

Human:
Looking for translation of
pseudogenes

Yan Zhang

20130924

Results and Examples from the **Relaxed List**

(pgene_unique_sorted.xlsx)

- 9882 individual MS hits (events) uniquely match to pseudogenes but not known proteins or splice variants.
- Among which 1430 annotated as transcribed.

Retired
example?

Example 1:

TranscriptID	FilePath	peptideSequence	previous	start	stop	SequenceName	RankCount	Type	Transcribed
ENST00000531139	A_MCF7_01.mgf	AAGTCPAAAR	M	1	11	Human_pgene_chr11_ORF3_ENST00000531139_plus_94800389_94801232_Pgene	1	PSSD	TRANS_YES

- AAGTCPAAAR , previousAA: M
 - Detected in one cell line
 - Matches to pgenes but not known proteins or splice variants
 - Pgene: ENST00000531139, was Transcribed PSSD; retired? (ensembl)
 - Parent gene: ENST00000358156

Exons

MSYGRPPPDVEGMTSLKVDNLTyrTSPDTLRRVFEKYGRVGDVYIPRDRYTKESRGFAFV
RFHDKRDAEDAMDAMDGAVLDGRELRVQMARYGRPPDSHHSRRGPPPRRYGGGGYGRRSR
SRSRSRSRSRSR

Pgene AA sequence has no overlap at all with the parent gene.

Example 2:

Possibly good example

- AAQLALR, previousAA: R
 - Detected in multiple cell lines multiple replicates
 - Matches to pgenes but not known proteins or splice variants
 - Pgene: ENST00000469993, known as PSSD, although TRANS_NO (what cell line?)
 - Parent gene: ENST00000273223

Exons Alternating exons Alternating exons Residue overlap splice site

MAALRPLVKPKIVKKRTKKFIRHQSDRYVKIKEEFALCTDANPPVYYFLQRNWRKPRGID
NRVRRRFKQILMPNIGYGSNKKTKHMLPSGFRKFLVHNVKELEVLLMCNKSYCAEIAHN
VSSKNRKAIVERAAQLAIRVTNPNARLRSEENE

Results and Examples from the **Strict List**

(pgene_unique_strict_sorted.xlsx)

- More strict restrictions:
 - Number of events detected (at least two)
 - Must be TRANS_YES (but it may be different for different cell lines)

 303 unique peptide sequences left

Good example

Example 3 from a more strict list

Might be cell line-specific

- ALFWNEEIVPQIK, previousAA: R
 - Detected in 3 replicates of HepG2
 - Matches to pgenes but not known proteins or splice variants
 - Pgene: ENST00000526979, transcribed PSSD
 - Parent gene: ENST00000334828

Exons

Alternating exons

Alternating exons

Residue overlap splice site

```
MAAYKLV LIRHGESA WNLNRFSGWYDADLSPAGHEEAKRGGQALRDAGYEFDICFTSVQ
KRAIRTLWTVLDAIDQMWLPVVRTWRLNERHYGGLTGLNKAETA AKHGEAQVKIWRRSYD
VPPPPMEPDHFPYSNISKDRRYADLTEDQLPSCESLKDTIARALPFWNEEIVPQIKEGKR
VLIAAHGNSLRGIVKHLEGLSEEAIMELNLP TGIPIVYELDKNLKPIKPMQFLGDEETVR
KAMEAVAAQ GKAKK
```

Example 4 from a more strict list

- FRVSPDLLWR, previousAA: R
 - Detected multiple times
 - Matches to pgenes but not known proteins or splice variants
 - Pgene: ENST00000521412, transcribed, PSSD
(different annotation: ensembl - transcribed unprocessed)
 - Parent gene: ENST00000316519

Blastp found no significant sequence similarity

Exons

Alternating exons

Alternating exons

Residue overlap splice site

```
MSKEERPGREEILECQVMWEPDSKKNTQMDRFRAAVGAACGLALESYDDLYHWSVESYSYD
FWAEFWKFSGIVFSRVYDEVVDTSKGIADVPEWFKGSRLNYAENLLRHKENDRVALYIAR
EGKEEIVKVTFEELRQEVALFAAAMRKMVGKKGDRVVGYPNSEHAVEAMLAAASIGAIW
SSTSPDFGVNGVLDLDRFSQIQPKLIFSVEAVVYNGKEHNHMEKLQQVVKGLPDLKKVVVIP
YVSSRENIDLKIPNSVFLDDFLATGTSEQAPQLEFEQLPFSHPLFIMFSSGTTGAPKCM
VHSAGGTLIQHLKEHLLHGNMTSSDILLCYTTVGWMMWNMVSLLATGAAMVLYDGSPLV
PTPNVLWDLVDRIGITVLVTGAKWLSVLEEKAMKPVETHSLQMLHTILSTGSPLKAQSYE
YVYRCIKSSILLGSISGGTDIIISCFMGHNFSLPVYKGEIQARNLGMAVEAWNEEGKAVWG
ESCFI VCFKBTDCQDTHFWNDENGNKYRKA VESKEDCTWAHGDVCFDTRDKTGCTVMTGRS
```