

Progress in PAINT-Powered Protein Property Propagation

Mike Livstone

Transfer (“Propagate”) GO terms from annotated proteins to less-annotated homologs

FAMILY NOT NAMED (PTHR11829)

Protein Family Tree Browser

Molecular Function Cellular Component Biological Process

Tree

The tree shows the evolutionary relationship between FOXA2 and its orthologs. Major clades include:

- Clade 1: MACMU_ENSMUG00000010410 (FOXA2)
- Clade 2: BOVIN_ENSBTAG00000012407 (Foxa2)
- Clade 3: CANFA_ENSCAFG00000005168 (Foxa2)
- Clade 4: MONDO_ENSMODG00000006213 (HNF-3beta)
- Clade 5: ORNAN_ENSOANG00000015108 (foxa2)
- Clade 6: FUGRU_ENSTRUG00000015568 (foxa2)
- Clade 7: FUGRU_ENSTRUG00000008187 (foxa2)

Table

Species	Description
homo sapiens	Hepatocyte
macaca mulatta	Hepatocyte
mus musculus	Hepatocyte
rattus norvegicus	Hepatocyte
bos taurus	
canis familiaris	Hepatocyte
monodelphis domestica	Hepatocyte
ornithorhynchus anatinus	Hepatocyte
gallus gallus	Hepatocyte
xenopus tropicalis	Forkhead b
fugu rubripes	
fugu rubripes	Hepatocyte
danio rerio	Forkhead b

MSA

Associations, Evidence, Annotations

Annotation Matrix

GO Terms (partial list):

- :0042826 histone deacetylase binding
- :0051525 NFAT protein binding
- :0008134 transcription factor binding
- :0016563 transcription activator activity
- :0030528 transcription regulator activity
- :0043565 sequence-specific DNA binding
- :0003677 DNA binding
- :0003676 nucleic acid binding
- :0035035 histone acetyltransferase binding
- :0016565 general transcriptional repressor
- :0016564 transcription repressor activity
- :0042803 protein homodimerization activity
- :0042802 identical protein binding
- :0046983 protein dimerization activity
- :0003700 transcription factor activity
- :0003714 transcription corepressor activity
- :0003712 transcription cofactor activity
- :0003705 DNA polymerase II transcription

PAINT Products

- GAF file
 - Annotations to extant proteins (RefGenome & non-RG)
 - 1° annotations to ancestral nodes
 - 2° annotations inherited by ancestral nodes
- Evidence file
 - Brief description of phylogeny & scope of annotations
 - Reasoning behind annotations
 - Questions
 - MOD curators
 - Ontology curators

These and some other accessory files are deposited in CVS.

<http://cvsweb.geneontology.org/cgi-bin/cvsweb.cgi/go/gene-associations/submission/paint/>

Initial SOP

http://gocwiki.geneontology.org/index.php/RefG_Princeton_April_12-13_2010

- Select the most specific GO term possible.
- Apply the term as widely as possible.
- When necessary, suggest changes to annotations or ontologies
- Use or address *every* experimental annotation.

Initial effort: Forkhead family (PTHR11829)

- Originally chosen for the lung development project.
- Large family (672 proteins) of transcription factors involved in multiple developmental pathways.
- Generated ~15,000 new annotations
- Time required to annotate with PAINT:
 - MF: 3 days
 - CC: 2 days
 - BP: 4 weeks
- *Highly productive, but much too slow*

Q: What was the major obstacle?

A: The need to address *every* annotation (especially in BP), often by reading the paper.

Solutions

- Abandon the need to address every unused EXP annotation.
 - “response to” IEP’s
 - Downstream effects (IMP’s)
- Impose deadlines: Keep pace with the literature curation.
Don’t PAINT for longer than the curation period.
- Prioritize the most important EXP annotations

Impose deadlines

Active annotation period (Dates)	Panther Family	Gene Symbol (Human protein ID)	PAINT due to begin
Week 1-2 (1st-16th July)	PTHR12027:SF15, PTHR12027:SF18, PTHR12027:SF25	wnt1 (P04628) wnt3 (P56703)	Week 5
	PTHR12027:SF21, PTHR12027:SF23	wnt2 (P09544) wnt5 (P41221, Q9H1J7)	Week 5
	PTHR12027:SF4, PTHR12027:SF6, PTHR12027:SF7, PTHR12027:SF19	wnt4 (P56705) wnt8 (Q9H1J5, Q93098) wnt9 (O14904, O14905) wnt11 (O96014)	Week 5
	PTHR12027:SF14	wnt6 (Q9Y6F9)	Week 5
	PTHR12027:SF16, PTHR12027:SF22, PTHR12027:SF26	wnt7 (O00755, P56706) wnt16 (Q9UBV4)	Week 5
	PTHR12027:SF11	wnt10 (WNT10A Q9GZT5, WNT10B O00744)	Week 5
Week 3 (17th-23rd July)	PTHR23315	CTNNB1 (P35222)	Week 7
	PTHR16505	CTNNBIP1 (Q9NSA3)	Week 7
Week 4 (24th-31st July)	PTHR10845:SF11	Axin1 (O15169)	Week 8
	PTHR10845:SF103	Axin2 (Q9Y2T1)	Week 8
Week 5-6 (1st-13th August)	PTHR11309:SF31, PTHR11309:SF32, PTHR11309:SF33, PTHR11309:SF34, PTHR11309:SF49	fzd1 (Q9UP38) fzd2 (Q14332) fzd7 (O75084)	Week 9
	PTHR11309:SF28, PTHR11309:SF29	fzd5 (Q13467) fzd8 (Q9H461)	Week 9
	PTHR11309:SF23	fzd4 (Q9ULV1)	Week 9
	PTHR11309:SF24	fzd10 (Q9ULW2)	Week 9
	PTHR11309:SF25	fzd9 (O00144) fzd3 (Q9NPG1)	Week 9
	PTHR11309:SF21	fzd6 (O60353)	Week 9
Week 6-7 (14th-20th August)	PTHR11309:SF7, PTHR11309:SF10	SFRP1 (Q8N474) SFRP2 (Q96HF1) SFRP4 (Q6FHJ7)	Week 10
	PTHR11309:SF10	SFRP5 (Q5T4F7)	Week 10
Week 8 (21st-31st August)	PTHR10878	dvl1 (O14640) dvl2 (O14641) dvl3 (Q92997)	Week 11
Week 9-10 (1st-10th September)	PTHR10529:SF108	LRP5 (O75197)	Week 13
	PTHR10529:SF109	LRP6 (O75581)	Week 13
Week 11 (11th-17th September)	PTHR12607	APC (P25054) APC2 (O95996)	Week 14
	PTHR10373:SF11	LEF1 (Q9UJU2)	Week 15
Week 12-16 (18th September - 22nd October)	Review annotations with respect to PAINT inferences)		

http://gocwiki.geneontology.org/index.php/Wnt_signaling_Pathway#Families_to_annotate_28BATCH_1.29

Establish Priorities

Example: the Wnt family

- 1) Signaling pathways (especially those mentioning “Wnt” by name, such as GO:0060070 “canonical Wnt receptor signaling pathway”)
- 2) Signaling pathway/developmental pathway cross-products
- 3) Regulation of transcription as a result of signaling
- 4) Developmental terms
 - a) organs and tissues
 - b) high-level pattern formation

New handling of questions: Wiki instead of SourceForge

- Easier access
- All discussions are public
- Better structure and organization
- Easily exported to Evidence file

Examples:

<http://gocwiki.geneontology.org/index.php/PTHR23315>

<http://gocwiki.geneontology.org/index.php/PTHR12027>

PTHR23315

Questions that arose during PAINT curation

Contents [hide]

1 Questions for MOD curators

- 1.1 Phylogeny
- 1.2 MF
 - 1.2.1 Multiple organisms
 - 1.2.2 Mouse
- 1.3 CC
 - 1.3.1 Human
 - 1.3.2 Mouse
 - 1.3.3 Fly
- 1.4 BP
 - 1.4.1 Mouse

2 Questions for ontology curators

- 2.1 MF
- 2.2 CC
- 2.3 BP

Statistics

April-August 2010

Family	Description	# proteins	Inferences	Annotations	Time elapsed
PTHR10046	LonP	115	20	1085	(SOP exemplar)
PTHR22573	HPRT	233	46	1900	MF+CC 3 days, BP 1 week
PTHR11829	Forkhead	672	387	14858	5 weeks
PTHR11447	p53	43	30	973	1 week
PTHR10202	presenilin	42	26	784	1 week
PTHR11361	MutS & MSH's	203	57	557	3 days
PTHR24221	TAP2	434	44	2492	1 week
PTHR12027	Wnt's	294	241	5151	2 weeks
PTHR23315	Beta-catenin	172	94	1796	4 days
PTHR16505	Beta-catenin interacting protein	32 (15)	7	98	1 day
PTHR10845	Axin1 & -2	383 (29)		1735	1 day