

Work Update

Caroline Monteiro

Ma'ayan Lab

Nov-2016

RNA-Seq

Mouse

	GENE	Drug	Disease	Total
Total accessions	511	192	131	834
Unique accession	268	74	26	368
Samples Control	1917	646	656	3219
Unique Control	1130	393	196	1719
Samples Condition	1702	339	464	2505
Unique Condition	1608	333	463	2404
Total samples				5724
Total Unique samples				4123

Rat

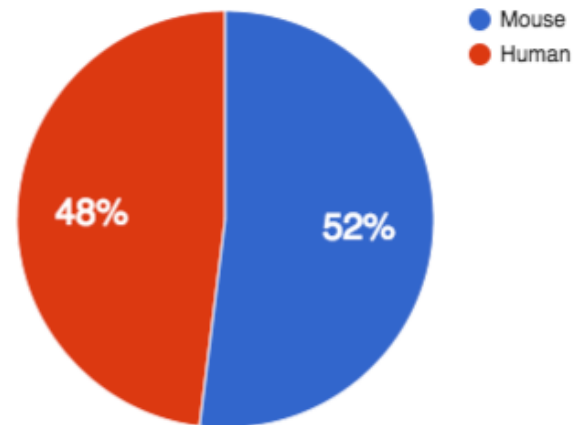
	GENE	Drug	Disease	Total
Total accessions	13	46	4	63
Unique accession	5	15	1	21
Samples Control	30	41	38	109
Unique Control	30	41	38	109
Samples Condition	30	41	38	109
Unique Condition	30	41	38	109
Total samples				218
Total Unique samples				218

Current Tally:

Accessions: 897
Samples: 5942

Platforms

GPL13112	Illumina HiSeq 2000	55317	Mus musculus
GPL17021	Illumina HiSeq 2500	23879	Mus musculus
GPL16791	Illumina HiSeq 2500	25273	Homo sapiens
GPL11154	Illumina HiSeq 2000	47941	Homo sapiens
	Total	152410	



Total : 158303

GSM = 157881

Unique GSM = 121627

Duplicates = 36 675 (Sample runs)

Gen3va report - IFNgamma

Total signatures: 28

Unique accessions: 13

Tag:

<http://amp.pharm.mssm.edu/gen3va/tag/IFNgamma>

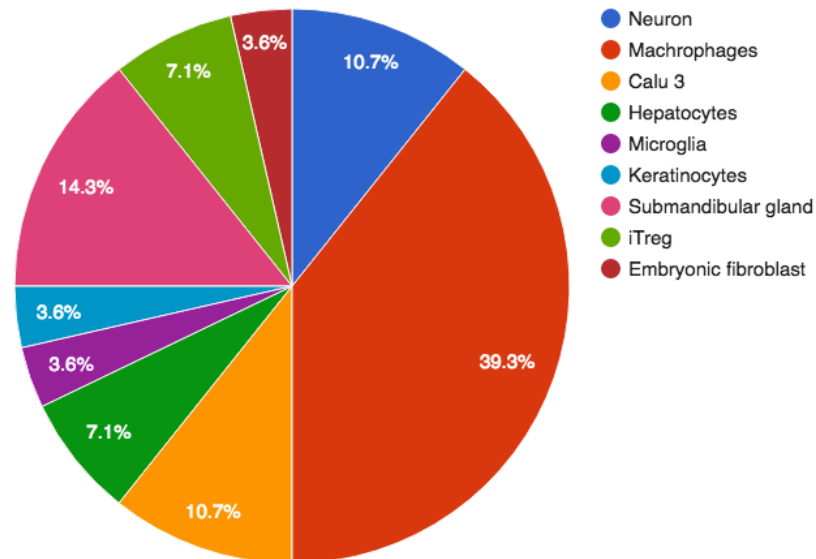
General Report:

<http://amp.pharm.mssm.edu/gen3va/report/IFNgamma>

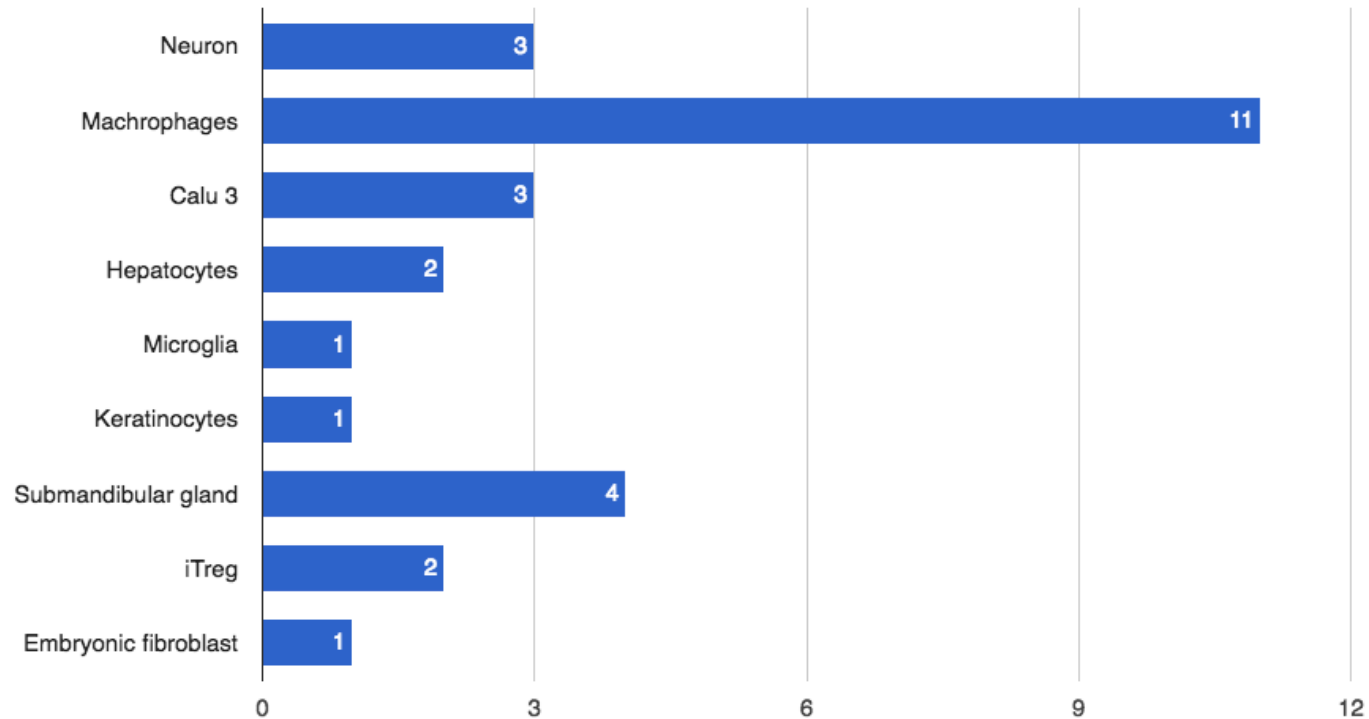
Report by cell/tissue:

<http://amp.pharm.mssm.edu/gen3va/report/291/IFNgamma>

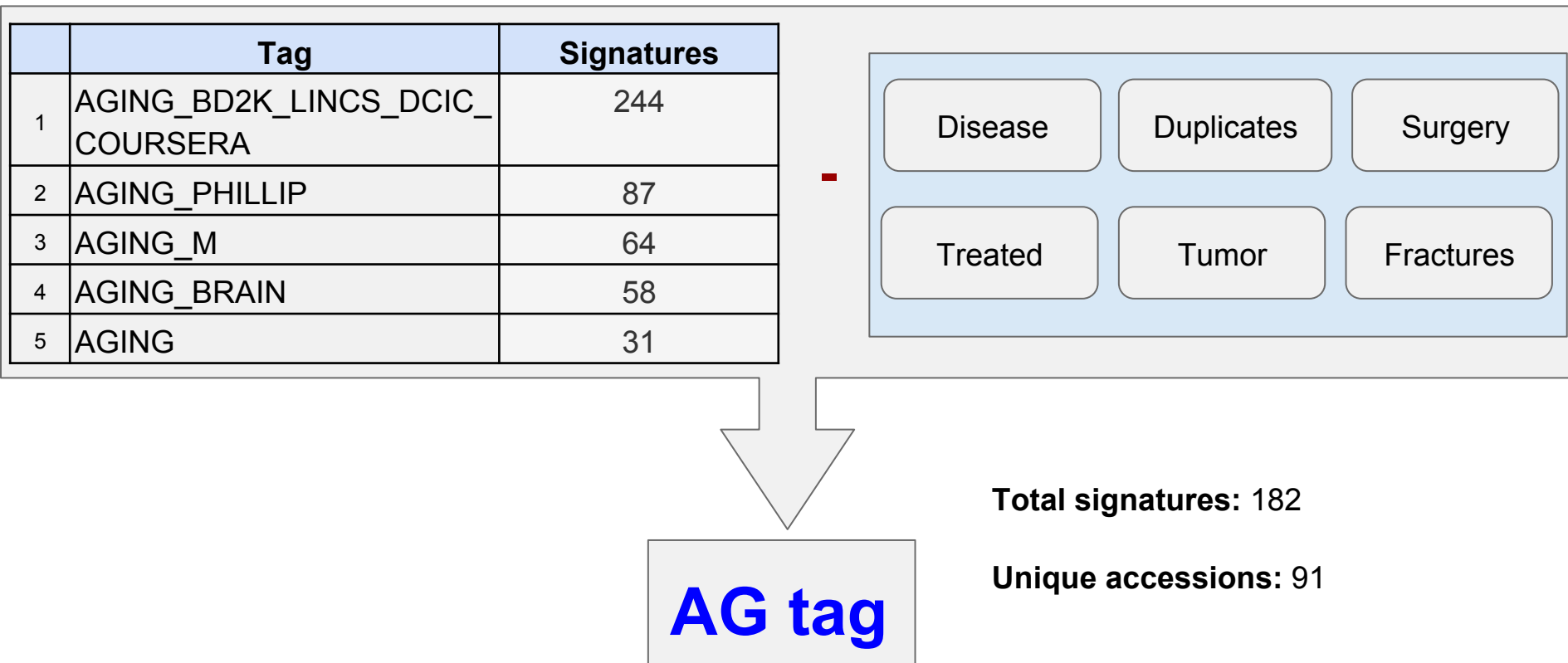
Sample distribution per tissue

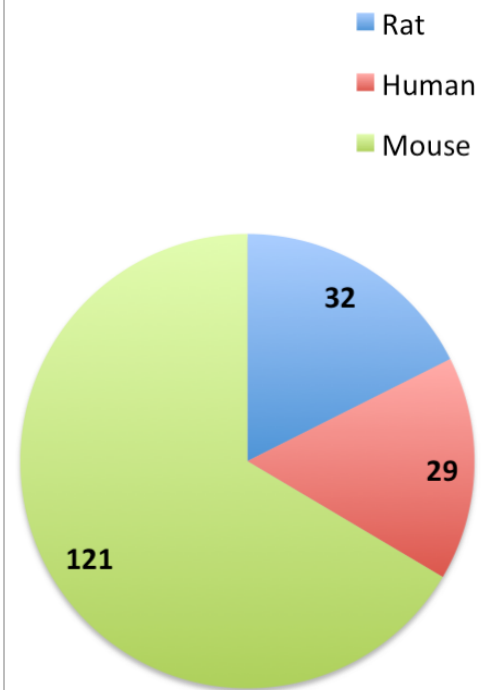
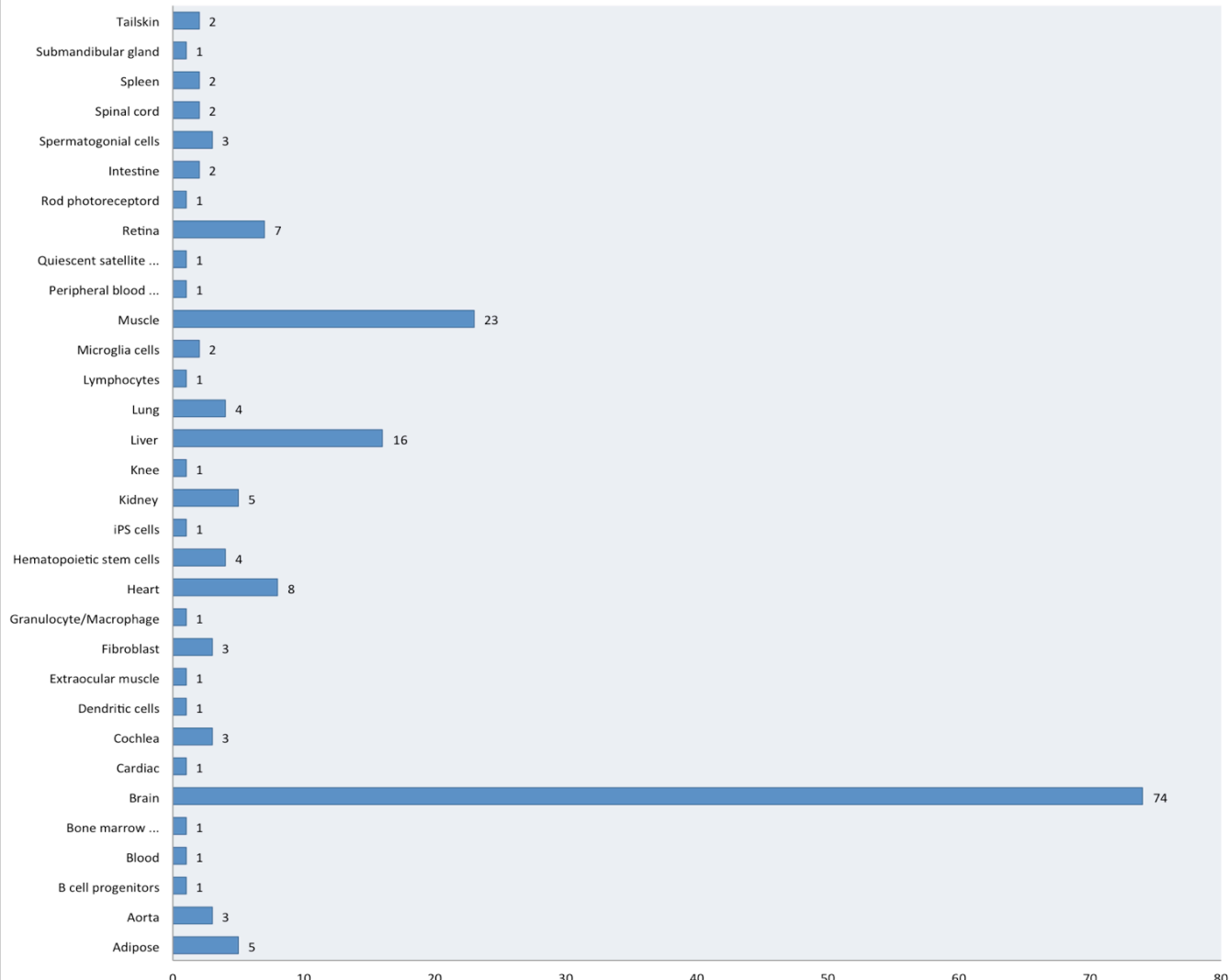


Tissue distribution:



Aging tags





ChEA Update

Large-Scale Quality Analysis of Published ChIP-seq Data

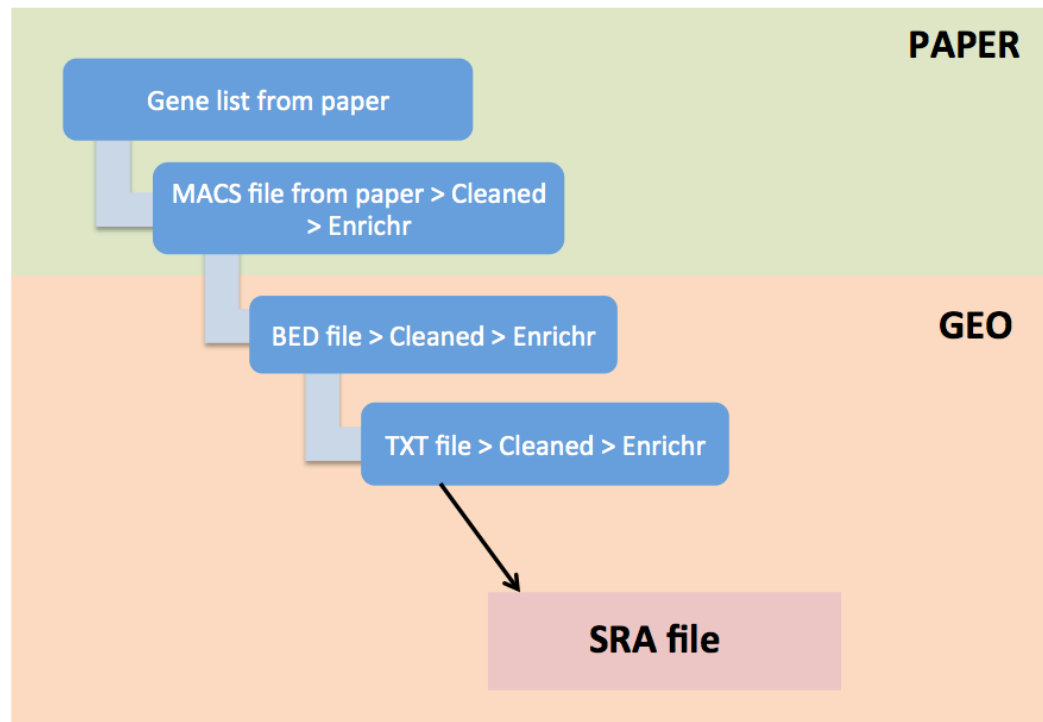
Georgi K. Marinov^{*}, Anshul Kundaje^{**,††,1}, Peter J. Park^{†‡§} and Barbara J. Wold^{*,2}

Not added:

- Histones
- DNA polimerase
- Drug treated samples
- Immunoglobulins
- Multiple replicates
- Different organisms

> 900 Chip-seq samples

													Available files									
Link to source	Gene list origin	Source	Org	Samples in paper	PMID	Accession	Sample	Platform	TF	Type	Should exhibit read clustering	Cell/ tissue	txt link	bed file	sra link	enrichr link						
https://www.ncbi.nlm.nih.gov/papers/	bed file from geo paper	Lister	human	7	19829295	GSE17917	GSM447584	GPL9052; GPL9115	KLF4	ChIP	yes	Embryonic stem cells	ftp://ftp.ncbi.nlm.nih.gov/pub/	ftp://ftp.ncbi.nlm.nih.gov/pub/	ftp://ftp-tr.ncbi.nlm.nih.gov/pub/	http://amp.ensembl.org/	2000 genes	HEXIM2	BTBD19	ESX1	BTBD16	C20ORF194
																	From paper 2000 genes	HEXIM2	BAD	CACNA2D2	C20ORF191	CACNA2D3



Multiple pubmed IDs
Platform >> sample platform

Extras

Last check : Nov - 11

Lincsproject.org

- Links
- Tools
- Center pages
- Workflows
- Publications
- Marketplace links

- Gen3va

- GEO2Enrichr

- L1000CDS2

- Links

-

Functions

Data mining course

- Syllabus Needs update
- “Box of goodies”

Here is the new version of the Chip-seq update.

<https://docs.google.com/spreadsheets/d/1oF-nq9IXcZKomwySLmIxTJ4TrLUCrOIQ4RRibP0XJU/edit?usp=sharing>

Adding both tabs (phase1 & phase 2) there are a total of **517 entries** and **247 different TF targets**.
I was in doubt about a few (and didn't add them) so this number might change slightly.

Explanations:

First in the tab named **Phase1**: Are all the entries from the previous table, but the matrix has been transposed. This table is color coded for the chip-seq studies with 2000 gene lists, shorter gene lists and those with only sra files.

Next in the tab named **Phase2**: Are the chip-seq samples from the paper.

OBS

1471-2164-12-11-S2.TXT

Additional File 2. Common genes from the two SOX2 Chip_seq datasets from mouse ES cells*

mouse gene name	human homologue
1-Sep	1-Sep
A2bp1	A2BP1
A4galt	A4GALT
Abca4	ABCA4
Abcc4	ABCC4
Abcd3	ABCD3
Abcg1	ABCG1
Abcg2	ABCG2
Acp5	ACP5
Acp6	ACP6
Acss1	ACSS1
Actn1	ACTN1
Acvr1b	ACVR1B
Adipor2	ADIPOR2
Adrb3	ADRB3
Aff3	AFF3
Aftph	AFTPH
Agpat5	AGPAT5
Ahcy	AHCY
Ahsa1	AHSA1
Akap1	AKAP1
Akap12	AKAP12
Aldh1a2	ALDH1A2
Alg8	ALG8
Alkbh6	ALKBH6
Alpk3	ALPK3
.....

ABout the CHEA

Missing org in a few