

# Pseudogenes in the ENCODE

Baikang Pei  
05/03/2011

# Pseudogenes in GENCODE Annotation File

Pseudogenes are divided into three categories in gencode v3c annotation file:

- Level 1: 3 way consensus of Yale, UCSC and HAVANA annotations;
- Level 2: manually curated by HAVANA;
- Level 3: predicted by ENSEMBL pipeline

Level 1 (3171)

processed_pseudogene	3018
unprocessed_pseudogene	65
transcribed_processed_pseudogene	12
transcribed_unprocessed_pseudogene	2
unitary_pseudogene	2
pseudogene	24
IG_pseudogene	48

Level 2 (5661)

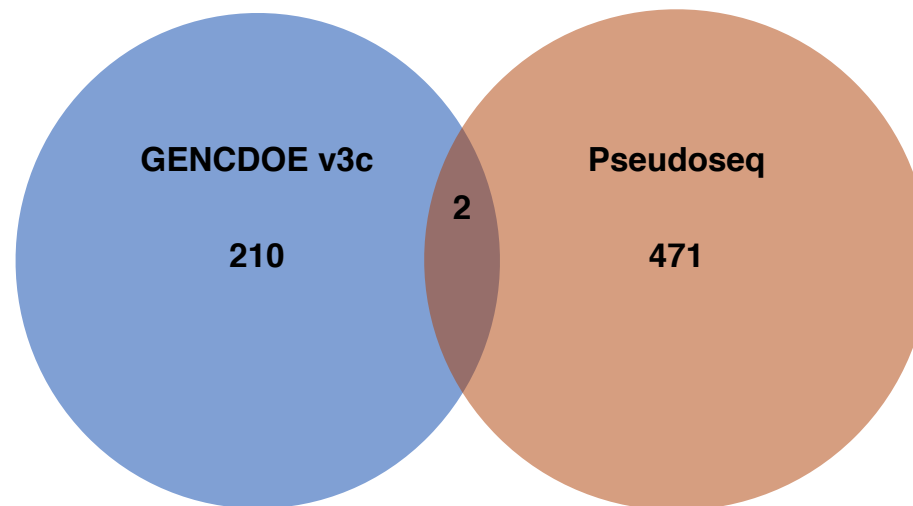
processed_pseudogene	3211
unprocessed_pseudogene	1201
transcribed_processed_pseudogene	50
transcribed_unprocessed_pseudogene	146
unitary_pseudogene	120
pseudogene	609
IG_pseudogene	112
polymorphic_pseudogene	20
TR_pseudogene	19
retained_intron	2
retrotransposed	5
processed_transcript	166

Level 3 (3610)

processed_pseudogene	1
unprocessed_pseudogene	4
pseudogene	601
retrotransposed	114
tRNA_pseudogene	128
miRNA_pseudogene	20
snoRNA_pseudogene	484
rRNA_pseudogene	338
snRNA_pseudogene	494
misc_RNA_pseudogene	7
Mt_tRNA_pseudogene	580
scRNA_pseudogene	839

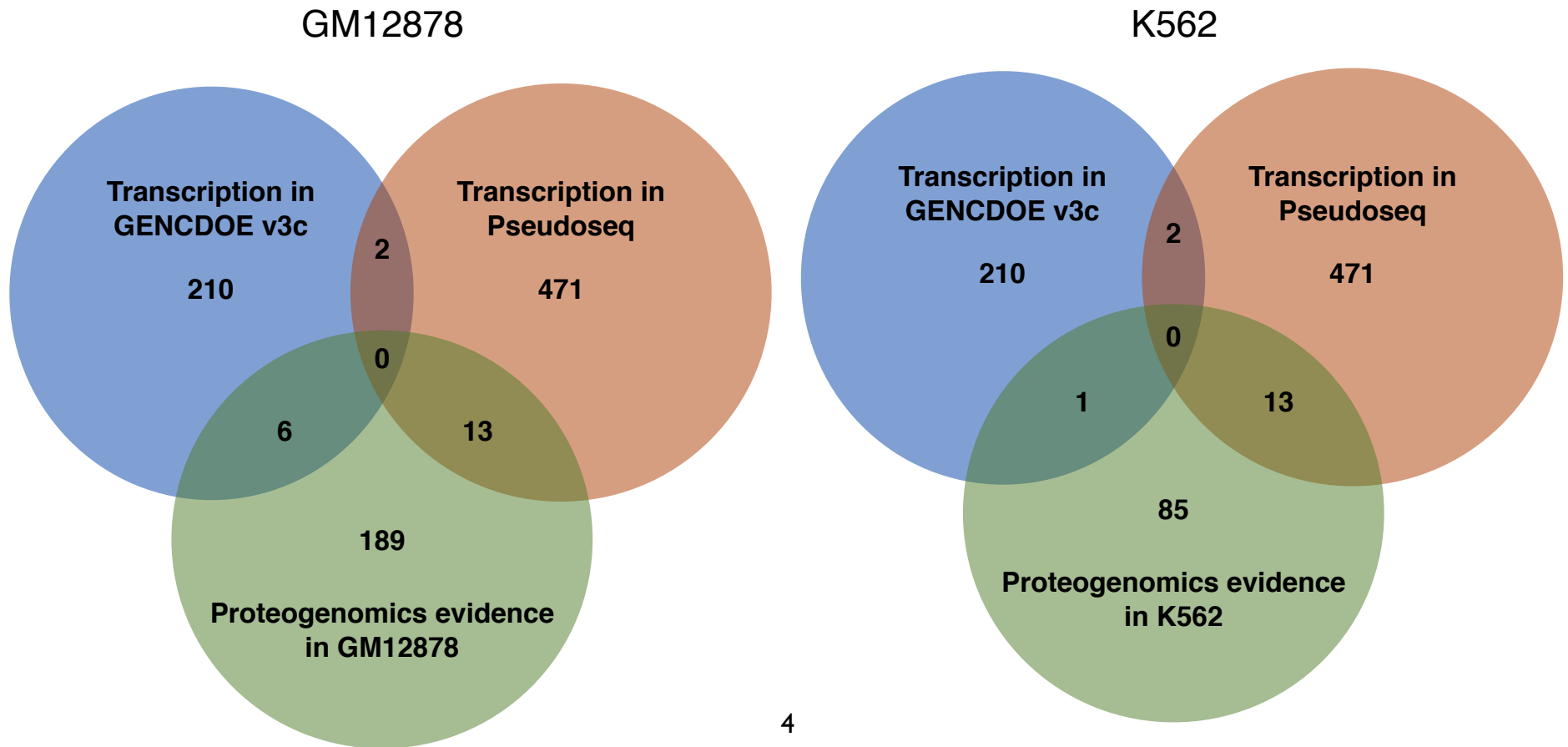
# Transcribed Pseudogenes

1. Transcribed pseudogenes from GENCODE annotation: totally 210 pseudogenes.
2. Transcribed pseudogenes from Pseudoseq using BodyMap data, lifteOver to Hg19. Totally 471 pseudogenes.



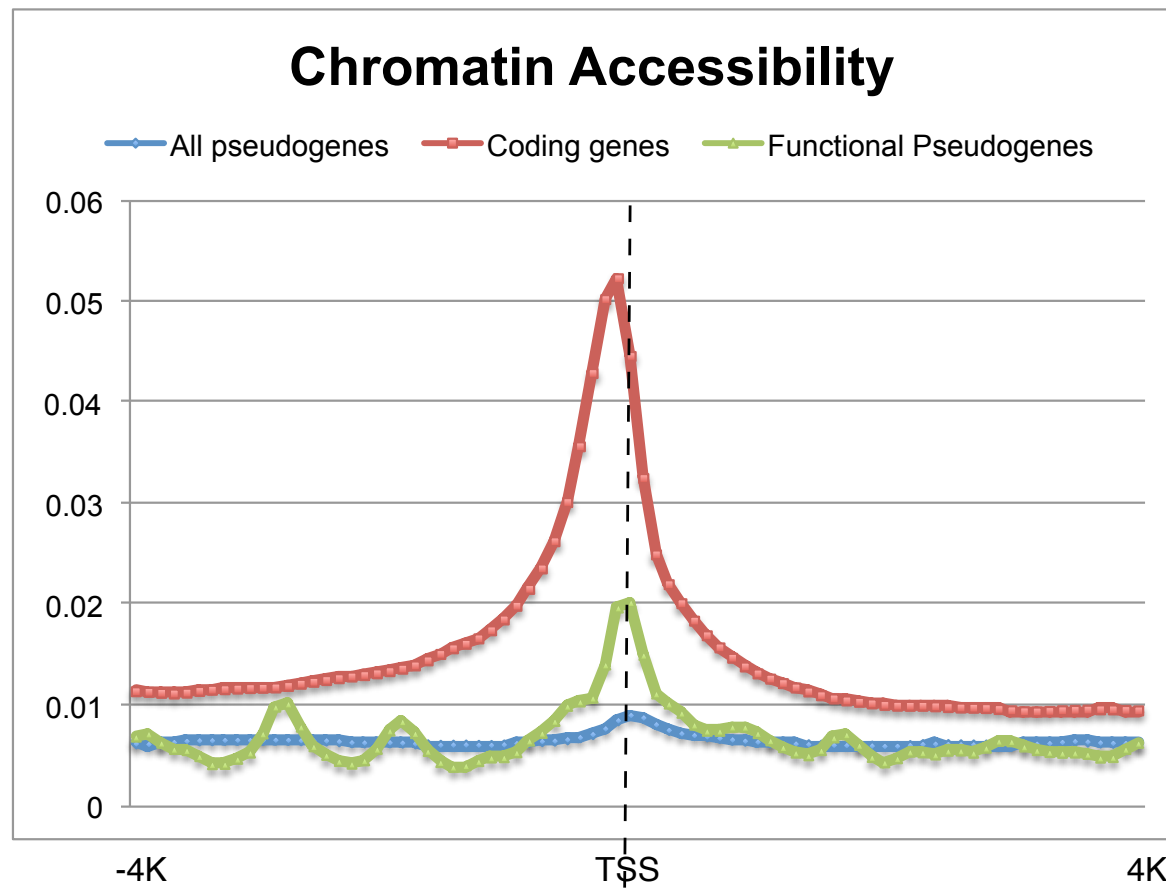
# “Translated” Pseudogenes

Proteogenomics data: from Morgan Gidding’s lab for nuclear fraction of GM12878 and cytosol and membrane of K562.



# Chromatin Accessibility

DnaseI Hyper-sensitivity data on GM12878 by Duke University



# Personal Variants (GM12878) in Pseudogenes

	For all variants	SNPs	Indels	Deletions
<b>Reference</b>	8832	8832	8832	8832
<b>(M)aternal</b>	2688	2552	597	20
<b>(P)aternal</b>	2647	2524	595	14
<b>Both</b>	1936	1822	432	10

# To Do...

- Personal genomic variants vs. pseudogene expression;
- TF, polymerase and/or histone markers associated with open chromatin regions;
- Activity with duplication;
- Selection;
- conservation;