Annotation Analytics for Gene and Protein functions

Nigam Shah, MBBS, PhD nigam@stanford.edu

NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

Annotation service

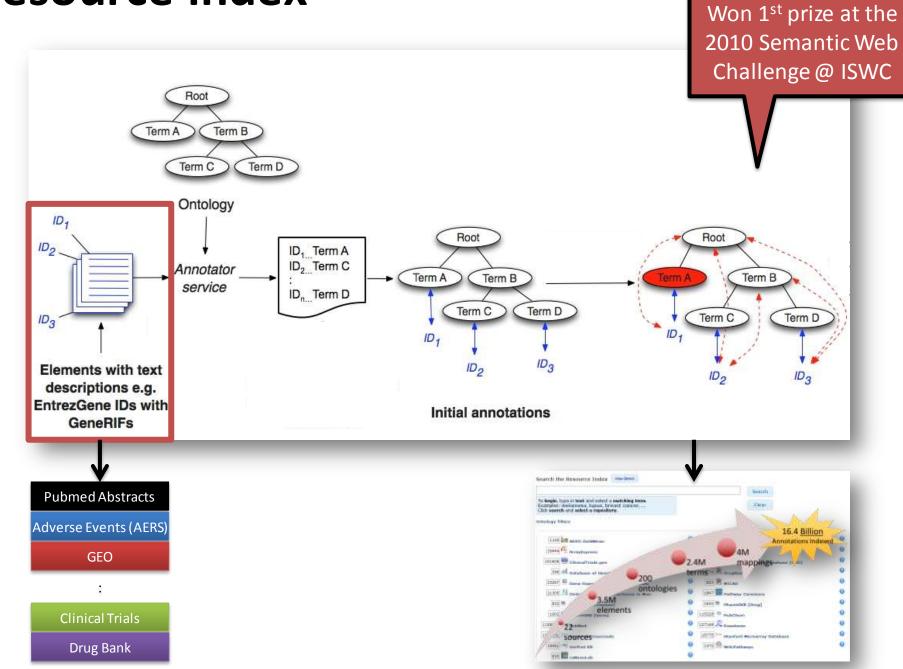
Process textual metadata to automatically tag text with as many ontology terms as possible.

Accession:	GDS906 @ View Expression (GEO profiles)									
Title:	Bladder smooth muscle cell response to mechanical stretch									
DataSet type:	gene expression array-based (RNA / in situ oligonucleotide)									
Summary:	Expression profiling of cultured bladder smooth muscle cells subjected to repetitive mechanical stimulation for 4 hours. Chronic overdistension results in bladder wall thickering, associated with loss of muscle contractility. Results identify genes whose expression is altered by mechanical stimuli.									
Platform:	GPL95: Alfymethx Genechip Human Genome U133 Array Set HG-U133A									
Citations:	Adam RM, Eaton SH, Estrada C, Nimgaonkar A et al. Mechanical stretch is a highly selective regulator of gene expression in human bladder smooth muscle cells. Physiol Genomics 2004 Dec 15:20(1):36-44. PMID: 15:467014									
Sample organism:	Homo sapiens	Platform organism:	Homo sapiens							
Feature count:	22283	Value type:	count							
Series:	GSE1595	Series published:	07/25/2004							
Last GDS update:	12/29/2004		20-							



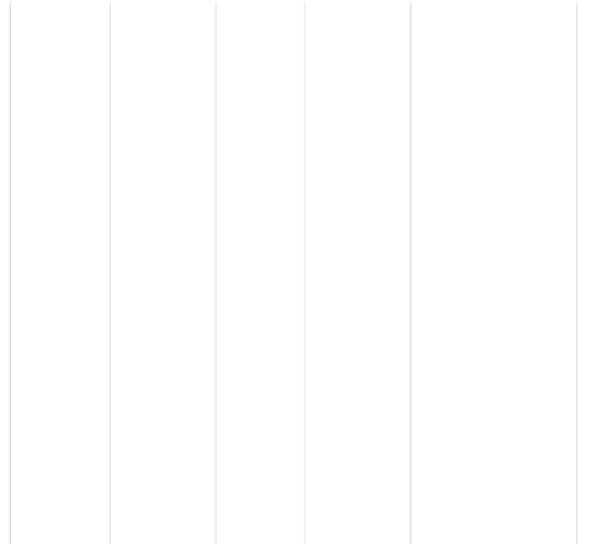
107 million calls, ~1000 GB data Expression, Expression of bladder, bladder, smooth, bladder muscle, muscle, smooth muscle, cells, mechanical, mechanical stimulation, stimulation, Chronic, results, bladder overdistension, associated, associated with, with, loss, genes, altered

Resource index



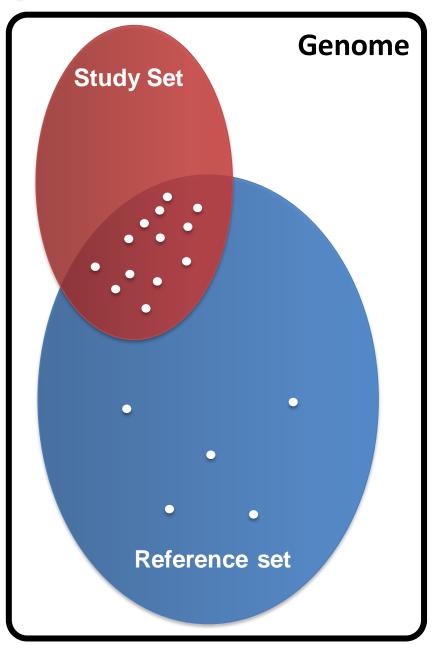
Understanding the genome

- Units of study range in length from 'whole chromosome' to 'singe nucleotide'
- E.g. three copies of Chr.
 21 → Down's syndrome
- The focus in on finding the functional associations of strings in the genome

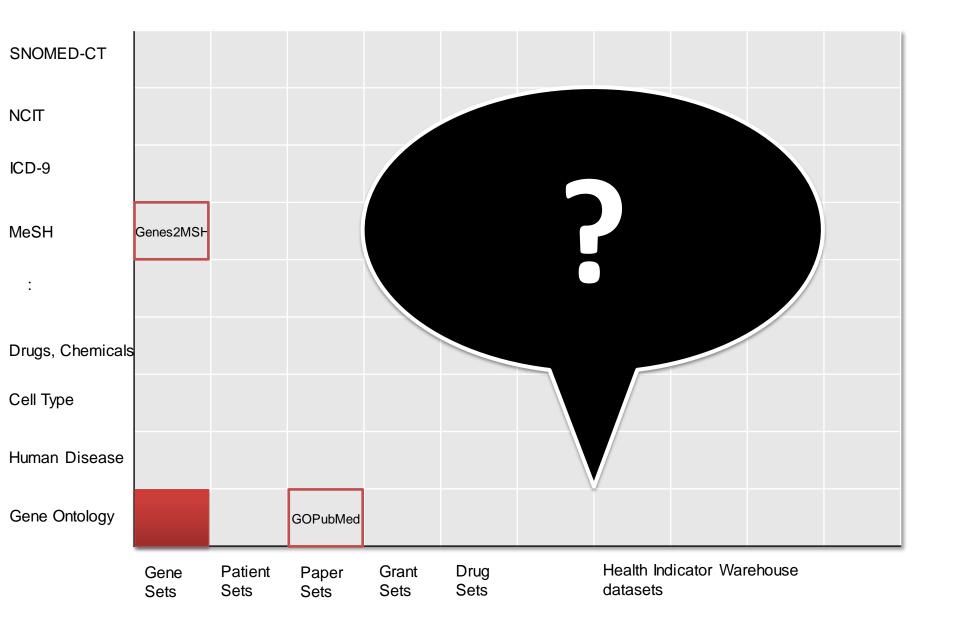


Generic GO based analysis routine

- Get annotations for each gene in a set
- Count the occurrence of each annotation term in the study set
- Count the occurrence of that term in some reference set (whole genome?)
- P-value for how surprising their overlap is.



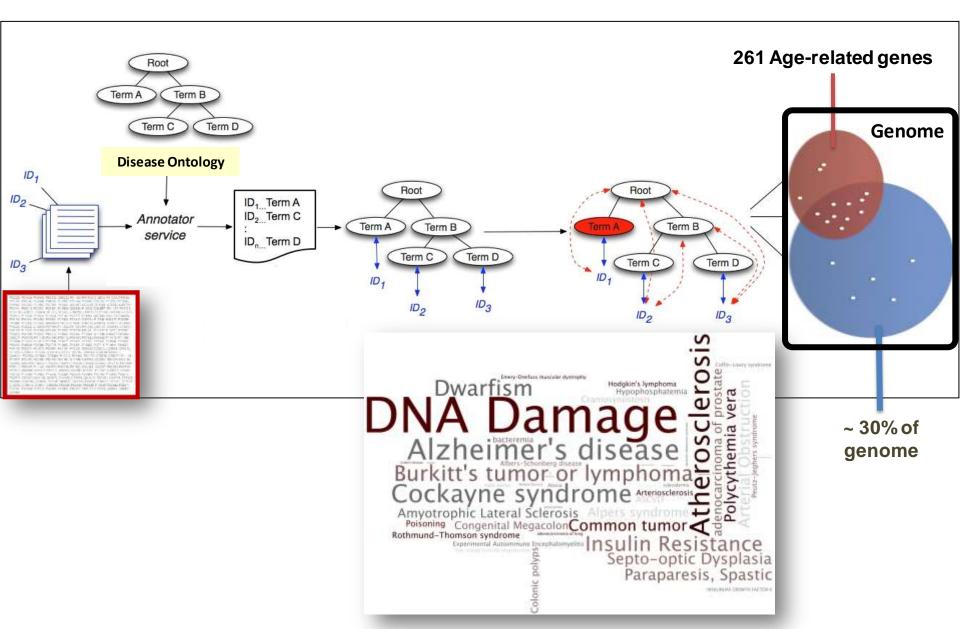
Annotation Analytics Landscape



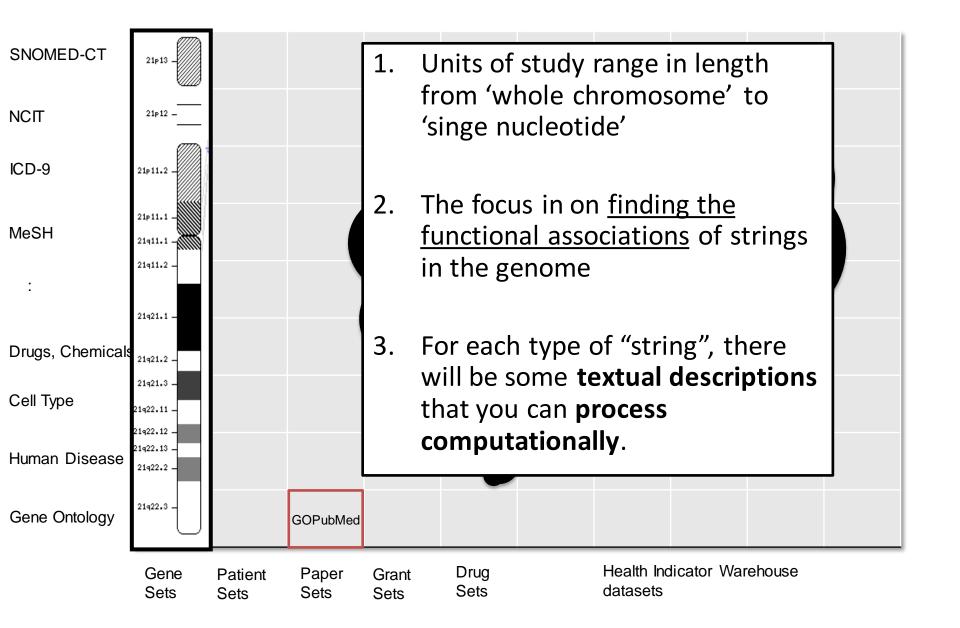
Mutation enrichment

	PHOS	METH	GLY	ACC	COIL	DIS	UBIQ	SHEET	CALM	CAT	HELIX	BUR	SIFT	STAB	
Immune Function	-1.57	-1.31	-1.75	-2.44	-1.56	-3.00	-2.64	1.56	-2.19	-1.02	1.46	1.87		1.11	
Neoplasm	1.15	-1.23	-2.22	<u>-1.59</u>	-1.21	-1.12	-1.77	-1.08	-1.62	1.09	1.50	1.54	2.69	1.12	
Musculoskeletal	-3.60	-1.84	-8.02	-2.19	-1.15	1.53	-4.01	-1.11	1.40	1.21	1.41	1.80		1.11	
Metabolic	-3.18	-3.44	-4.61	-2.55	-1.19	-1.31	-4.52	-1.13	-1.60	1.30	1.50	1.89		1.11	
Dermatological	-1.56	-2.08	-3.66	-2.19	-1.19	1.15	-3.67	-1.25	-1.30	-1.05	1.56	1.80		1.05	
Digestive System	-1.50	-1.40	-3.50	-1.76	-1.39	-1.87	-1.75	-1.22	1.35	1.24	1.86	1.63	2.80	1.13	
Cardiovascular	-1.12	-1.49	-2.40	-1.52	-1.36	-1.03	-1.20	-1.25	2.58	1.07	1.83	1.50	2.91	1.13	
Ophthalmological	-1.84	-3.27	-1.40	-2.04	-1.22	-1.85	-4.93	-1.15	1.51	1.05	1.57	1.75	2.96	1.10	
Endocrine	-1.56	-2.08	-2.50	-2.00	-1.47	-1.91	-6.26	-1.45	1.44	1.30	2.06	1.73	3.1.0	1.13	
Genitourinary	-2.59	-5.17	-3.46	-2.06	-1.32	-1.95	-2.60	1.08	2.20	1.17	1.56	1.75	2.95	1.16	
Nervous System	-2.07	-1.02	-3.45	-1.95	-1.39	-2.18	-4.61	1.07	1.47	1.23	1.66	1.71	2.99	1.17	
HGMD Disease Mutations	-2.12	-1.80	-3.58	-1.99	-1.28	-1.41	-2.29	1.05	1.40	1.16	1.51	1.73		1.11	
Connective Tissue	-2.97	-2.97	0.00	-2.42	1.05	2.17	-8.95	-1.20	1.28	1.11	1.01	1.86		-1.00	
Blood	-3.30	-1.46	-4:41	-2.17	-1.52	-2.11	0.00	1.02	-1.10	1.10	1.85	1.79	3.23	1.10	
HGMD Functional Polymorphisms	-2.70	-1.35	-5.42	<u>-1.42</u>	-1.48	-2.24	2.58	-1.73	-1.13	-1.09	2.16	1.43	2.41	1.05	
Blood Coagulation	-2.38	-2.38	-14.31	<u>-1.92</u>	-1.13	-3.94	-1.79	1.81	1.07	1.32	-1.55	1.70		1.17	
Nutritional	-4.89	-3.26	-2.18	-2.66	-1.49	-7.11	0.00	-1.03	-1.09	1.33	1.85	1.91	3.20	1.17	
Ear, Nose, Throat	-1.78	-1.19	0.00	-1.55	-1.35	4.74	0.00	1.40	2.70	1.32	1.33	1.52	2.69	1.14	
Lymphatic	0.00	0.00	0.00	-1.44	-1.21	-2.11	0.00	2.04	0.00	0.00	-1.67	1.45	3.20	1.00	
Respiratory	-1.26	0.00	2.37	-1.49	-1.14	-1.01	2.36	1.07	3.33	-1.50	1.26	1.48	2.32	1.04	
Psychiatric	1.22	2.56	1.10	-1.37	-1.19	1.30	0.00	-1.61	4.10	1.17	1.72	1.39	2.52	1.09	
Developmental	1.43	2.14	-1.41	<u>-1.38</u>	1.01	1.34	-1.39	-1,20	4.00	-1.15	1.12	1.40	2.91	1.08	
Cancer	2.21	2.26	-1.33	-1.05	-1.12	1.67	-1.11	-1.34	3.01	-1.36	1.47	1.06	1.95	1.00	
Kinase	3:41	-2.10	-2.11	-1.05	1.04	1.19	-1.06	-1.64	1.71	-1.05	1.21	1.07	2.24	-1.07	
HGMD Disease Polymorphisms	2,18	-1.53	1.52	-1.08	-1.06	1.32	0.00	-1.43	1.31	1.05	1.39	1.10	1.00	-1.03	
		Fold Depletion							Fold Enrichment						
		>-5 -5 -3 -2 -1 0							2	2	. Sha E	4<			
					-2	-	U	-	-						

Profiling a set of Aging genes



Annotation Analytics Landscape



The team @ www.bioontology.org/project-team

NIH Roadmap grant U54 HG004028

