HIGH PERFORMANCE COMPUTING AND INFORMATICS OFFICE (HPCIO)

Calvin A. Johnson

DCB Operations Review

July 15, 2011

Division of Computational Bioscience

Center for Information Technology | National Institutes of Health

HPCIO DOMAIN AREAS

- Portfolio Analysis
- High Performance Computing
- Genomics
- Proteomics
- Imaging
- Biomedical Informatics
- Cheminformatics
- “Special Projects” emphasis,
  - + e.g., vocabulary research activity

HPCIO CAPABILITIES AND COMPETENCIES

- Numerical and statistical methodology
- Machine learning
- Natural language processing
- Semantics and linguistics
- Information retrieval
- Visualization
- Knowledge management
- Agile, rapid-response team

HPCIO STAFF

- Federal Employees
  - Huey Cheung
  - Anthony Fletcher
  - Michelle Ji
  - Calvin Johnson
  - William Lau
  - Daniel Russ
  - Giun Sun
  - Alex Wang
  - Kai Wang
- SRA and subcontractors
  - Shuxing Cheng
  - Sarada Chintala
  - Krishna Collie
  - Tin Doan
  - Adam Frazin
  - Stephen Glanowski
  - Shahar Goldin
  - Beecher Greenman
  - Arun Ravindran
  - Mark Roth
  - Jigar Shah
  - Guoli Wang

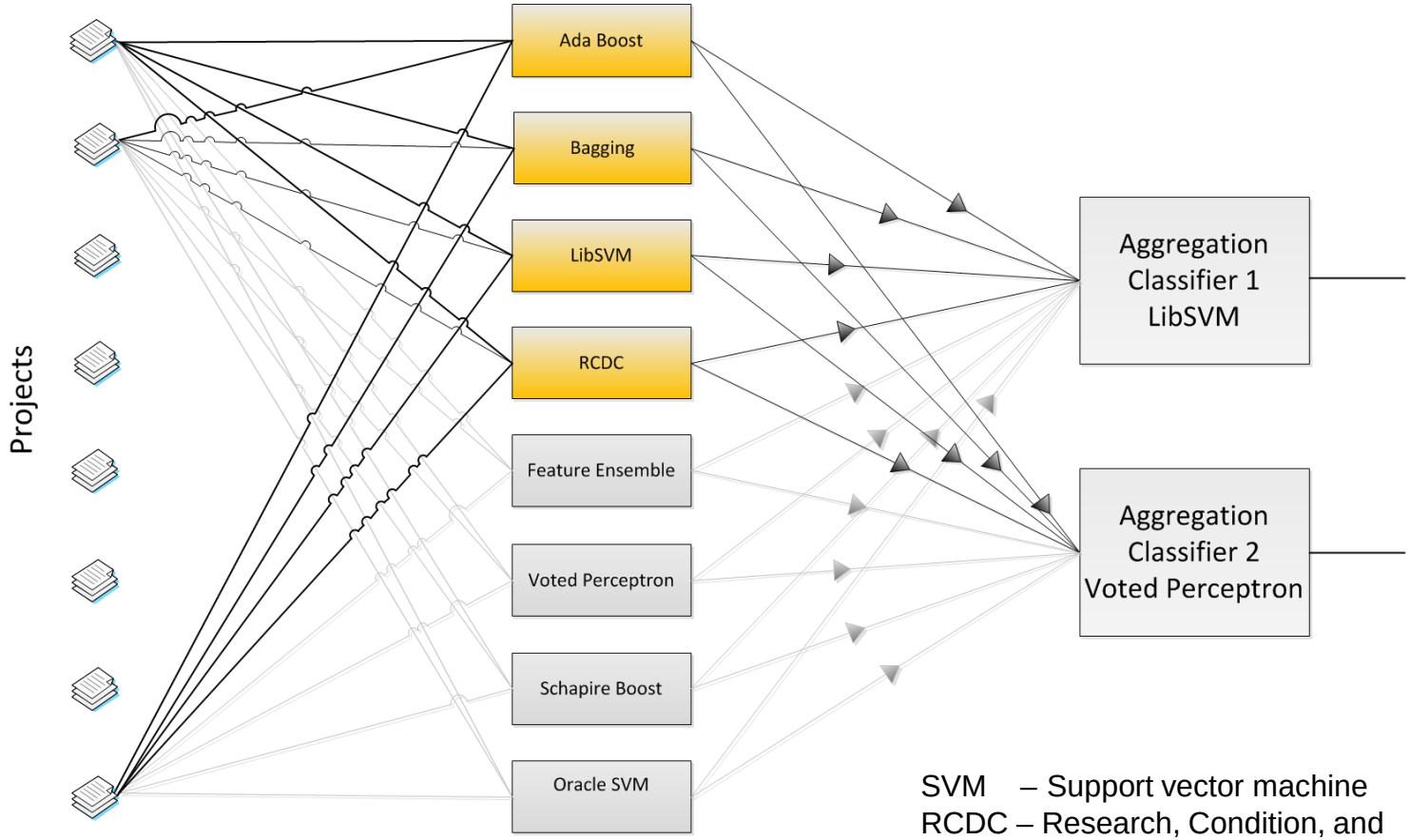
PORTFOLIO ANALYSIS ACTIVITY COLLABORATOR LIST (1 OF 2)

IC	Collaborators	Projects/Goals
NATIONAL CANCER INSTITUTE	Lisa Krueger Michele Vos Maria Bukowski (KMSPB)	• Analyzed Behavioral Research portfolio • Clustering on cancer prevention • Map ClinicalTrials.gov protocols to IMPACII grants. • Enhanced classification for IRP, including auxiliary classifiers to provide evidence
DPCPSI/OD NATIONAL INSTITUTES OF HEALTH	Chuck Lynch Lora Kutkat Geetha Senthil Carole Christian Faye Austin	• Development of Comparative Effectiveness Research (CER) classifier and retrospective identification of CER projects. • High-Risk, High-Reward demonstration project
National Institute of Allergy and Infectious Diseases	Marie Parker Dolan Ghosh-Das	• Visualize Anti-Microbial Resistance portfolio • Mapping HIV/AIDS priorities, objectives, and initiatives
NIMH National Institute of Mental Health	David Armstrong	• High-Risk, High-Reward demonstration project

PORTFOLIO ANALYSIS ACTIVITY COLLABORATOR LIST (2 OF 2)

IC	Collaborators	Projects/Goals
NIGMS	Paul Sheehy Lisa Dunbar Elena Makareeva (NICHD) Peter Lyster Jim Deatherage	• Facilitate a comprehensive understanding of the NIGMS portfolio including the development of basic research categories. • Mine literature to discover emerging trends in cell biology and biophysics. • Develop classifiers for BICB.
OER/OD	Rick Ikeda Pete Morton Judy Riggie Israel Lederhendler Patty Gaines	• Portfolio visualization and clustering platform. • Decision support for thesaurus enrichment (e.g., cancer genomics). • Metric-informed enhancement of visualization. • Prediction of study section assignment.
NHLBI	Carl Roth Melissa Antman Susan Scolnik Zophia Gajdos	• High-Risk, High-Reward (HRHR) demonstration project.

ENSEMBLE CLASSIFIER

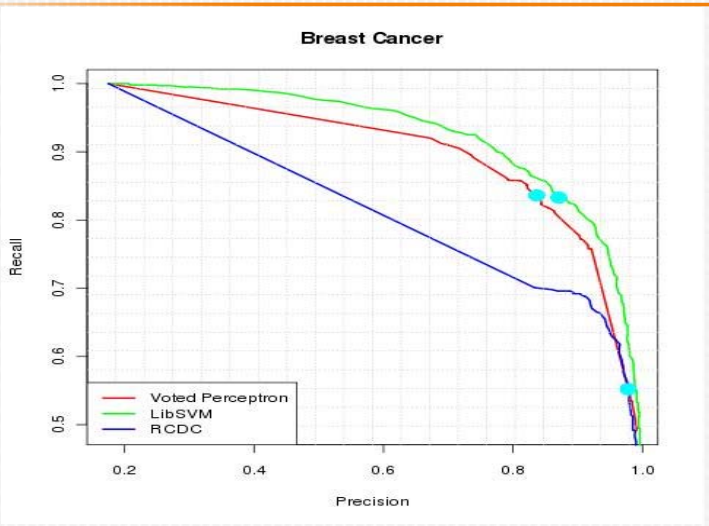


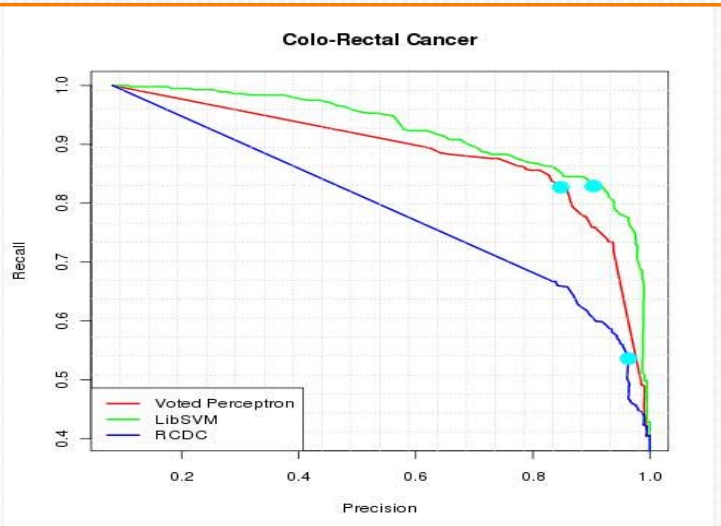
SVM – Support vector machine
RCDC – Research, Condition, and Disease Categorization

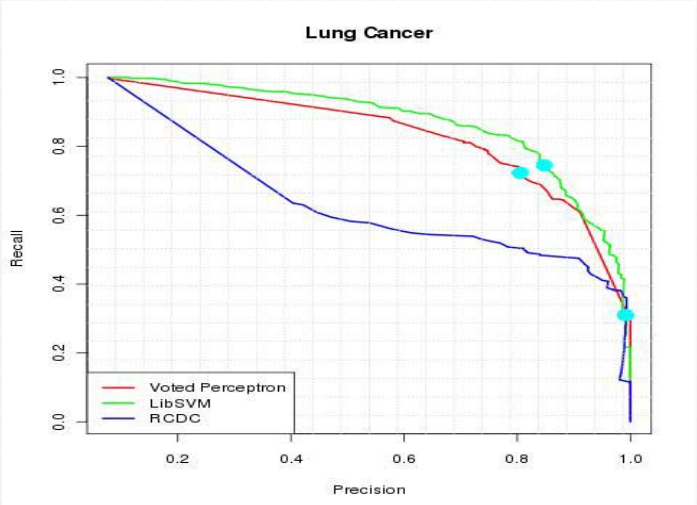
PERFORMANCE ON CANCER CATEGORIES

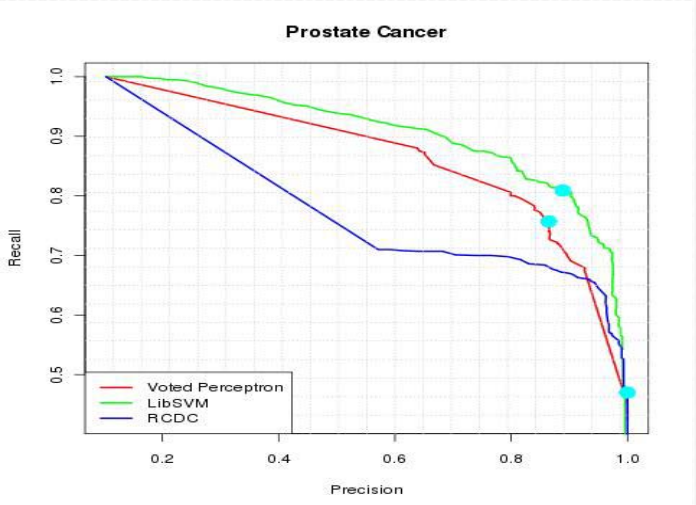
Disease Category	2006 RCDC			LibSVM Ensemble		
	Recall	Precision	F-score	Recall	Precision	F-score
Breast Cancer	**0.55**	**0.98**	**0.71**	**0.83**	**0.87**	**0.85**
Cervical Cancer	**0.35**	**0.93**	**0.51**	**0.75**	**0.89**	**0.81**
Childhood Leukemia	**0.40**	**0.22**	**0.28**	**0.67**	**0.81**	**0.73**
Colo-Rectal Cancer	**0.54**	**0.96**	**0.69**	**0.83**	**0.90**	**0.87**
Hodgkins Disease	**0.18**	**0.64**	**0.28**	**0.79**	**0.77**	**0.78**
Liver Cancer	**0.35**	**0.92**	**0.50**	**0.66**	**0.87**	**0.75**
Lung Cancer	**0.31**	**0.99**	**0.47**	**0.75**	**0.85**	**0.79**
Lymphoma	**0.46**	**0.93**	**0.62**	**0.64**	**0.82**	**0.72**
Ovarian Cancer	**0.37**	**0.99**	**0.54**	**0.66**	**0.88**	**0.75**
Prostate Cancer	**0.47**	**1.00**	**0.64**	**0.81**	**0.89**	**0.85**
Uterine Cancer	**0.21**	**0.93**	**0.34**	**0.54**	**0.86**	**0.67**
Weighted average, cancer	**0.45**	**0.94**	**0.60**	**0.77**	**0.87**	**0.81**
Weighted average, all categories	**0.58**	**0.91**	**0.69**	**0.77**	**0.86**	**0.81**

RECALL-PRECISION CURVES (2006 DATA)

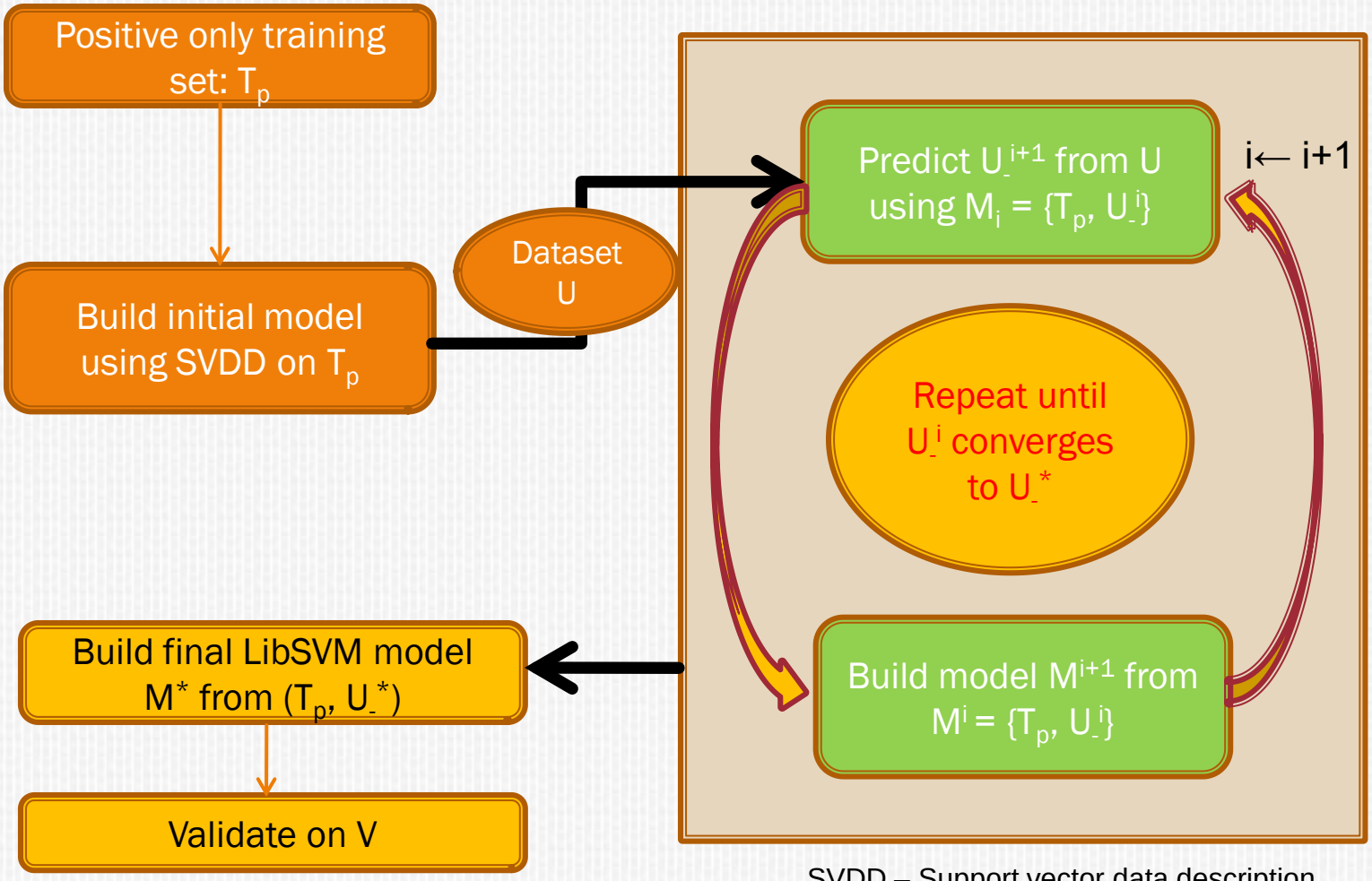








ONE-SIDED SVM-BASED CLASSIFIER



SVDD – Support vector data description

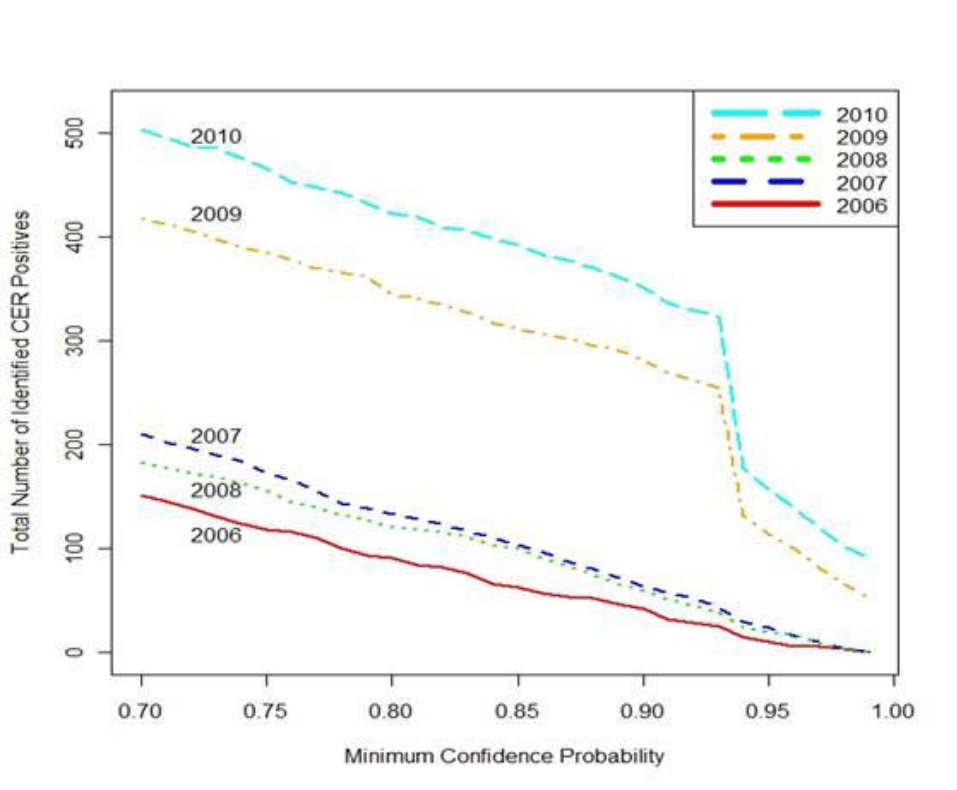
ONE-SIDED CLASSIFICATION ON STUDY SECTIONS

SS	Train #P	Test #P	Test #N	Recall	Precision	F-score
Train and test on 2006-2010 awarded only[1]						
BMIT	663	443	5000	0.930	0.800	**0.860**
GCAT	437	295	5000	0.885	0.808	**0.845**
Train on 2006-2009, test on 2010 awarded and un-awarded[2]						
BMIT	2029	571	1983	0.723	0.797	**0.758**
GCAT	1284	391	2163	0.813	0.795	**0.804**

1. Test negatives were drawn from awarded grants in all study sections.
2. Test negatives were awarded and un-awarded grants in 2010 from all other Biomedical Informatics and Computational Biology study sections.

BMIT: Biomedical Imaging Technology
GCAT: Genomics, Computational Biology and Technology

COMPARATIVE EFFECTIVENESS RESEARCH



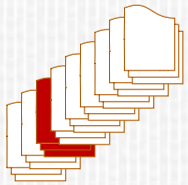
Parameter	Value
# positives in training	229
True positives	80
False positives *	8
False negatives	22
True negatives *	79
Recall	0.784
Precision *	0.909
F-score *	0.842

* - Assuming suspected negatives are actual negatives (RCDC Non-Defensible's and Unambiguous Error's)

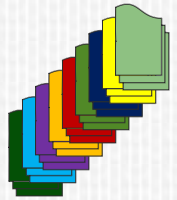
Matching IMPACII to ClinicalTrials.gov

ClinicalTrials.gov
A service of the U.S. National Institutes of Health

IMPAC II
National Institutes of Health
> Information for Management, Planning, Analysis, and Coordination



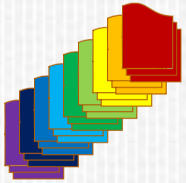
Search IMPACII for Project



Candidates Found

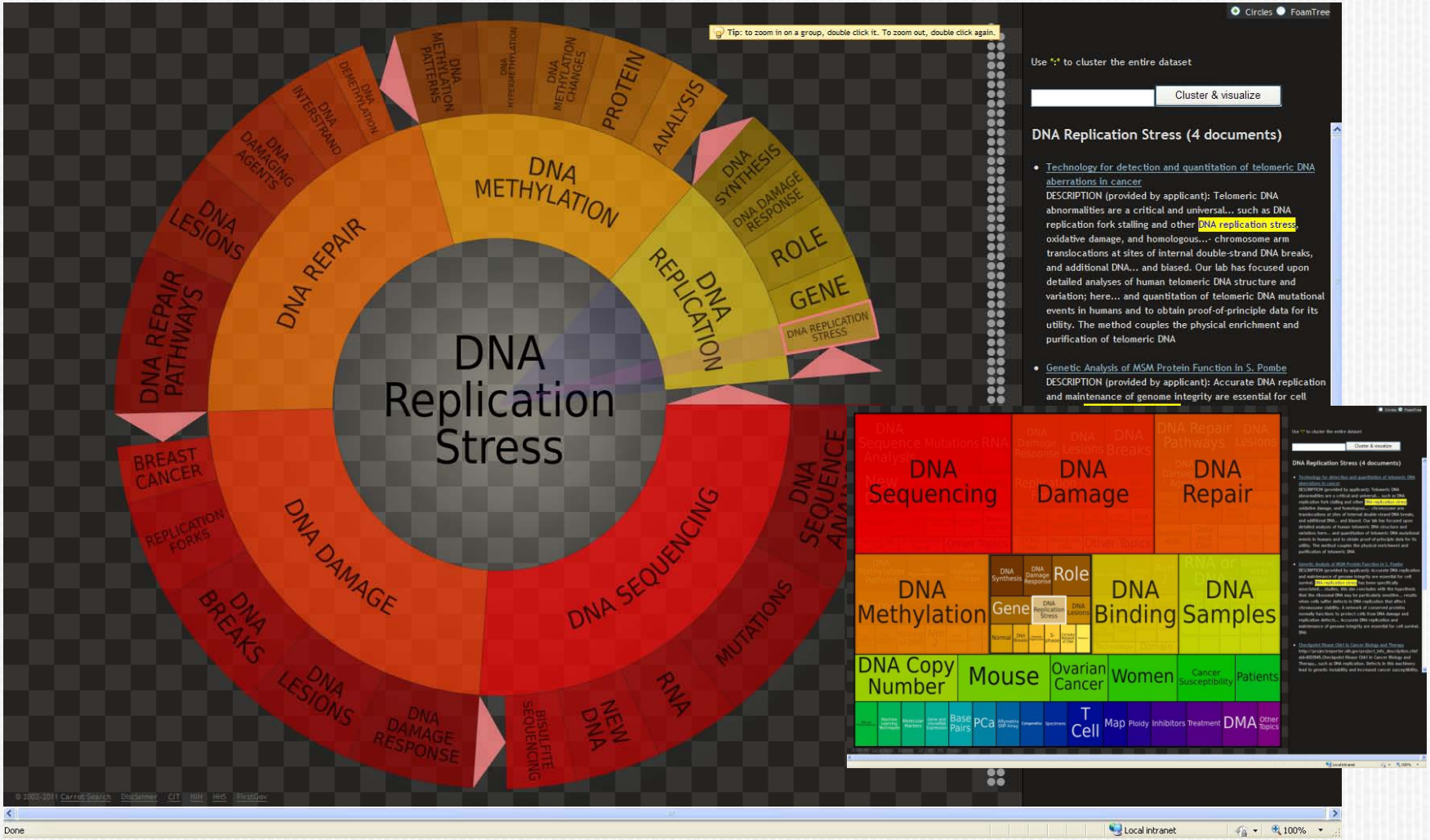
Rank Results

- Project title edit distance
- Abstract/Specific Aims versus Protocol cosine similarity from fingerprint
- Edit distance on Location
- Edit distance on Organization
- Adjust score based on Phase III ClinicalTrials flag in IMPAC II

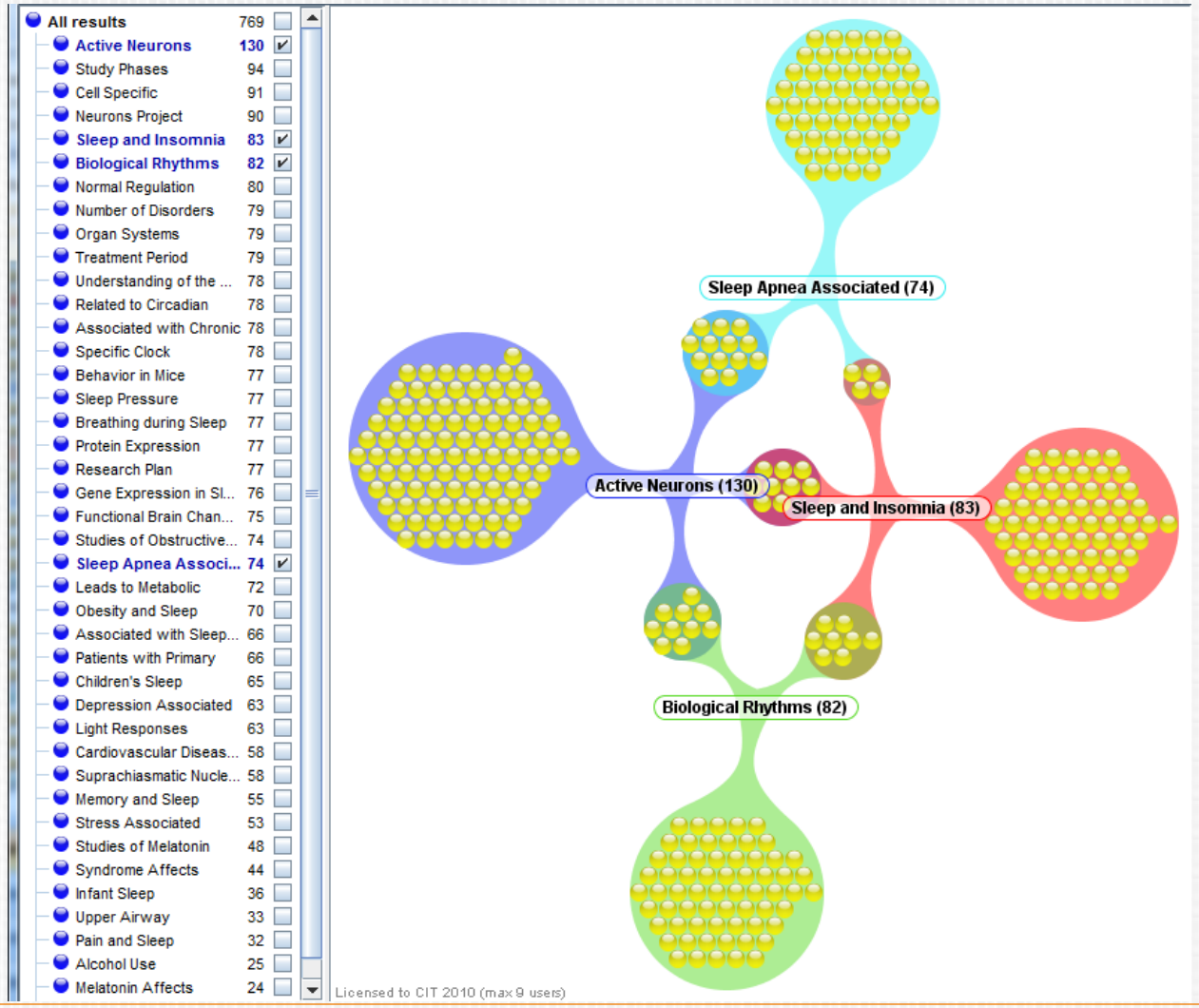


Ranked IMPACII Projects returned.

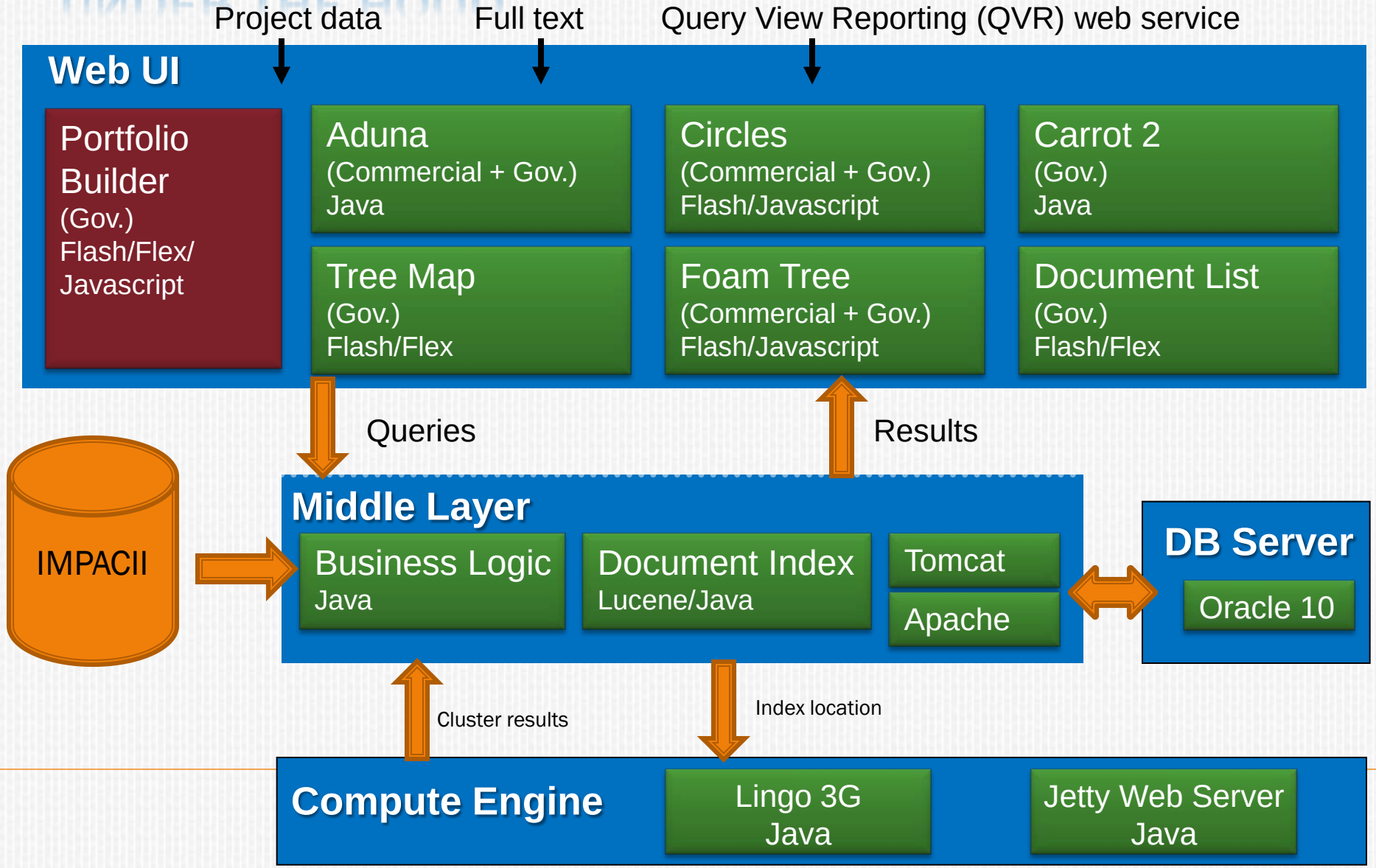
CARROT2 WITH LINGO 3G CLUSTERING



ADUNA VISUALIZATION



PVIZ (PORTFOLIO VISUALIZATION)
UNDER THE HOOD



HIGH-RISK HIGH-REWARD (HRHR) PROJECT

- Division of Computational Bioscience/CIT
  - Daniel Russ, Stephen Glanowski, Calvin Johnson
- Office of the Director/NHLBI:
  - Carl Roth (Acting Deputy Director, NHLBI), Melissa Antman, Susan Scolnik, Zophia Gajdos
- Division of Program Coordination, Planning, and Strategic Initiatives/OD
  - Faye Austin (contractor)
- Division of Extramural Activities/NIMH
  - David Armstrong (Chief, Extramural Review Branch)

CIT HRHR TEXT MINING TOOL

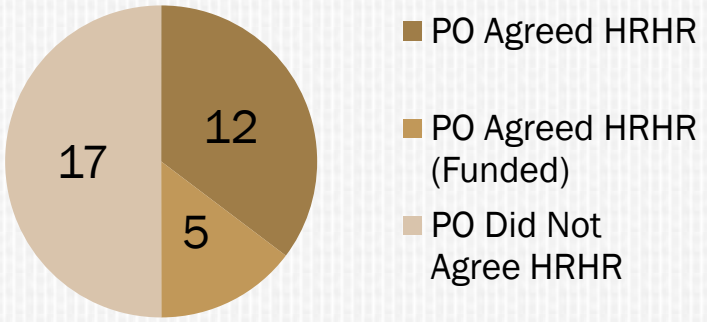
- NIH Reform Act of 2006
  - Required HRHR Demonstration Project
  - CIT Text Mining Tool specifically mentioned in 2009 Report to Congress as an HRHR Demonstration Project
- Natural language processing to find words in context
- Can identify HRHR research.
  - Recall ~80-90%, Precision ~30-60%
- Can show with statistical significance that NIH funds HRHR research.
- Can show effectiveness of specific funding mechanism for attracting HRHR research.

NHLBI 10/2007 and 1/2008 Council Rounds

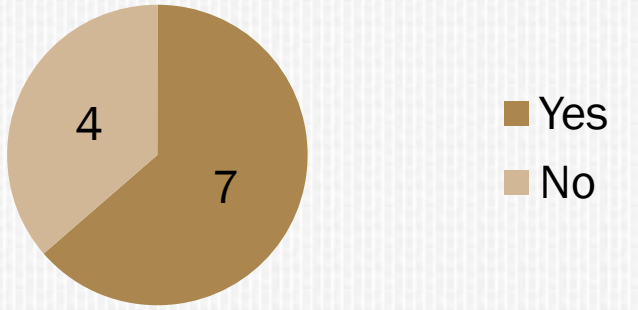
HRHR+			HRHR-		
Funded	Unfunded		Funded	Unfunded	
	Scored	Unscored		Scored	Unscored
18.8%	42.0%	39.2%	12.7%	36.2%	51.1%

PROGRAM OFFICIAL (PO) FEEDBACK: UNSOLICITED R01 OCT. 2010 NHBLI

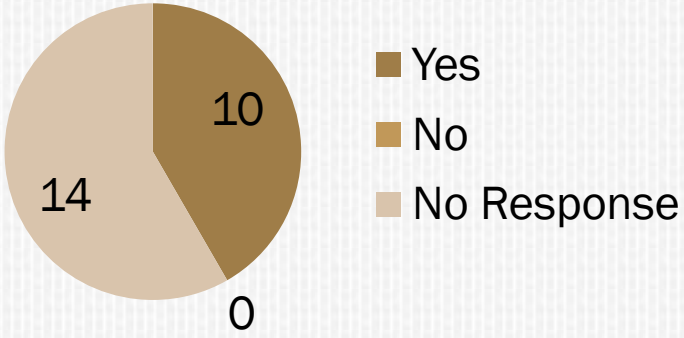
HRHR+ Applications
(34 HRHR+ Apps)



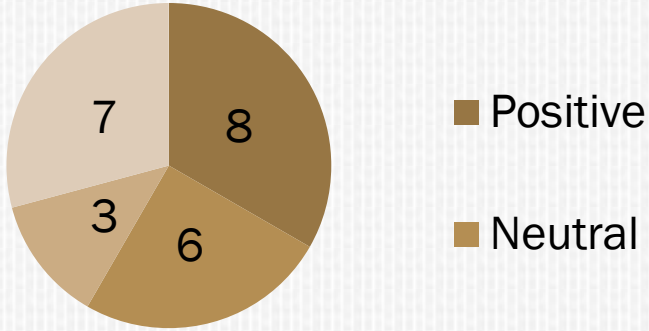
Consider for Select Pay
(11 HRHR+ Apps Discussed, Not Funded, not amended and resubmitted)



HRHR Status as Tiebreaker
(24 POs)

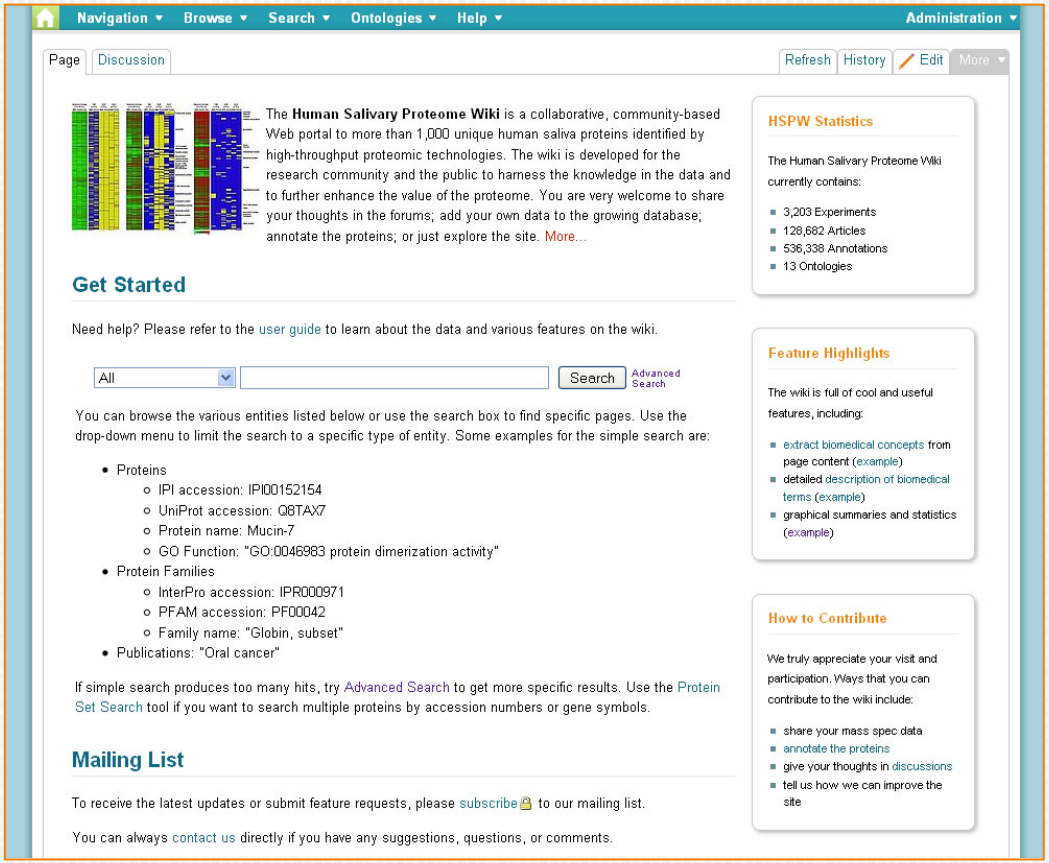


Feelings on Usefulness
(24 POs)



HUMAN SALIVARY PROTEOME WIKI

- HPCIO staff: William Lau [E], Giun Sun [E], Beecher Greenman [C]
- Collaborators:
  - Dr. Lillian Shum, Dr. Penny Burgoon, Dr. Jason Wan, Integrative Biology and Infectious Disease Branch, Division of Extramural Research, NIDCR
  - John Prue, NIDCR CIO



http://salivaryproteome.cit.nih.gov/training - Training site w/new design

GOALS

- Make the salivary proteome catalog, including its associated experimental data, fully and easily accessible to researchers;.

PROTEIN CATALOG

Pages in category "Salivary Proteins"

Sort by: Peptide Hits | Records per page: 20 | Tissue: Parotid gland

Export to Spreadsheet

1 to **20** of **1839** Next › Last »

Protein	Gene	Protein Name	Peptide Hits	Data Source
HSPW:PD0564E	AMY1A AMY1B AMY1C	Alpha-amylase 1	499	Swiss-Prot
HSPW:PD06661	AMY2B	Alpha-amylase 2B	389	Swiss-Prot
HSPW:PD07CCA	AMY2A	Pancreatic alpha-amylase	351	Swiss-Prot
HSPW:PE918CD	PRB1	PRB1 protein	214	TrEMBL
HSPW:PE9B0A3	PRB1	Basic salivary proline-rich protein 1	205	Swiss-Prot
HSPW:PE8C8CE	PIGR	Polymeric immunoglobulin receptor	194	Swiss-Prot
HSPW:PEB4900		cDNA FLJ14473 fis, clone MAMMA1001080, highly similar to Homo sapiens SNC73 protein (SNC73) mRNA	179	TrEMBL
HSPW:PEDBDCE	PRH1 PRH2	PROLINE-RICH PROTEIN HAEIII SUBFAMILY 1	173	IPI
HSPW:PE90567	PRH1 PRH2	Salivary acidic proline-rich phosphoprotein 1/2	172	Swiss-Prot
HSPW:PDEC681	IGHA1	Ig alpha-1 chain C region	169	Swiss-Prot
HSPW:PE9934C	DKFZp686L19235	Putative uncharacterized protein DKFZp686L19235	164	TrEMBL

Jump to: Name and Origin | Sequence Attributes | Comments | Features | Proteomics | Cross References | Keywords | References | Entry Information

Alpha-amylase 2B [Homo sapiens]

Names and Origin

Protein names	*Official name:* Alpha-amylase 2B *Alternative name(s):* 1,4-alpha-D-glucan glucanohydrolase 2B, Carcinoid alpha-amylase
Genes	AMY2B [EntrezGene:280]
Organism	Homo sapiens
Taxonomy	Eukaryota > Metazoa > Chordata > Craniata > Gnathostomata > Mammalia > Euarchontoglires > Primates > Haplorrhini > Simiiformes > Catarrhini > Hominoidea > Hominidae > Homininae > Homo

Sequence Attributes

Identifier	Name	Aliases	Sequence length	Molecular mass	Sequence
P19961-1	Canonical sequence		511	57,709.5	HSPW:PD06661/1

Comments

Sort by: Feature key Records per page: 20

Export to Spreadsheet

(All 15 results shown)

Feature key	Description	Evidence code	Evidence reference	Variant	Reported by
Biological process	GO:0007586 digestion	TAS			UniProt
Biological process	GO:0005975 carbohydrate metabolic process	NAS			UniProt
	Endohydrolysis of (1->4) alpha-D-glucosidic linkages in				

Available Gadgets

- Cluster Membership
- BLAST Search
- Sequence Signatures
- Protein Interactions
- Proteomics Identifications

PROTEOMICS DATA

Experiments

Sort by: Accession Records per page: 20

Export to Spreadsheet

1 to 20 of 64 Next › Last »

Accession	Title	Species	Tissue	Disease	Protein Count	Peptide Count
1	Experiment HSPP_SF_20, subexperiment 0, scan 0 (2011-02-17 10:43:47)	Homo sapiens (Human)	Submandibular gland, Sublingual gland		185	560
6	Experiment HSPP_SF_20, subexperiment 5, scan 0 (2011-02-17 10:43:47)	Homo sapiens (Human)	Submandibular gland, Sublingual gland		123	442
16	Experiment HSPP_SF_28, subexperiment 2, scan 0 (2011-02-17 10:43:47)	Homo sapiens (Human)	Submandibular gland, Sublingual gland		322	1222
28	Experiment HSPP_SF_48, subexperiment 2, scan 0 (2011-02-17 10:43:47)	Homo sapiens (Human)	Submandibular gland, Sublingual gland		83	164
38	Experiment HSPP_SF_49, subexperiment 2, scan 0 (2011-02-17 10:43:47)	Homo sapiens (Human)	Submand Sublingua			
42	Experiment HSPP_SF_49, subexperiment 6, scan 0 (2011-02-17 10:43:47)	Homo sapiens (Human)	Submand Sublingua			
99	Experiment HSPP_DW_2151, subexperiment 0, scan 0 (2011-02-17 10:46:08)	Homo sapiens (Human)	Sublingua			
289	Experiment HSPP_DW_2011, subexperiment 0, scan 0 (2011-02-17 10:46:09)	Homo sapiens (Human)	Submand			

Peptides

```
  1 MKFFLLLFTI GFCWAQYSPN TQQGRTSIVH LFEWRWVDIA LECERYLAPK
 51 GFGGVQVSPP NENVAIYNPF RPWWERYQPV SYKLCTRSGN EDEFRNMVTR
101 CNNVGVRIYV DAVINHMCGN AVSAGTSSTC GSYFNPGSRD FPAVPYSGWD
151 FNDGKCKTGS GDIENYNDAT QVRDCRLTGL LDLALEKDYV RSKIAEYMNH
201 LIDIGVAGFR LDASKHMWPG DIKAILDKLH NLNSNWFPAG SKPFIYQEVI
251 DLGGEPIKSS DYFGNGRVTE FKYGAKLGTV IRKWNGEKMS YLKNWGEGWG
301 FVPSDRALVF VDNHDNQRGH GAGGASILTF WDARLYKMAV GFMLAHPYGF
351 TRVMSSYRWP RQFQNGNDVN DWVGPPNNNG VIKEVTINPD TTCGNDWVCE
401 HRWRQIRNMV IFRNVVDGQP FTNWYDNGSN QVAFGRGNRG FIVFNNDDWS
451 FSLTLQTGLP AGTYCDVISG DKINGNCTGI KIYVSDDGKA HFSISNSAED
501 PFIAIHAESK L
```

Highlight peptides from Submandibular gland in SPW:C000186

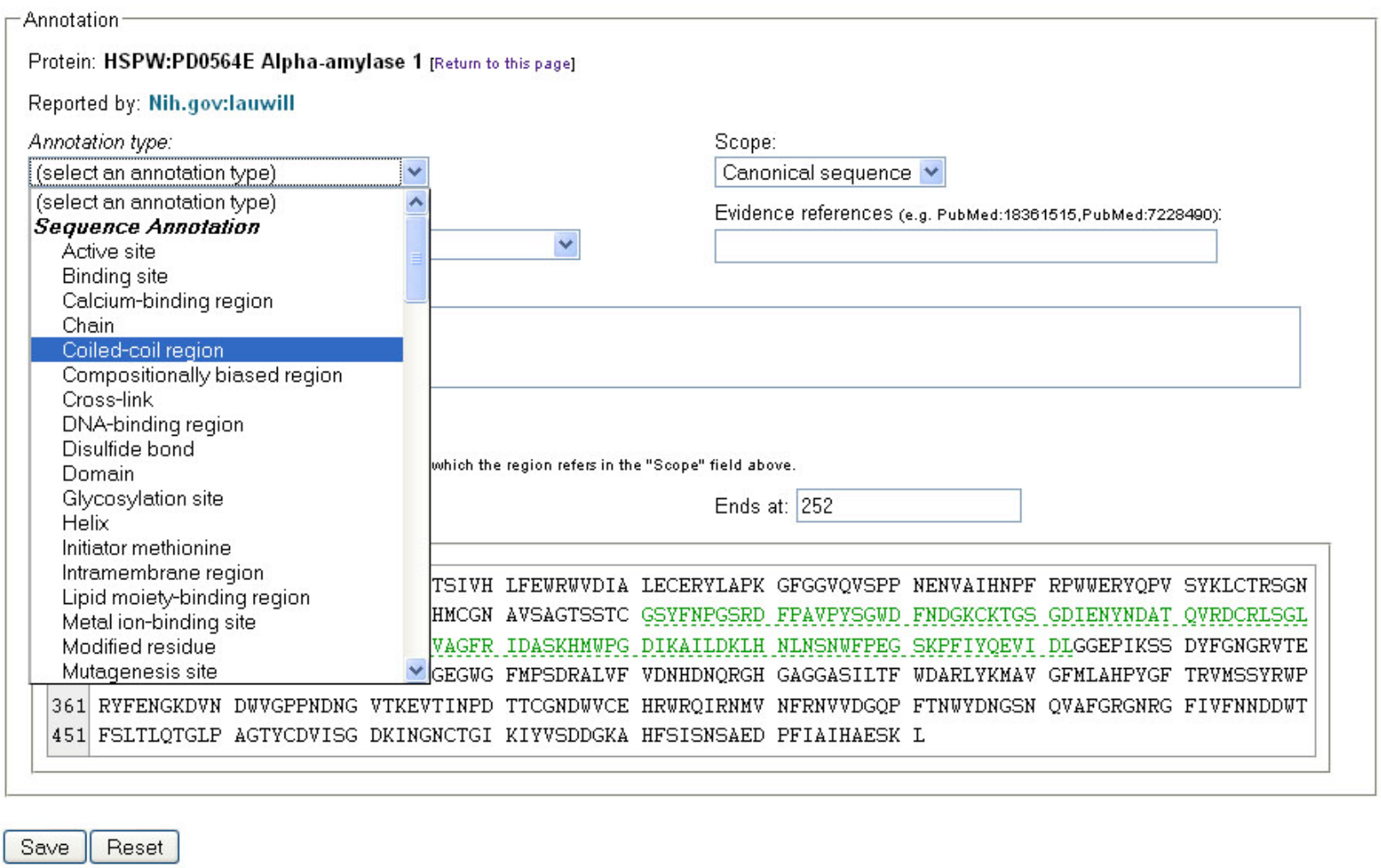
Sequence Coverage = 84.34%

PRIDE – Proteomics IDEntification database (ProteomeExchange consortium)

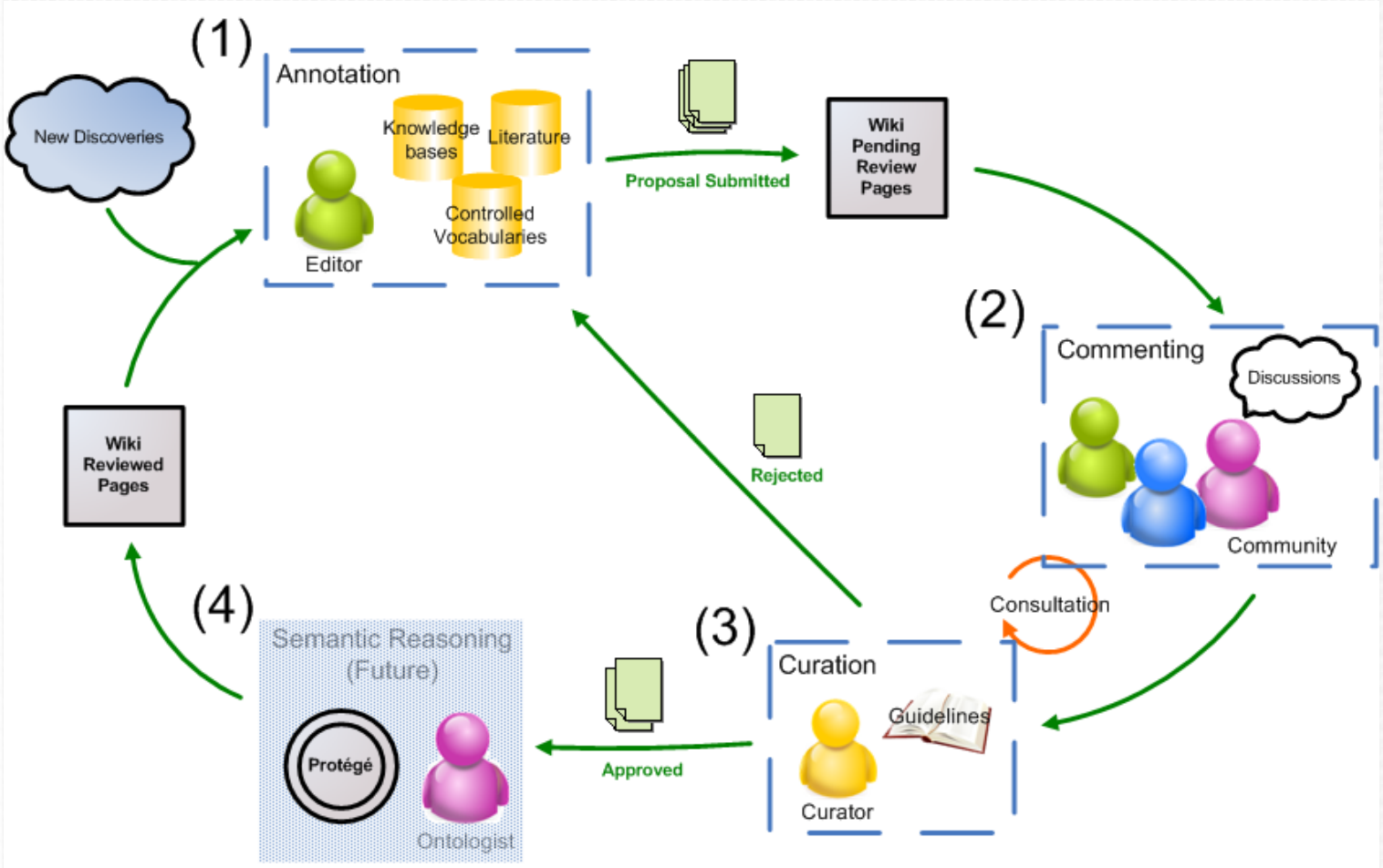
GOALS

- Make the salivary proteome catalog, including its associated experimental data, fully and easily accessible to researchers;
- Encourage community-driven refinement of the catalog through the deposition of new data and annotation on existing content;

ANNOTATION INTERFACE



CURATION WORKFLOW



ANNOTATIONS

Comments

Sort by: Feature key Records per page: 20

Export to Spreadsheet

(All 7 results shown)

Feature key	Description	Evidence code	Evidence reference	Variant	Reported by
Cellular component	GO:0005615 extracellular space	TAS			UniProt
Function	PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.	NR			UniProt
Molecular function	GO:0005515 protein binding	IPI			UniProt
Polymorphism	Sequence shown is that of allele PRH2-2, also known as PR-2; Allele PRH2-1 is also known as PR-1 or protein C, and allele PRH2-3 as PR-1'. The PRH1-DB allele (about 16% of the population) has an insertion of 21 repeated amino-acids compared to the more frequent PRH1-PIF allele (68%). In contrast to all other PRH1 and PRH2 alleles, the PRH1-PA allele (16%) is not proteolytically cleaved.	NR			UniProt
PTM	An hexuronic acid was shown to be linked to Ser-33 in about 40% of the polypeptides. Neither the structure of the carbohydrate (whether glucuronic acid or an isomer of), nor the linkage (whether a glycoside or an ester) has been definitely established.	NR			UniProt
PTM	Proteolytically cleaved; PRP-2, PRP-1, PIF-S and Db-S yield PRP-4, PRP-3 (protein A), PIF-F and Db-F, respectively.	NR	PubMed:18463091		UniProt
Subcellular location	Secreted.	NR			UniProt

(All 7 results shown)

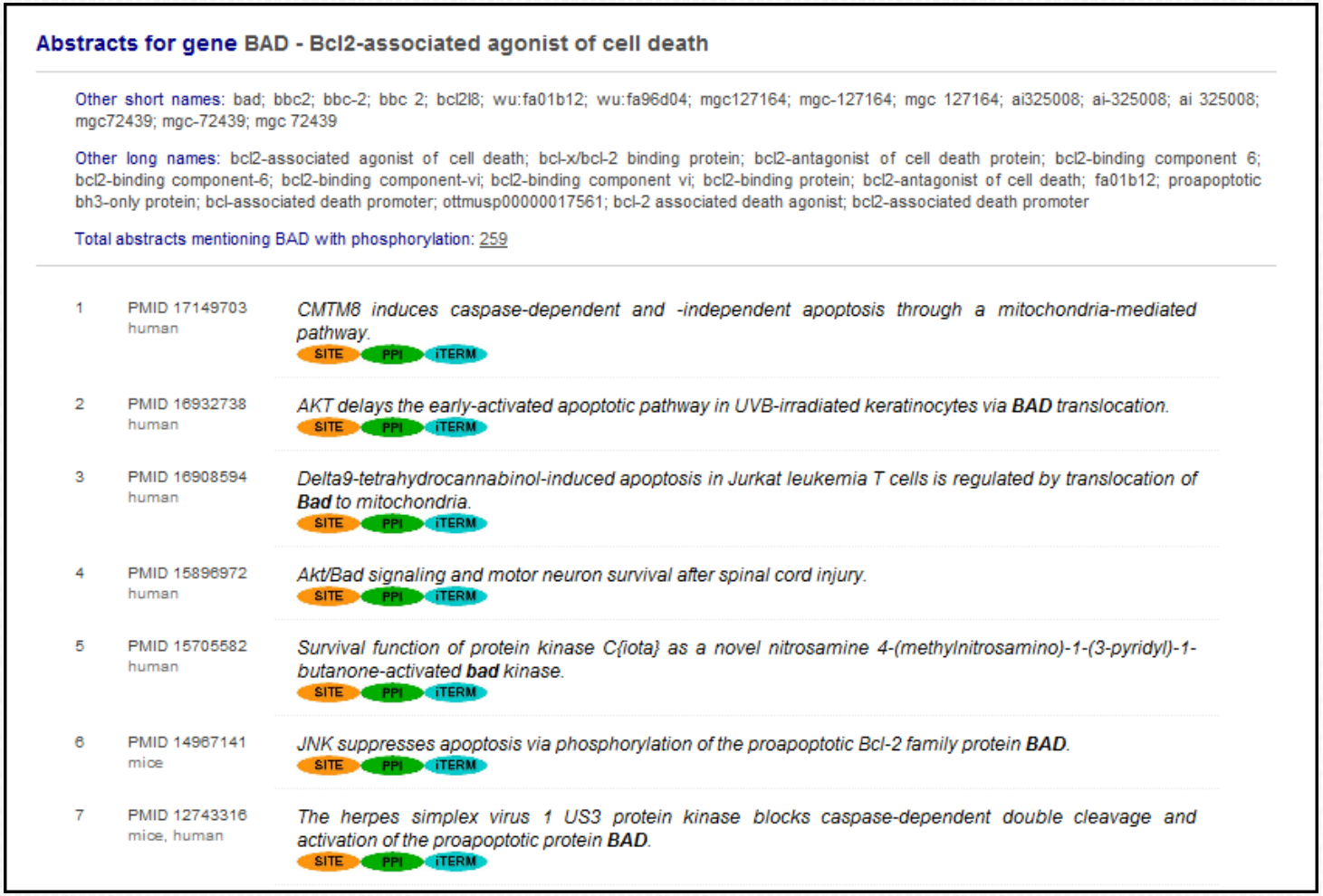
GOALS

- Make the salivary proteome catalog, including its associated experimental data, fully and easily accessible to researchers;
- Encourage community-driven refinement of the catalog through the deposition of new data and annotation on existing content; and
- Facilitate the discovery of therapeutic targets for both oral and systemic diseases.

FUTURE PLAN

- Established communication with the Protein Information Resource (PIR) group at Georgetown University.
- PIR has expertise in both protein curation and text mining.
- Next step is to develop a comprehensive annotation-curation framework/guideline.
- NIDCR is considering the appropriate funding mechanism to support the curation work.

EGIF INTERFACE



Abstracts for gene BAD - Bcl2-associated agonist of cell death

Other short names: bad; bbc2; bbc-2; bbc 2; bcl2l8; wu:fa01b12; wu:fa96d04; mgc127164; mgc-127164; mgc 127164; ai325008; ai-325008; ai 325008; mgc72439; mgc-72439; mgc 72439

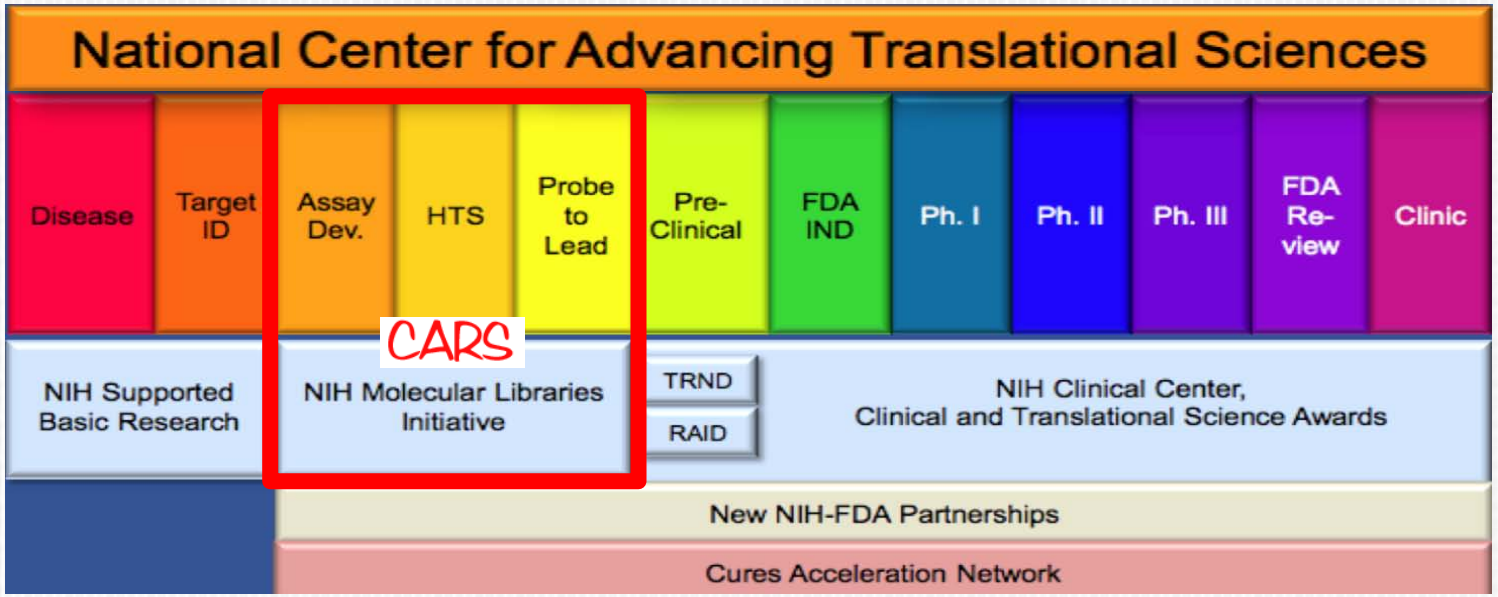
Other long names: bcl2-associated agonist of cell death; bcl-x/bcl-2 binding protein; bcl2-antagonist of cell death protein; bcl2-binding component 6; bcl2-binding component-6; bcl2-binding component-vi; bcl2-binding component vi; bcl2-binding protein; bcl2-antagonist of cell death; fa01b12; proapoptotic bh3-only protein; bcl-associated death promoter; ottmusp00000017561; bcl-2 associated death agonist; bcl2-associated death promoter

Total abstracts mentioning BAD with phosphorylation: 259

1	PMID 17149703 human	*CMTM8 induces caspase-dependent and -independent apoptosis through a mitochondria-mediated pathway.* SITE PPI iTERM
2	PMID 16932738 human	*AKT delays the early-activated apoptotic pathway in UVB-irradiated keratinocytes via **BAD** translocation.* SITE PPI iTERM
3	PMID 16908594 human	*Delta9-tetrahydrocannabinol-induced apoptosis in Jurkat leukemia T cells is regulated by translocation of **Bad** to mitochondria.* SITE PPI iTERM
4	PMID 15896972 human	*Akt/Bad signaling and motor neuron survival after spinal cord injury.* SITE PPI iTERM
5	PMID 15705582 human	*Survival function of protein kinase C{iota} as a novel nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-activated **bad** kinase.* SITE PPI iTERM
6	PMID 14967141 mice	*JNK suppresses apoptosis via phosphorylation of the proapoptotic Bcl-2 family protein **BAD**.* SITE PPI iTERM
7	PMID 12743316 mice, human	*The herpes simplex virus 1 US3 protein kinase blocks caspase-dependent double cleavage and activation of the proapoptotic protein **BAD**.* SITE PPI iTERM

CARS : COMMON ASSAY REPORTING SYSTEM

- HPCIO Staff: Huey Cheung, Adam Frazin, and Sarada Chintala
- Collaborators: Dr. Ajay Pillai of NHGRI and Dr. Linda Brady of NIMH
- A Component of the Molecular Libraries Program (a Commons Fund /NIH Roadmap Project)



NHGRI NIMH National Institute of Mental Health CIT

MOLECULAR LIBRARIES PROGRAM

Technology Development

Chemical Diversity
- Small Mol. Repository
- Library Enrichment

Assays Diversity
- Development
- HTS ready

Instrumentation
- Development

Predictive ADMET
- Algorithms

Screening

Screening and Chemical Probe Generation

Screening/Informatics
- Comprehensive Centers
- Specialized Screening Centers

Chemistry/Pharmacology
- Specialized Chemistry Facilities

Metadata → **Assay Reporting** CARS

Informatics

Screening Data

Data Analysis/Dissemination
- PubChem
- Cheminformatics Research Centers

Data → Public

CARS COMMON ASSAY REPORTING SYSTEM

- Track bioassay project and assays status

Bioassay Projects

Drill Down

Assays Within a Project

View Assays	Delay	Grant Number	Probe Type	UID	Project Name	Project Type	Cycle	Assay Provider	Science Officer	Status ?
	X (41)	MH085683-01	Inhibitor	328	HTS: Redox in mitochondria	Probe	Cyc 10	Carla Koehler	Ron Margolis	P D V(0/0) C
	R (64)	NS059380-01	Inhibitor	528	Mex-5: in vivo probe	Probe	Cyc 9	Sean Ryder	Carson Loomis	P D V(0/0) C
	R (265)	MH085698-01	Inhibitor	333	HTS: RasCE	Probe	Cyc 10	Walter Schmidt	Min Song	P D V(3/3) C
	R (268)	NS061738-01	Inhibitor	504	HTS: Kaposi LANA FP	Probe	Cyc 12	Kenneth Kaye	Min Song	P D V(1/4) C
		MH084117-01	Inhibitor	272	Ras VDAC HTS	Probe	Cyc 8	Brent Stockwell	Dan Zaharevitz	P D V(3/3) C R(3) S(4/4)
		NS059380-01	Inhibitor	297	HTS: MEX-5 with TCR2	Probe	Cyc 9	Sean Ryder	Carson Loomis	P D V(2/2) C R(0) S(1/6)

Project Status

Probe Project: MH085683-01 Inhibitor HTS: Redox in mitochondria

Center: Broad (Allows < 300,000 Compounds)

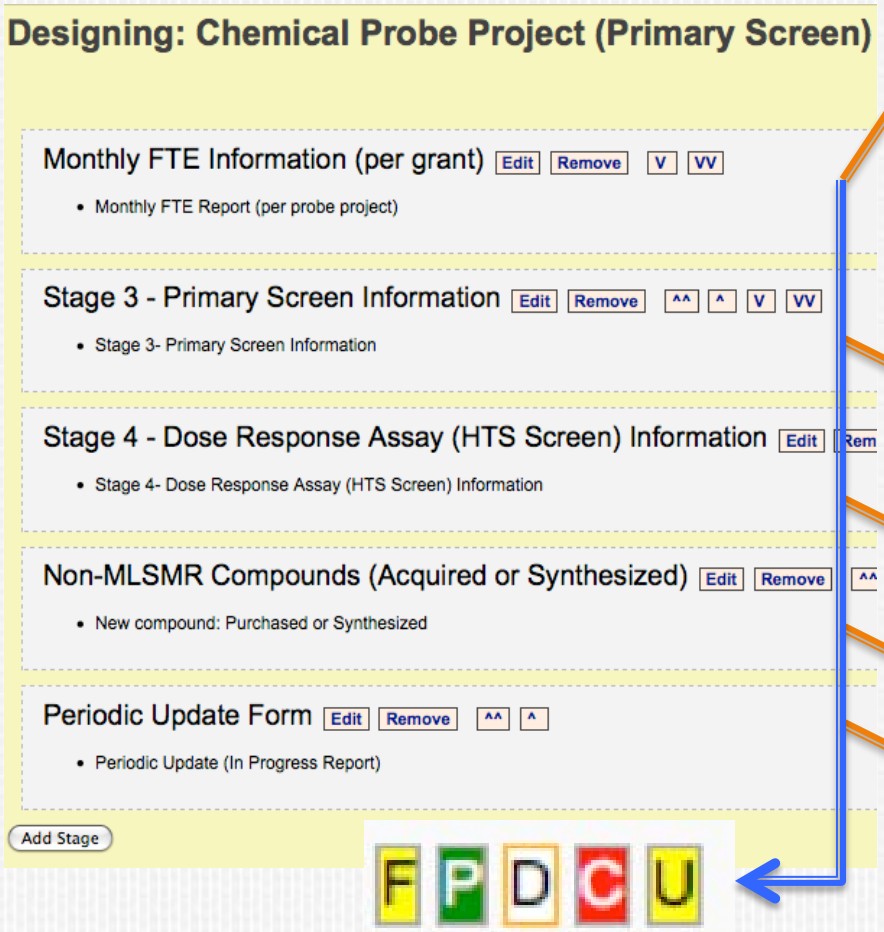
Add Assay | Closeout Report / Closeout | Delete Project | Project Status ? : P D V(0/0) C

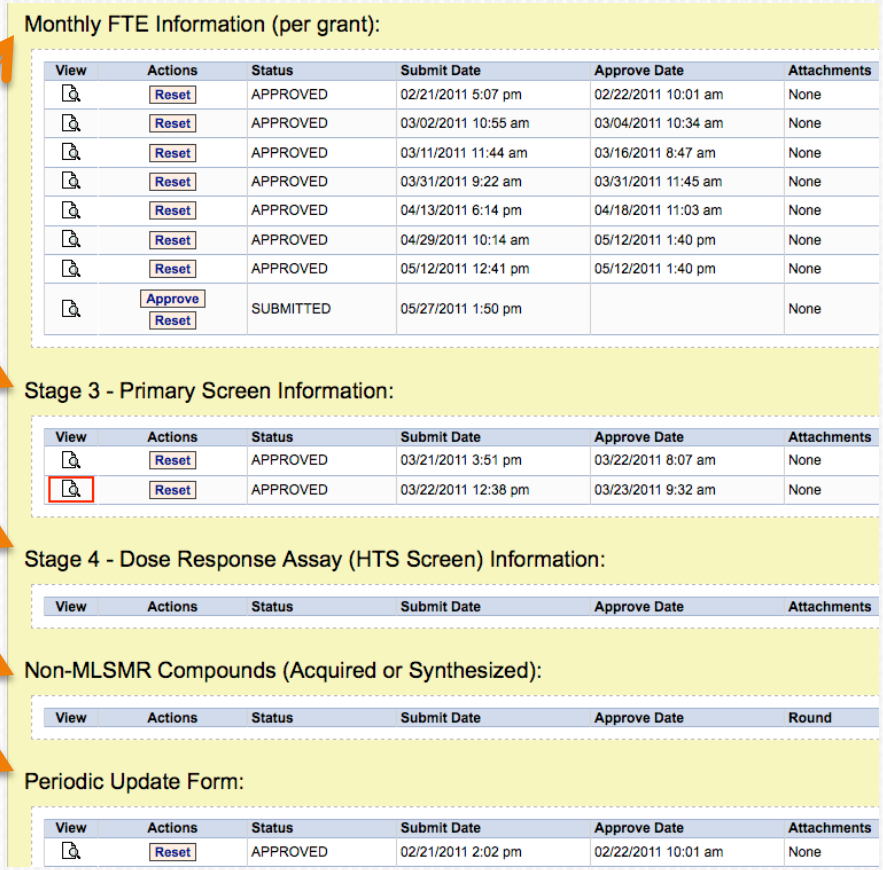
Project Plan	Edit	Delete	Assay	Assay Type	#Days since Submission	Status
			HTS-Fluorescent assay for inhibitors of ALR by detection of H2O2 Production	Primary [HV]	21	F P D C U
			>>> Cell viability assay using MTT with a colorimetric read-out in HUVEC cells	Secondary		S
			>>> Fluorescent counter screen for the activity of H2O2 in the absence of ALR and DTT	Secondary		S

Assay Status

CARS COMMON ASSAY REPORTING SYSTEM

- From Designing a Tracking Workflow
- To Collecting Data on each Stage





NEI/NNRL COLLABORATION

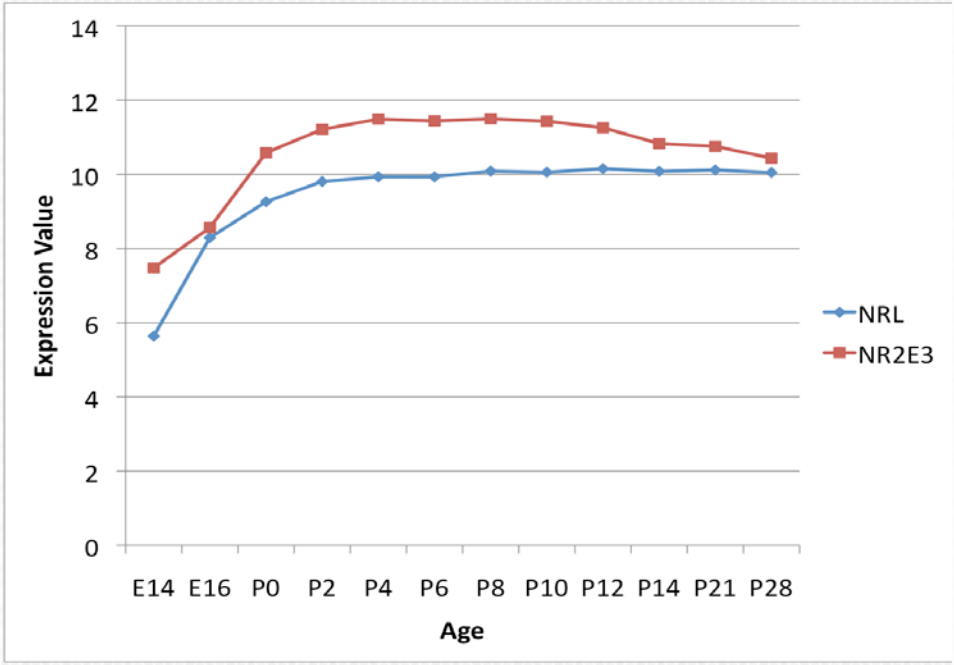
- Division of Computational Bioscience/CIT
  - Daniel Russ, Guoli Wang, Stephen Glanowski, Calvin Johnson
- Neurobiology, Neurodegeneration & Repair Laboratory/NEI
  - Anand Swaroop (Chief, NNRL), Harsha Rajasimha (contractor)

NNRL Research Goals

- Differentiation of retinal neurons from progenitors or stem cells
- Synaptogenesis in the retina.
- Gene regulatory networks in retinal differentiation and disease.
- Genes and pathways underlying retinal degenerative diseases.

INFERRING CO-REGULATION

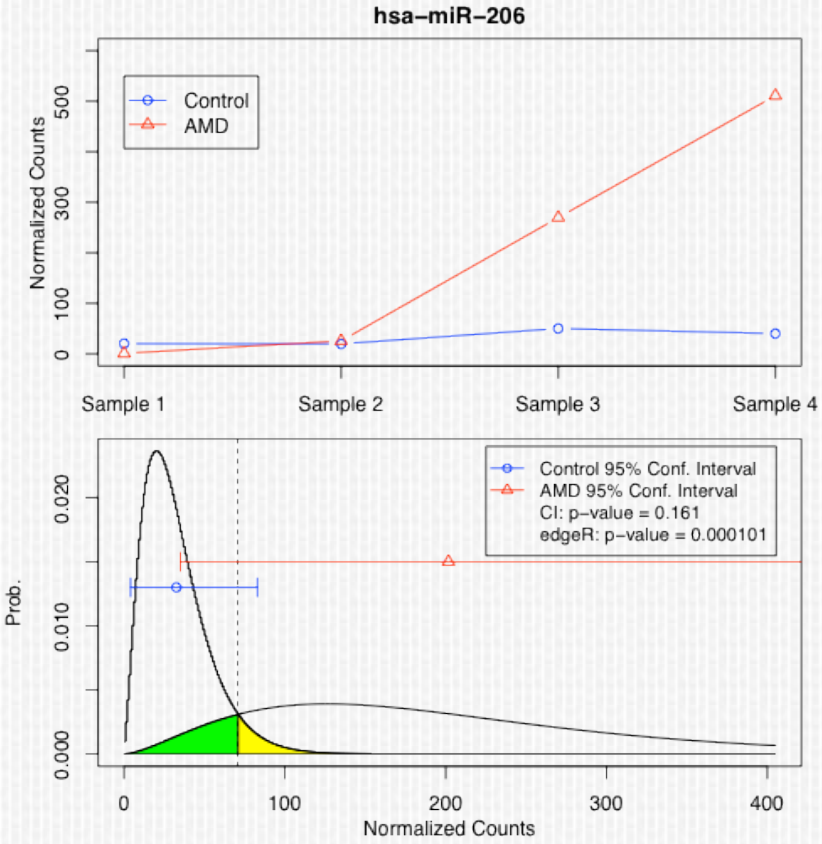
- Identify potentially co-regulated genes by similarities in time sequence data
- Find similarities in shape of curve
- Search for similar responses shifted in time using
  - Matched filter
  - Cross-correlation
  - Cross-covariance
  - Power spectral density
- Estimate gene regulatory network through Kalman filter or extensions.



NRL and NR2E3 are two genes known to co-regulated

IDENTIFYING DIFFERENTIAL EXPRESSION IN HIGH THROUGHPUT SEQUENCING

- DESeq/edgeR inadequate
  - Are the counts the counts from AMD samples different from the control samples?
  - Is the uncertainty within the AMD group small compared to the differences with the control?
- CIT – Confidence Interval Test
  - Build the large non-overlapping confidence interval



IDENTIFICATION OF LYSOSOMES IN MICROSCOPIC IMAGES

- HPCIO staff: William Lau
- Collaborator: Dr. Joseph Mindell (NINDS)

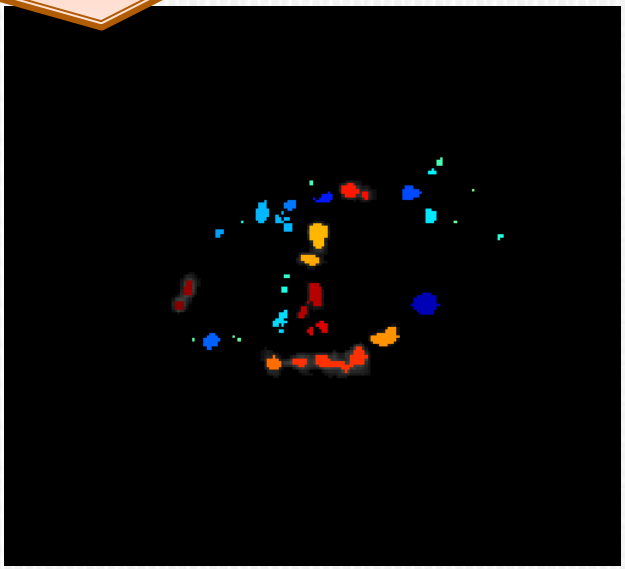
- The goal of this project is to be able to accurately measure the pH of each lysosome in the cell
- The pH of the lysosomes can be calculated by comparing their emission intensity from images taken at different wavelengths

Steps involved:

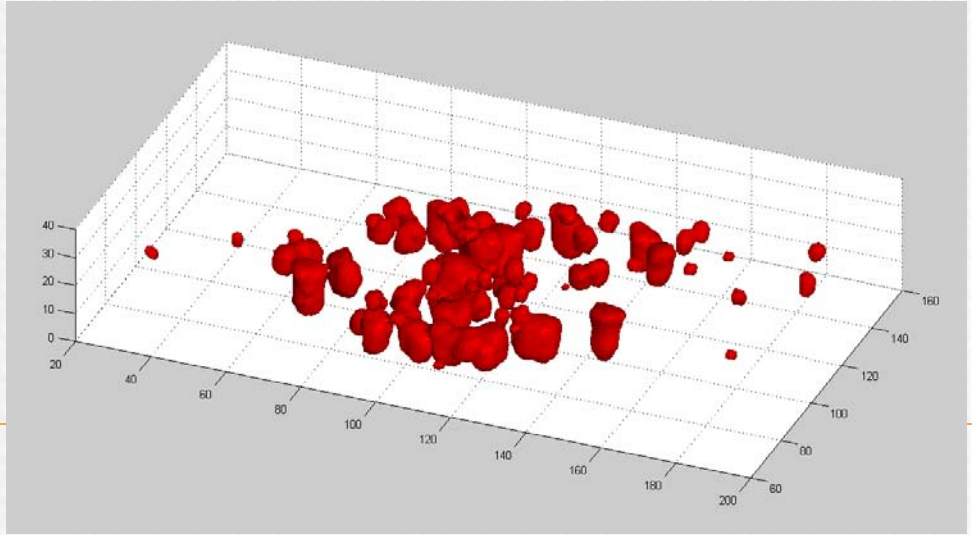
1) Noise Removal
2) Identification
3) Calculation of Intensity Ratio

The tool we developed is able to identify individual objects from the images and highlight them with different colors.

Lysosomes are stained with fluorescence probes that emit light at a certain wavelength.



A 3-D model of the lysosomes is created using the coordinates found.



GENETIC ASSOCIATION DATABASE (GAD)

- HPCIO investigators: Dr. Alex Wang
- Collaborators: Dr. Kevin Becker, Dr. Yongqing Zhang, Dr. Supriyo De (NIA)

- Accumulated ~84,000 records of whether a gene is associated with a particular disease.
- The GAD web site receives millions of page hits every year.
- More than 100 citations for the original GAD paper published in 2004.
- 2008 NIH Director's Award

Genetic Association Database

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

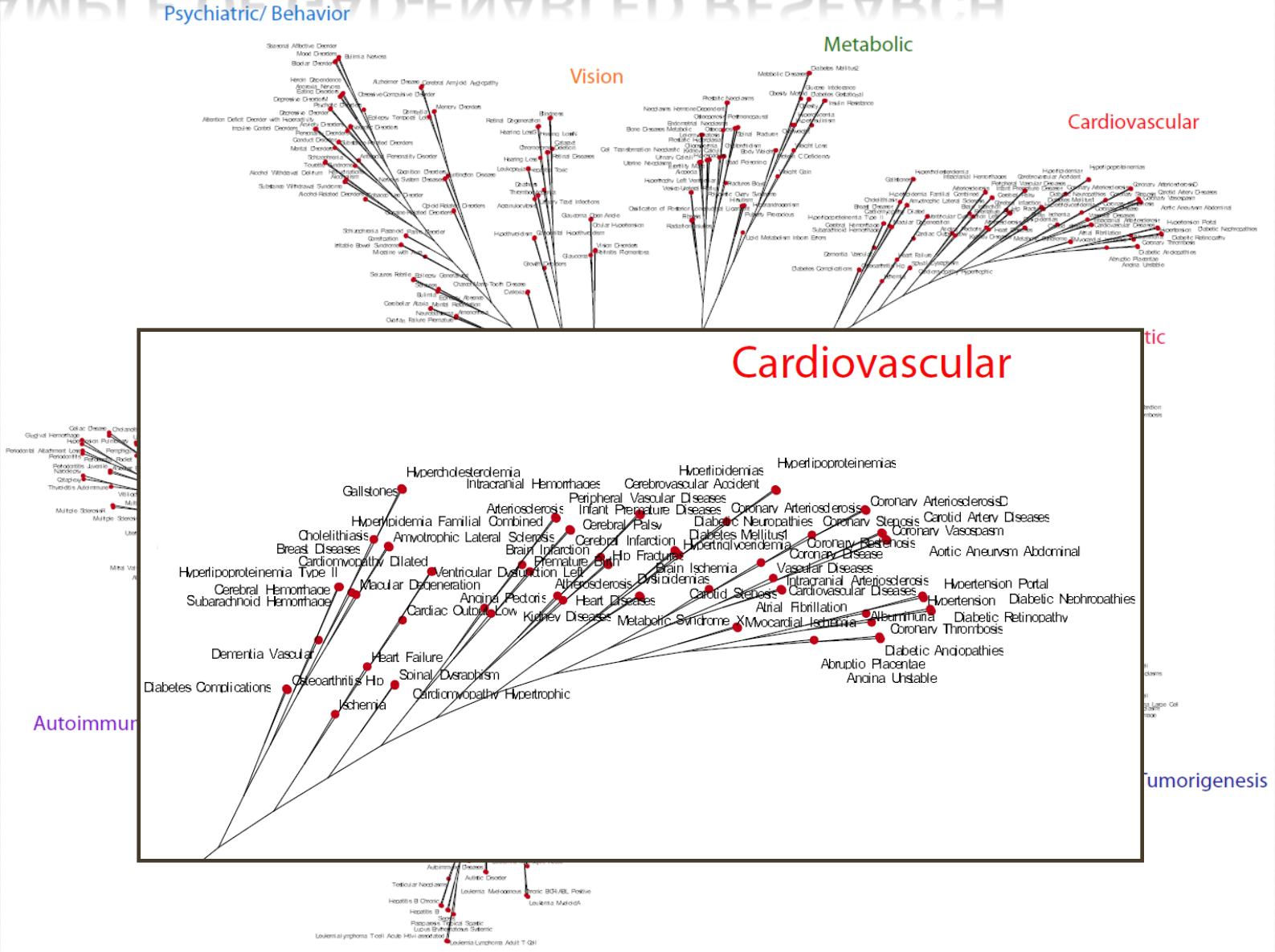
For a complete gene list, click here.

Prev 50 Next 50

Gene View **Search for All** **Record found: 894**

	Assoc? YorN	Gene Symbol	OMIM	Gene Expert	Gene Name	Unigene Cluster	Entrez GeneID	Chr	Ch-Band	DNA Start(bp)	DNA End(bp)	Rep seq	EG	GC	Ace View	BBID	e!	PUB MED	P Value	Disease Class	Broad Phenotype (Disease)
view	N	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM	n	CARDIOVASCULAR	coronary artery d
view	N	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM	n	OTHER	serum lipid levels
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		RENAL	polycystic kidney
view		ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		OTHER	angiotensin I con
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM	<0.02	OTHER	ARDS
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		CARDIOVASCULAR	Blood Pressure
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		OTHER	higher blood pres
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		CARDIOVASCULAR	myocardial infarc
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		METABOLIC	diabetes, type 2
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM	<0.001	NEUROLOGICAL	Alzheimer's Dise
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM	0.043	CARDIOVASCULAR	heart rate variabi
view		ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		IMMUNE	Atopy
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		CARDIOVASCULAR	increased vascul
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		CARDIOVASCULAR	hypertension
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		NEUROLOGICAL	Parkinson's disea
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		OTHER	antiproteinuric ef
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		METABOLIC	elevated ACE
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		CARDIOVASCULAR	carotid wall thicke
view		ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		RENAL	nephropathy dev
view	Y	ACE	106180	Ex	Angiotensin I converting enzym	Hs.298469	1636	17	17q23	61554433	61574979	R	EG	GC	AV	BBID	e!	PM		OTHER	leukoaraiosis

EXAMPLE OF GAD-ENABLED RESEARCH



THIRD-PARTY TOOLS THAT MAKE USE OF GAD DATA

Resource Name	URL *Organization* Description
WikiProteins	http://www.wikiprofessional.org *Department of Medical Informatics, Rotterdam, the Netherlands* WikiProteins enables community annotation in a Wiki-based system.
Disease/Phenotype web-PAGE	http://dpwebpage.nia.nih.gov/ National Institute of Aging Disease/ Phenotype PAGE is a disease focused gene set analysis web tool to analyze microarray gene expression data with predefined groups of disease related genes.
Genome Browser	http://genome.ucsc.edu/ *University of California, Santa Cruz* Genome Browser is a mature web tool for rapid and reliable display of any requested portion of the genome at any scale, together with several dozen aligned annotation tracks.
T1Dbase	http://t1dbase.org/page/Welcome/display *Wellcome Trust Diabetes and Inflammation Laboratory* T1DBase is a community web-based resource for type 1 diabetes research.
WholePathwayScope	http://www.abcc.ncifcrf.gov/wps/wps_index.php *Advanced Biomedical Computing Center, National Cancer Institute* WholePathwayScope is a comprehensive pathway-based analysis tool for high-throughput data.
SNPs3D	http://www.snps3d.org *University of Maryland Biotechnology Institute* SNPs3D is a website which assigns molecular functional effects of non-synonymous SNPs based on structure and sequence analysis.
Autoimmune Disease Database	http://www.sbi.uni-rostock.de/aidb/home.php *Institute for Medical Informatics and Biometry, University of Rostock* Description: Autoimmune Disease Database is a comprehensive literature-based database covering all known or suspected autoimmune diseases.
PolyDoms	http://polydoms.cchmc.org/polydoms *Cincinnati Children's Hospital Medical Center* PloyDoms is a whole genome database for the identification of non-synonymous coding SNPs with the potential to impact disease.
Rat Genome Database	http://rgd.mcw.edu Department of Physiology, Medical College of Wisconsin The Rat Genome Database is the model organism database for the laboratory rat.
GenomeTrafac	http://genometrafac.cchmc.org/genome-trafac/index.jsp *Cincinnati Children's Hospital Medical Center* GenomeTrafac is a comparative genomics-based resource for initial characterization of gene models and the identification of putative cis-regulatory regions of RefSeq Gene Orthologs.
UniProtKB	http://www.uniprot.org *UniProt Consortium*

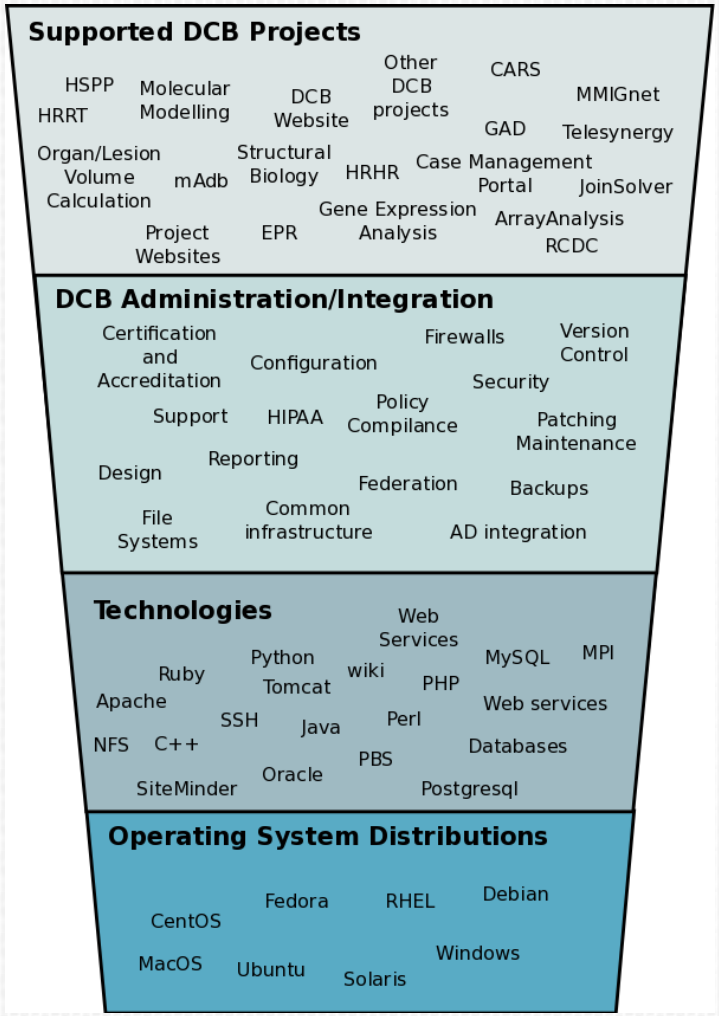
DCB SCIENTIFIC COMPUTING FACILITY

- HPCIO staff: Anthony Fletcher, Shahar Goldin [C], Mark Roth [C], Pamela Hill [C].
- Collaborators: DCB staff and their collaborators.
- Support DCB projects with NCI, NIAID, FDA, CC, NHGRI, NIMH, NIA, OD, NIDDK, NHLBI, NEI; also support NIH as a whole through MMIGnet, UNIX support, support of DCB resources.

DCB SCIENTIFIC COMPUTING FACILITY

- Server space in building 12A, with no guaranteed level of service. In practice, the service level is approximately 99%.
- Systems needing guaranteed levels of service are placed in the CIT Data Center.
- Variety of systems, ranging from the small (e.g., a 2-CPU, 1U web server) to medium (48 core, 64 GB RAM SMP computational server) and 60 cluster nodes in two clusters.
- Separate racks on a private network for systems with PHI.

DCB SCIENTIFIC COMPUTING RESEARCH FACILITY



DCB Computaional Facility