Steps to find pig human homology

**all the files have as first column the pig probe ID and the second column the matching human probe ID**

1. Use document "Of Pig and Men/platforms/AnnotationByIPA.xlsx"

downloaded from: http://www.animalgenome.org/pig/projects/array\_annotatn

citation: Couture et al Mammalian Genome 2009 20: 768-77 PMID 19936830.

- this contains pig microarrays blasted against IPA gene clusters. It match pig probe IDs to Entrez Gene IDs. (http://www.ingenuity.com/products/ipa)

2. Use script

/Users/nemquae/Of Pig and Men/scripts/matchPigGeneIDToHumanMultiples.pl

which matches the Entrez gene IDs given by the file in 1 to ensemble transcript IDs for pig, then to ensemble transcript IDs for human, then to the human microarray probe IDs.

* this also depends on the database we use: currently I use Ensemble Genes 73
* downloaded November 7, 2013
* human: Homo sapiens genes (GRCH37.p12)
* pig: Sus scrofa genes (Sscrofa10.2)
* and only the orthologous genes, which gives me 19279/25322 genes in pigs have human ortholog
* 17210/63605 human genes have pig ortholog
* but if we extend it to possible orthologs: 2625/25322 genes pigs
* possible orthologs in humans: 2880/63605 genes
* this workflow is also used for the agilent porcine and affymetrix porcine. In the case of the affymetrix porcine, porcine.db is an R package that gives the entrez gene ID for the probes. In the case of the Agilent porcine, I used the annotate porcine bioconductor package (org.Ss.eg.db)to translate genbank accession numbers to Entrez IDs, but only about 1/8 of them had a gene assocaited with them, probably because they were mostly ESTs from putative mRNAs that were not yet curated and named as a gene

getUnique.pl (to get unique pig to human id’s) l

I put these files in the “Matches” folder

3. If the Entrez Gene ID for the pig probe is not from pig, then query homologene database for the entrez gene ID and take the one in the homologene cluster that is human

scripts:

geneToHomoToGeneUID.pl <trialGeneID.txt (which is all the Entrez Gene IDs of the first file)

* homologene\_gene.idx
* homologene\_gene\_search.idx
* geneToHomoToGeneUID.txt

-the script queries the ncbi server, and I ran it on Nov 15,2013

fromPigGeneIDtoHumanGeneIDthruHomologene.pl

- though apparently if we restrict these to only the human genes that are orthologous gene to pig, we get 1/10 as many matching human probes as if we let any probes match

- I put these files in the “Weaker matches” folder

Relevant documentations on NCBI

Ensemble documentation of how homology is found for the biomart downloaded pig to human gene mappign:

<http://asia.ensembl.org/info/genome/compara/homology_method.html>

homologene build procedure

<http://www.ncbi.nlm.nih.gov/homologene/build-procedure/>

I think it’s release 67 that I used