Co-curation of Transcription Factors regulating Heart Development

A Reference Genome project

Lead curator – Varsha Khodiyar (BHF-UCL)

http://www.cardiovasculargeneontology.com

The motivation behind this project

- Heart development
  - Interesting from evolutionary viewpoint
  - High relevance to human cardiovascular (CV) disease
  - Model organisms vital to CV research

- Transcription
  - Relevant to all MODs

Aims

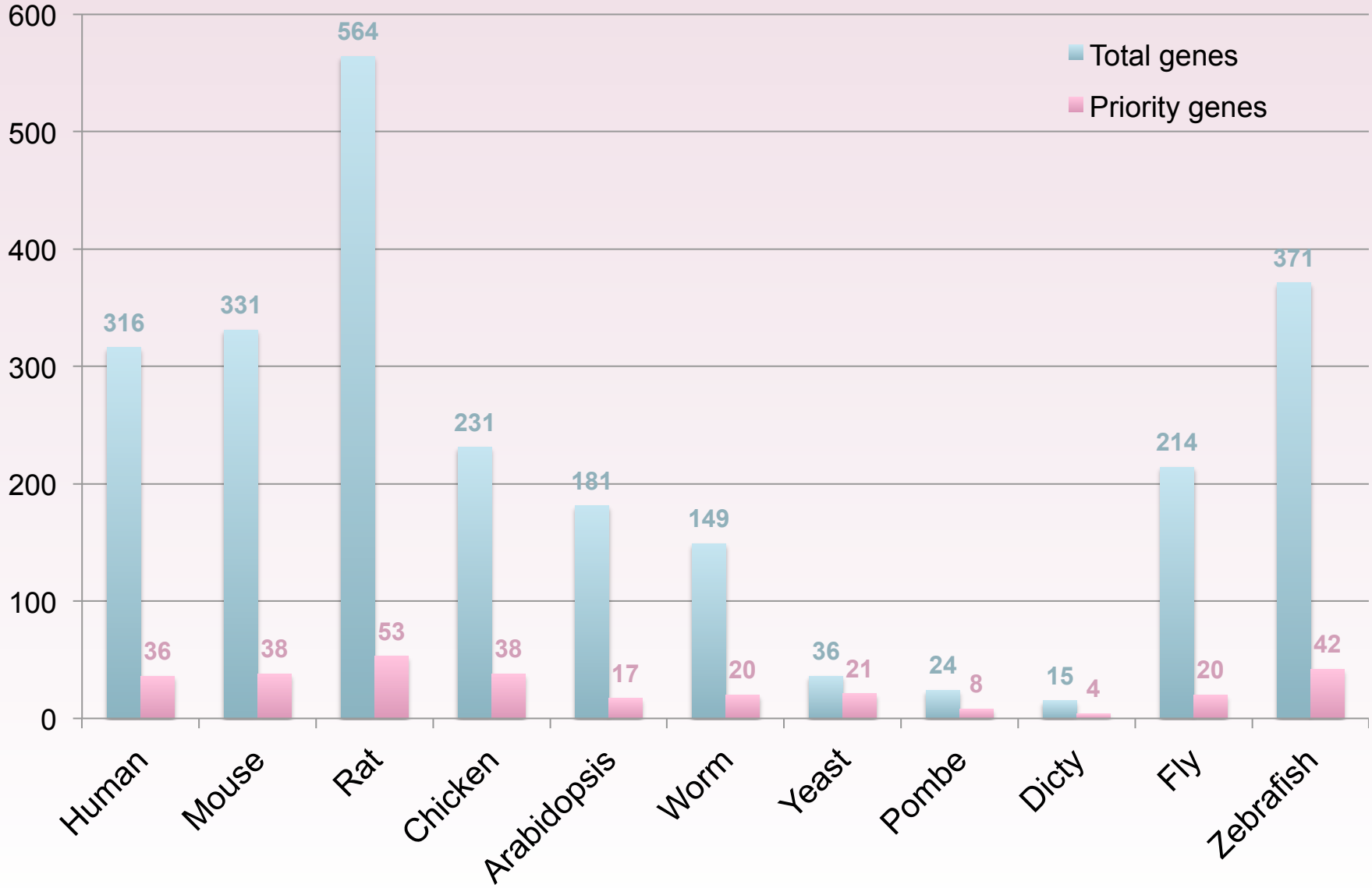
- Populate the heart development ontology with annotations

- Provide an opportunity for use of the new transcription ontology by Ref Genome annotation groups

Selection of annotation targets

- Original Gene list (60 human gene products)
  - Proposed by heart development experts
  - Majority of these may not be relevant to lower MOD curators
- Pared down original list to transcription factors only (36 human gene products)
  - Focus on transcription
  - Gene list included whole PANTHER families to ensure that lower organism MODs had targets to annotate
- Entire target list for whole families (316 human gene products)
  - Manually checked list for each MOD and identified high priority genes for annotation to assist curators with a large number of targets

Number of targets per MOD

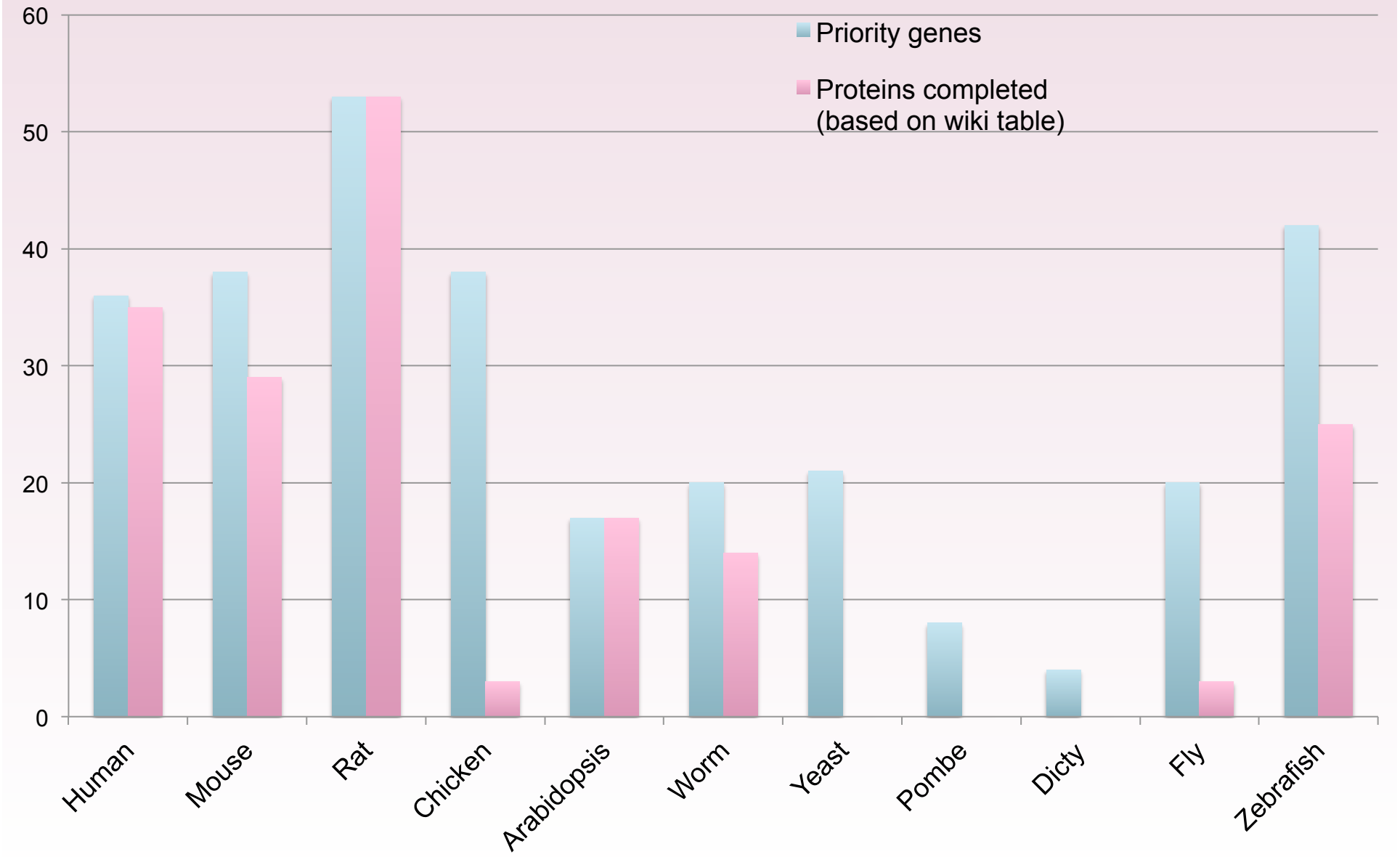


Wiki table to monitor progress (as at 2:30pm BST today)

UCL

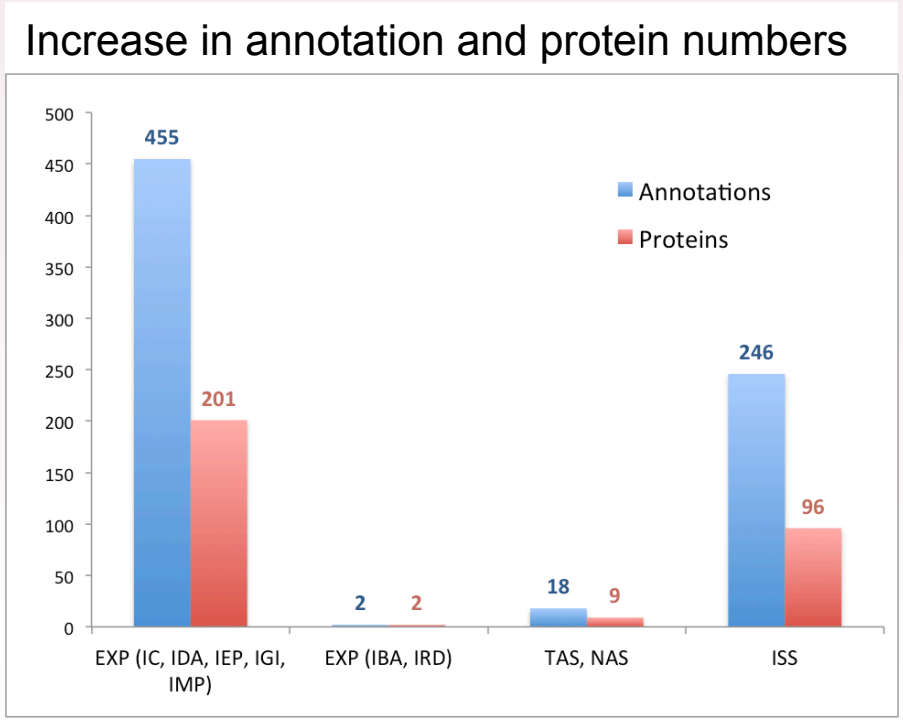
Annotation period (week beginning)	Panther Families	Gene Symbol of representative family members (Human protein ID)	PAINT due to begin	GOA / BHF-UCL	MGI	RGD	ZFIN	Wormbase	Flybase	Agbase	Dictybase	TAIR	SGD	Pombase
Week 1 (9th May 2011)	PTHR10071	GATA4 (P43694)		priority genes done	done	done	done	elt-1 - done	no priority genes		no priority genes	done	no priority genes	no priority genes
Week 2-3 (16th May 2011)	PTHR10032, PTHR24338	OVOL2 (Q9BRP0), MSX1 (P28360)		priority genes done	done	done	done	hbl-1 - done			no priority genes	no priority genes	no priority genes	no priority genes
Week 4 (31st May)	PTHR10270	SOX17 (Q9H6I2)		priority genes done	done	done	done	sox-3 - done	no priority genes		no priority genes	no priority genes	no priority genes	no priority genes
Week 5 (7th June)	PTHR10985, PTHR22793	HEY2 (Q9UBP5), MYCD (Q8IZQ8)		priority genes done	done	done	done	lin-22 - done			no priority genes	no priority genes	no priority genes	no priority genes
Week 6-7 (14th June)	PTHR11267, PTHR23349	TBX3 (O15119), HAND2 (P61296)		priority genes done	6/9 done	done	done	tbx-2 done			no priority genes	no priority genes	no priority genes	no priority genes
Week 8 (28th June)	PTHR11309, PTHR10812	SMO (Q99835), TFAP2A (P05549)		priority genes done	1/2 done	done	done	F28C6.1 - done, F28C6.2 - done			no priority genes	no priority genes	no priority genes	no priority genes
Week 9 (5th July)	PTHR11389	SNAI2 (O43623)		priority genes done	done	done	done	ces-1 - done				done		
Week 10-11 (12th July)	PTHR11945	MEF2C (Q06413)		priority genes done	done	done	done	mef-2 - done				no priority genes		
Week 12-13 (26th July)	PTHR11829, PTHR24340	FOXA2 (Q9Y261), NKX2-5 (P52952)		priority genes done	done	done		pha-4 - done		FOXA2 not a priority gene	no priority genes	no priority genes	no priority genes	no priority genes
Week 14 (2nd August)	PTHR12198, PTHR24126	PROX1 (Q92786), ANKR1 (Q15327)		priority genes done	1/2 done	done	done	ceh-26 - done			no priority genes	no priority genes	no priority genes	no priority genes
Week 15 (9th August)	PTHR12881, PTHR12958	MED1 (Q15648), FOG1 (Q8IX07)		FOG1 done	done	done	done	no priority genes		done	no priority genes	no priority genes	no priority genes	no priority genes
Week 16 (16th August)	PTHR13844, PTHR23341	SMARCD3 (Q6STE5), HMGA2 (P52926)		priority genes done	done	done	done	swsn-2.1 - done, swsn-2.2 - done, T24G10.2 - done	CG1240 done	SMARCD3 done		no priority genes		
Week 17 (23rd August)	PTHR13864, PTHR17045	TAL1 (P17542), CITED2(Q99967)		priority genes done	done	done		hlh-15 - done			no priority genes	no priority genes	no priority genes	no priority genes
Week 18 (30th August)	PTHR23043	HIF1A (Q16665)		priority genes done		done	done	no priority genes	no priority genes		no priority genes	no priority genes	no priority genes	no priority genes
Week 19 (6th September)	PTHR24204	ISL1 (P61371)		priority genes done	done	done		lim-7 - done	tup done		no priority genes	no priority genes	no priority genes	no priority genes
Week 20 (13th September 2011)	PTHR24329	PITX2 (Q99697)		priority genes done	done	done	done	unc-30 - done	Ptx1 done		no priority genes	no priority genes	no priority genes	no priority genes

% of priority genes annotated



Heart Development terms

- Ontology developed in Sept 2009 workshop
  - Expanded available terms from 12 to over 280
- Manual annotations to 'heart development' (or child terms)
  - Start of project (May 2011) **3574** annotations to **1655** proteins
  - End of project (Oct 2011) **4049** annotations to **1867** proteins



Transcription factor ontology

- Recent overhaul of terms by Karen and David
- Project enabled curators across multiple MODs to use and test the new ontology
- Electronic jamboree held as part of the Ref Genome project, to deal with transcription ontology queries

Electronic Jamboree discussion

- Some of the issues we discussed in the call:

- When do we have enough data for a MF annotation?
  - How to annotate a MF to repressing TF, as GO:0016564 transcription repressor activity is obsolete
  - Discussion ongoing between Ruth and Karen as to how best to capture this information

- How much curator judgment is appropriate?
  - Agreed curators can assume RNAPII is being used when TF is regulating mRNA encoding gene
  - Ambiguity in literature with regard to promoter and enhancer
  - Karen and David currently revising the enhancer terms to clarify the use of these terms
  - All definitions to be included in the transcription curation manual

Electronic Jamboree summary

- Asked curators to add examples to wiki page as they came across them
  - Proved a successful way of capturing issues as they arose
  - Wiki format allowed other curators to read relevant papers/SF items as needed prior to the call

- Last two issues did not get discussed fully due to lack of time
  - Perhaps schedule more than 1 hour for electronic jamborees, as these are not a regular annotation call and discussions can be quite involved

Limitations of the project

- Compliance with Ref Genome projects
  - Most MODs start off well, curators have good intentions
  - Lack of resources at each MOD
  - Lack of interest in project
    - Ref genome aims not directly relevant to MOD?
    - Selected gene families not high priority for MOD?
    - Other reasons?

- PAINT curation
  - Vital part of co-curation project aims
  - Lack of resources

Achievements

- Populate the heart development ontology with annotations
  - 721 annotations added to ***GO:0007507 heart development*** (or child terms) for 308 proteins across all species

- Provide an opportunity for use of the new transcription ontology by Ref Genome annotation groups
  - Ongoing refinement of transcription ontology (as with whole ontology) as curators annotate papers
  - Electronic jamboree provided Ref Genome groups the opportunity to focus on the transcription ontology

- Additional outcomes
  - 179 proteins fully annotated across all species (from 279 proteins on the priority list)

Possible future directions

- Increase number of annotations to heart development terms
  - Identify other heart development genes
- Continue to hold electronic jamborees on specific topics or areas of the ontology
- PAINT curation of Ref Genome project genes