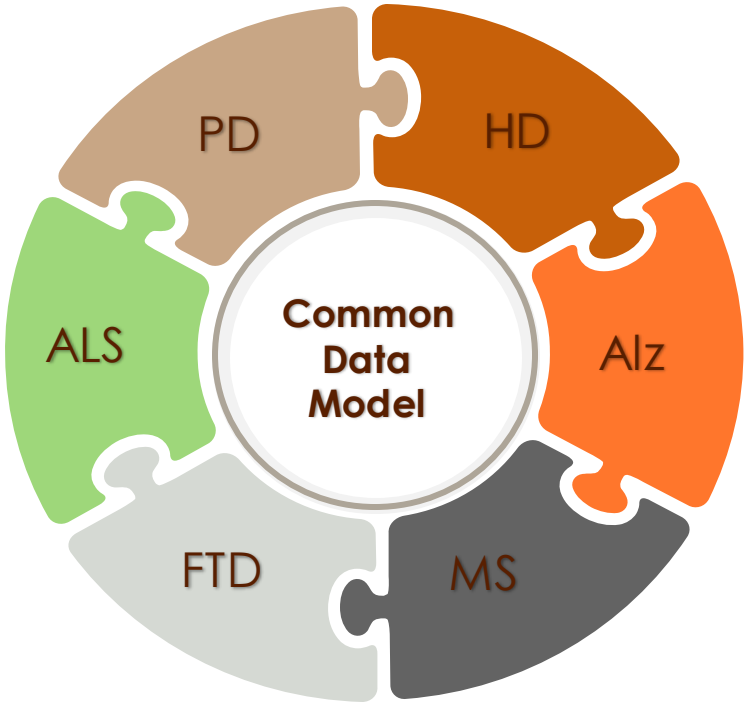
DATA MANAGEMENT

Create interfaces for data ingestion. Integrate public data with internal datasets.

04

NDC Membership Pricing Options

- 3 Year term
- All members join via the Common Data Model $25k per year for charter members
- Members then select disease areas of interest for "focused mini-consortia"



Disease Area	Charter Member fee/year	New Member Fee/year
ALS	$60k	$80k
ALZ	$75k	$100k
FTD	$60k	$80k
HD	$60k	$80k
MS	$75k	$100k
PD	$75k	$100k
Other	$75k	$100k

Charter Member has a signed Membership Agreement in place by time of Kick Off Meeting.

Questions?

Want to know more about how we can support your data needs? Ask us anything!

Contact us through:

Laura Brovold, Ph.D. Business Development
Laura.Brovold@RanchoBioSciences.com
Tel: 760-458-0692

Rancho biosciences

Extra Slides