

GeneHub-GEPIS: Digital Expression Profiling for Normal and Cancer Tissues Based on an Integrated Gene Database

Yan Zhang, Shih-Ming Luoh, Lawrence S. Hon, Robert Baertsch¹, William I. Wood and Zemin Zhang*

Department of Bioinformatics, Genentech, Inc., South San Francisco, CA 94080, USA

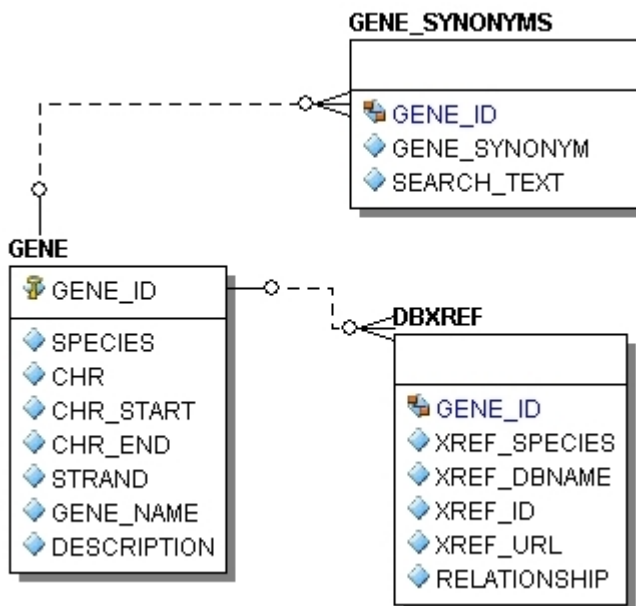
¹Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA

Supplemental Materials:

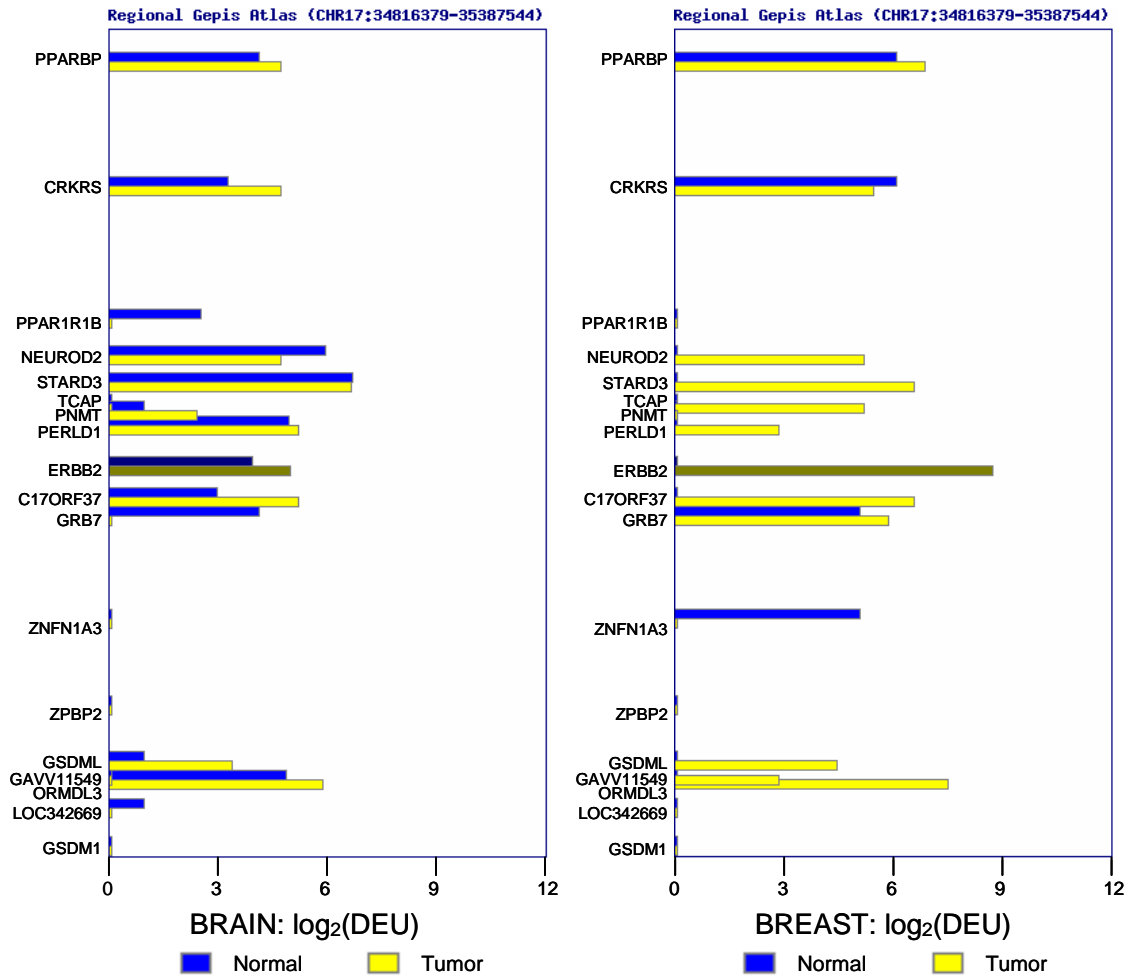
List of content:

- Supplemental Figure 1.
- Supplemental Figure 2.
- Supplemental Figure 3.
- Supplemental Table 1.
- Supplemental Table 2.

Supplemental Figure 1: Database schema. The gene attributes and synonyms are stored in two tables, GENE and GENE_SYNONYMS, respectively. Both homolog or ortholog cross-references between GeneHub gene indexes and database records are saved in the DBXREF table.



Supplemental Figure 3. Example of a GEPIS Regional Atlas result. Users can select multiple tissue types to study as well as the size of the genomic window, centering on the input gene, to view. In this example, genes located within 300 kb on each side of the ERBB2 gene on chromosome 17 are included in the analysis for brain and breast tissues. The log value of digital expression unit for each gene is a plotted along the x-axis, and the y-axis shows the location and name for every gene in the window.



Supplemental Table 1. Number of records in each of the database sources used by GeneHub.

Database Type	Database Name	Human	Mouse	
mRNA	RefSeq *	29,521	45,309	
	Ensembl Known *	29,387	27,617	
	Proteome *	19,343	20,622	
	FATNOM Known *	-	54,854	
	GenBank	138,360	110,359	
	Ensembl Novel	4,482	8,854	
Protein	UniProt	69,842	59,391	
	PDB	16,735	3,595	
Gene	Entrez Gene	36,013	51,065	
Microarray	Affy Human U133P	54,675	-	
	Affy Human U133A	22,283	-	
	Affy Human U133B	22,645	-	
	Affy Human U95Av2	12,479	-	
	Affy Human U95B	12,620	-	
	Affy Human U95C	12,646	-	
	Affy Human U95D	12,644	-	
	Affy Human U95E	12,639	-	
	Affy Human U35KA	8,934	-	
	Affy Human U35KB	8,924	-	
	Affy Human U35KC	8,928	-	
	Affy Human U35KD	8,928	-	
	Affy Human FL68	6,633	-	
	Affy Mouse Genome 430 2.0	-	45,101	
	Affy Mouse 430A	-	22,626	
	Affy Mouse 430B	-	22,511	
	Affy Murine 74A	-	12,488	
	Affy Murine 74B	-	12,477	
	Affy Murine 74C	-	11,934	
	Affy Murine 11KA	-	6,584	
	Affy Murine 11KB	-	6,002	
	Agilent Human 1A	17,086	-	
	Agilent Human 1Av2	20,173	-	
	Agilent Human 1B	19,673	-	
	Agilent Whole Humn Genome	41,000	-	
	Agilent Mouse V2	-	20,868	
	Agilent Whole Mouse Genome	-	41,174	
	Combined	GeneHub	31,999	34,794

(* core transcript sets used for deriving GeneHub genes)

Supplemental Table 2. Distribution of ESTs in usable libraries:

Organ	Total	Normal	Tumor	Total	Normal	Tumor
	Human			Mouse		
ADIPOSE_TISSUE	15533	13834	1699	<i>14</i>	<i>14</i>	<i>0</i>
ADRENAL	36922	19061	17861			
BONE	52633	12668	39965	30319	30319	0
BONE_MARROW	48113	20311	27802	113285	113285	0
BRAIN	705566	515939	189627	229874	229874	0
BREAST	195527	59033	136494	249503	10733	238770
CARTILAGE	36300	16211	20089			
CERVIX	39427	2576	36851			
COLON	221020	38203	182817	50336	49555	781
EAR				32199	32199	0
ESOPHAGUS	17717	87	17630			
EYE	104400	52863	51537	39643	39643	0
GENITOURINARY	43705	9586	34119	11696	9955	1741
GI_TRACT	43267	28590	14677	31945	31945	0
HEAD_N_NECK	183837	45646	138191	31869	31869	0
HEART	21198	21198	0	24734	24734	0
KIDNEY	128658	45782	82876	120192	117954	2238
LIVER	196342	108706	87636	83206	73368	9838
LUNG	219169	74876	144293	91011	23263	67748
LYMPH_NODE	46659	25462	21197	1578	1578	0
LYMPHORETICULAR	66580	19355	47225	9345	3390	5955
MUSCLE	65922	34215	31707	25555	25555	0
NERVOUS	34602	34602	0	125927	123715	2212
OVARY	116612	8902	107710	17901	17901	0
PANCREAS	211295	94668	116627	61115	61115	0
<i>PARATHYROID</i>	<i>53</i>	<i>0</i>	<i>53</i>			
PERIPHERAL_NERV	21290	20092	1198	14637	14637	0
PINEAL_GLAND	8534	8534	0	2987	2987	0
PITUITARY_GLAND	18649	16360	2289	8047	8047	0
PLACENTA	219794	176385	43409	27639	27639	0
PROSTATE	144727	60159	84568	24768	24768	0
RETINA	23182	23182	0	37368	37368	0
SALIVARY_GLAND	13568	2647	10921	20853	19880	973
SKIN	182179	44078	138101	75179	41525	33654
<i>SOFT_TISSUE</i>	<i>225</i>	<i>225</i>	<i>0</i>			
SPLEEN	49015	49015	0	73328	54477	18851
STOMACH	107774	32681	75093	14481	14481	0
SYNOVIUM	2542	278	2264	17859	17859	0
TESTIS	228165	95411	132754	121236	118272	2964
THYMUS	3099	2958	141	43108	21451	21657
THYROID	78482	18650	59832	4546	4546	0
UTERUS	174205	10151	164054	1119	1119	0
VASCULAR	49393	49393	0	<i>601</i>	<i>601</i>	<i>0</i>
Total:	4175880	1912573	2263307	1869003	1461621	407382

The tissues that are represented by less than 1,000 total ESTs are italicized and are not analyzed.