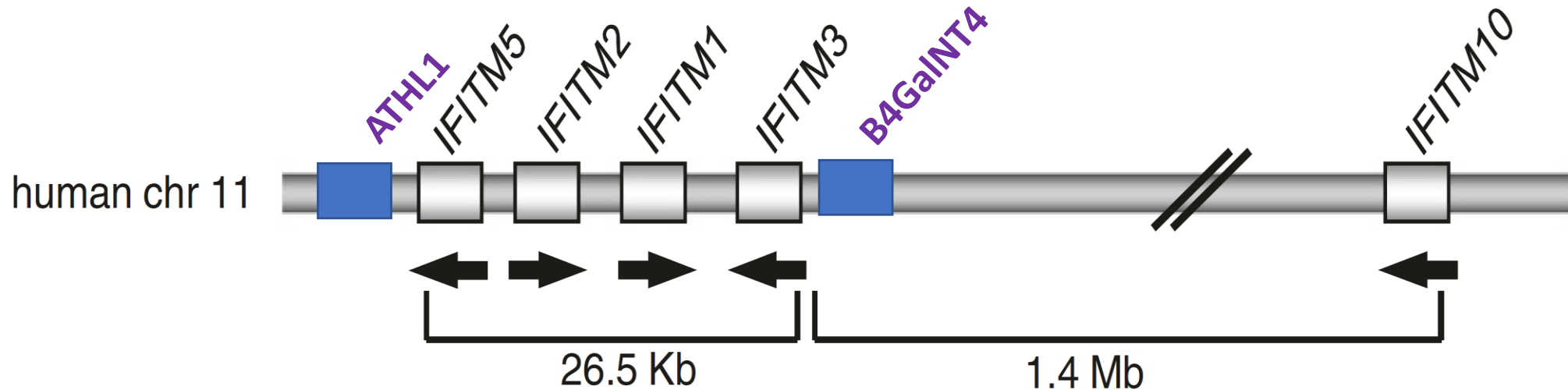


IFITM3 Retrogenes with Intact ORF: Are They Functional?

Kazi Rahman, PhD
National Cancer Institute
03/02/2018

IFITMs : Genomic and Syntenic Context

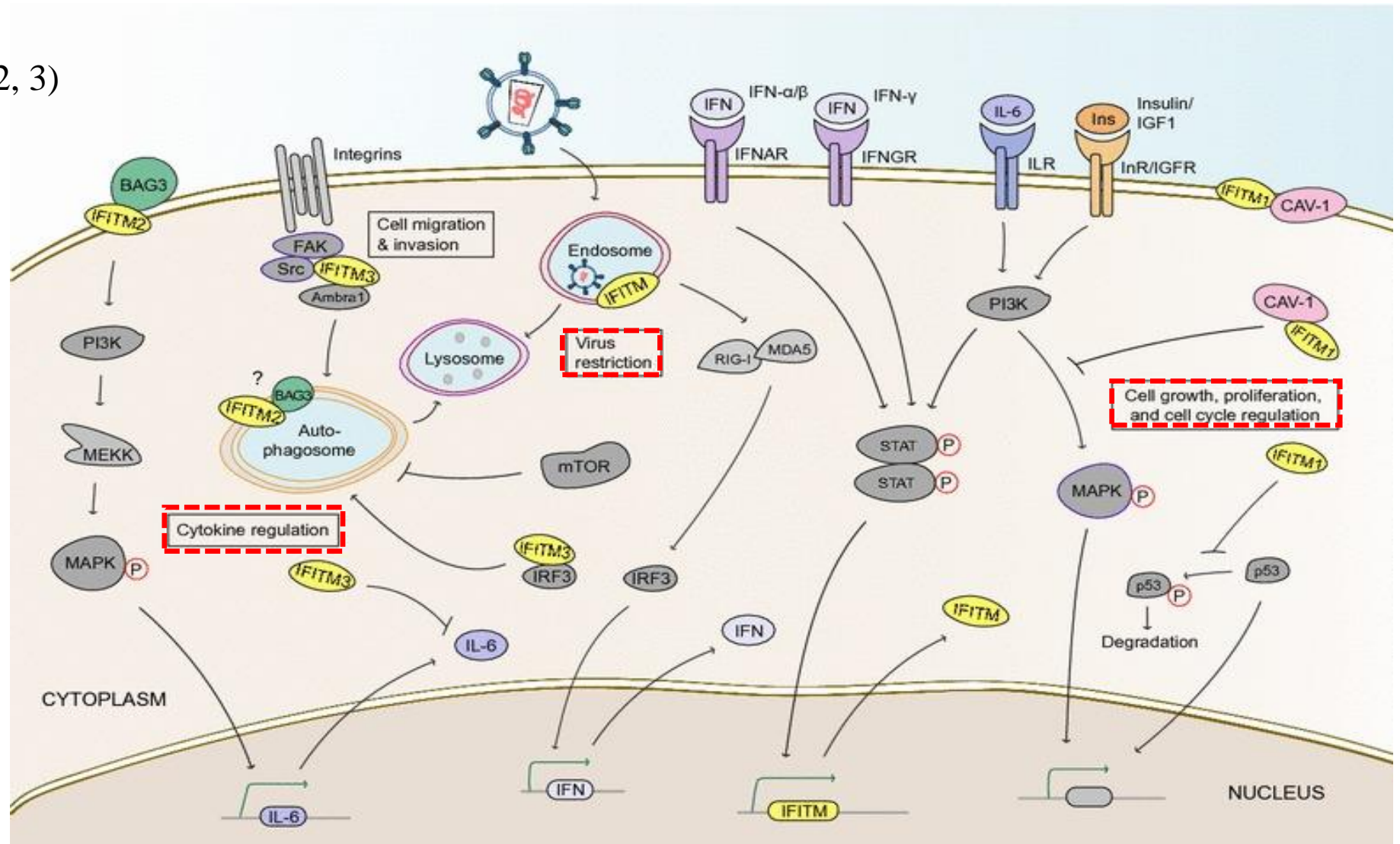
- Five different types: IFITM1, 2, 3, 5 and 10
- All are in Chromosome 11
- IFITM1, 2, and 3 are interferon-inducible and flanked by ATHL1 and B4GalNT4 genes



Role of IFITMs in Various Cellular Processes

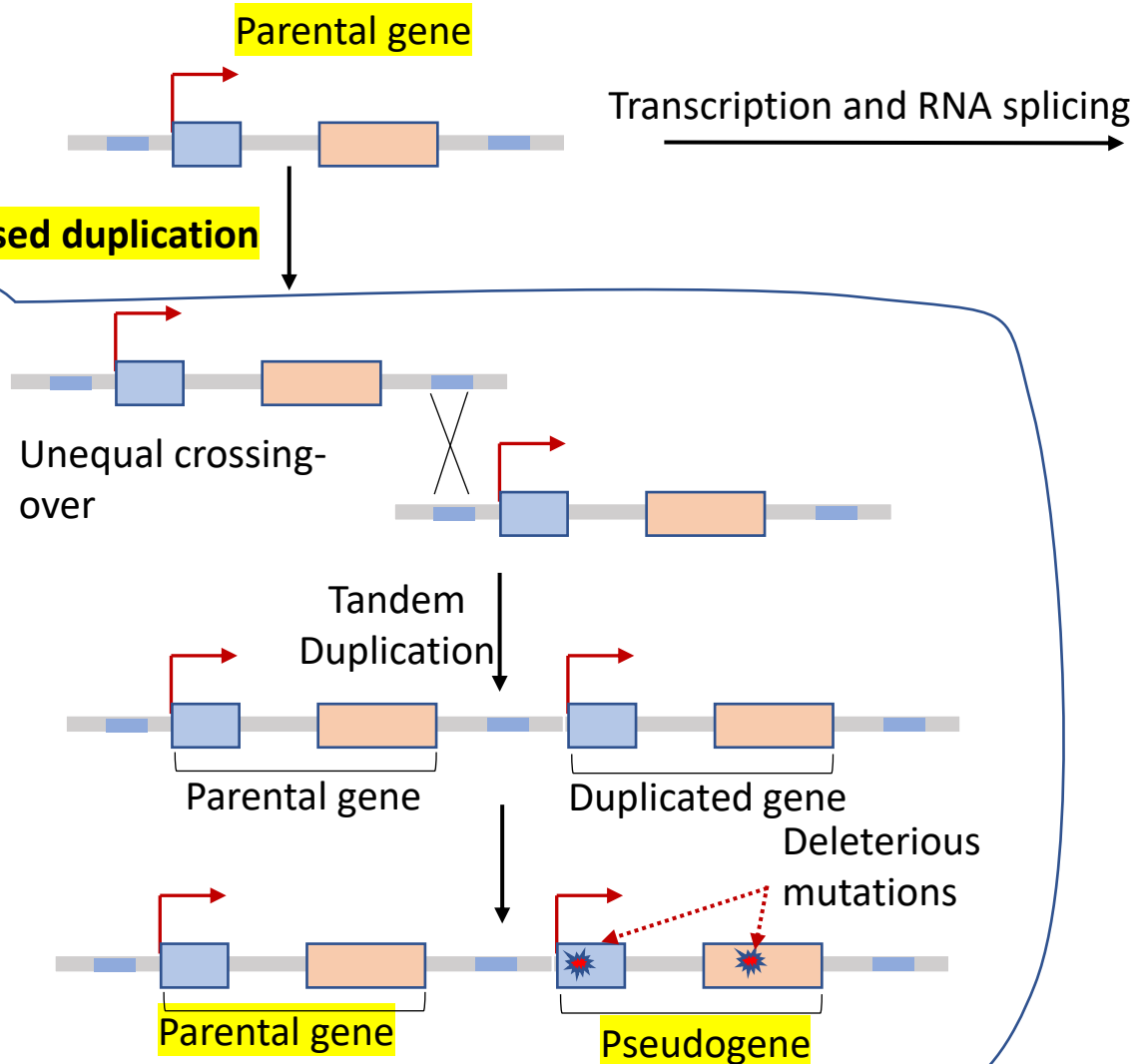
Involved in

- A. Virus Restriction (IFITM1, 2, 3)
- B. Cell Adhesion
- C. Anti-Proliferation
- D. Tumor Suppression
- E. Germ cell and embryo development

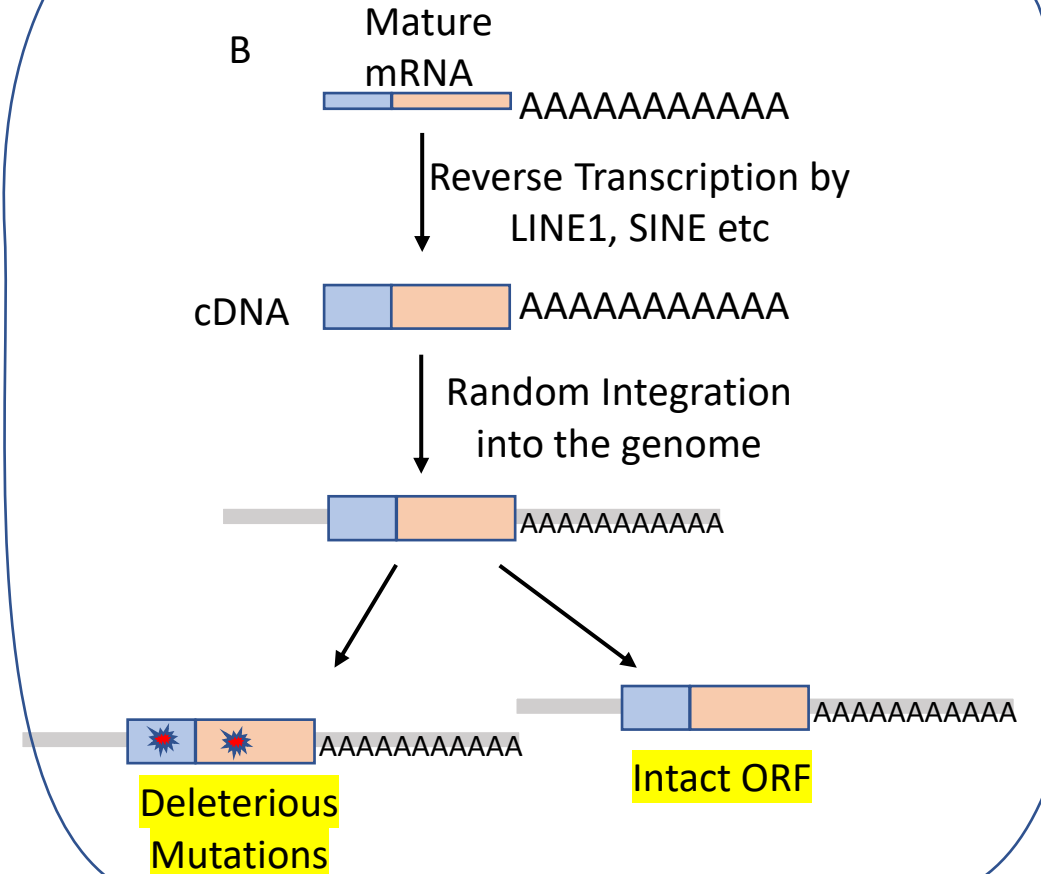


Origin of New Gene Copies through Gene Duplication

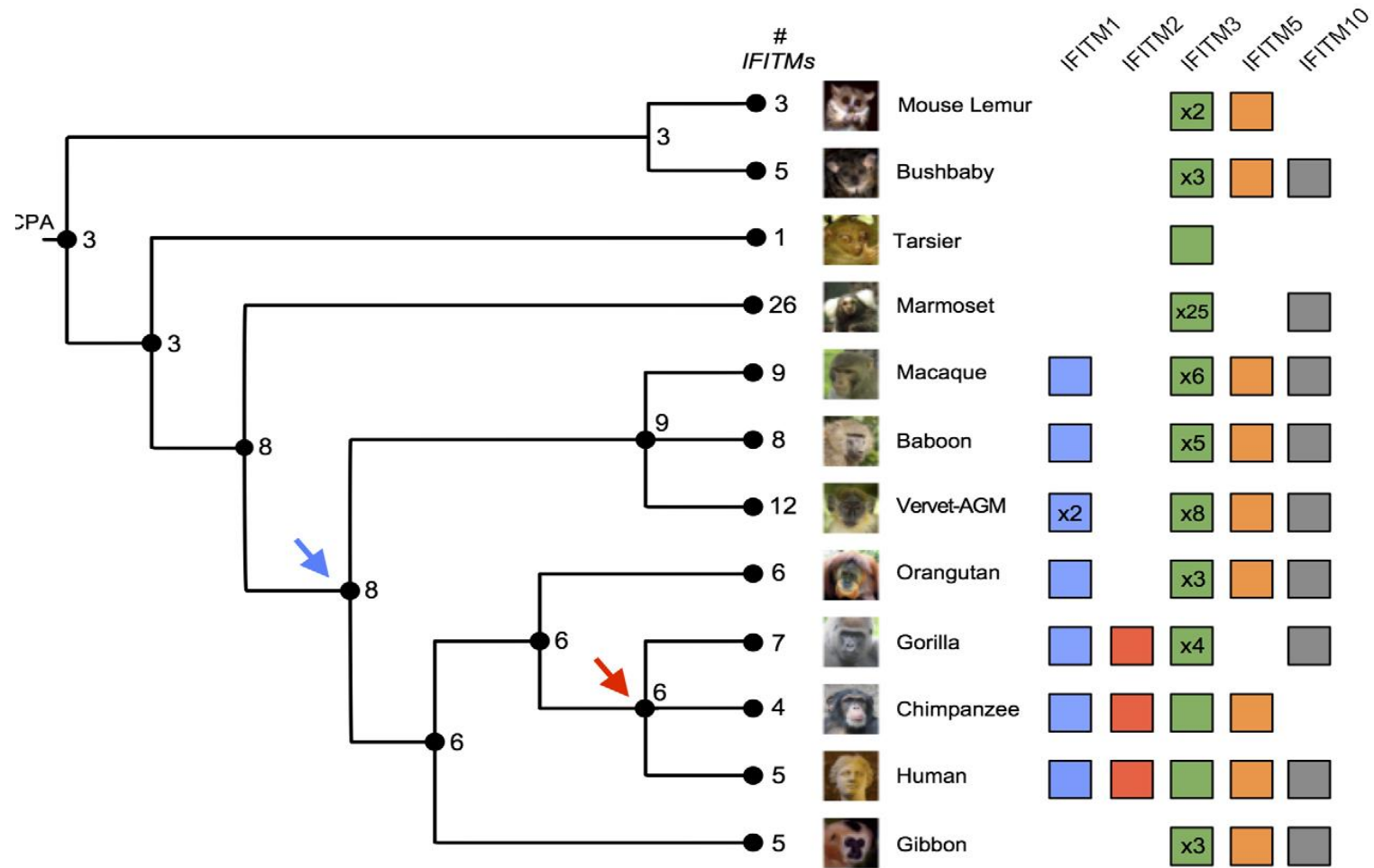
A. DNA-based duplication



B. RNA-based duplication



Recurrent Duplication of IFITM3 in Primates



Questions

Are those IFITM3 retrogenes with intact ORF functional?

- A. Do they undergo transcription and translation?
- B. What is their spatio-temporal expression pattern (i.e. specific cells/tissues and timing of expression)?
- C. Do they play redundant Roles of IFITM3 or gained neofunctionalism?

Investigating the function of human IFITM3 retrogenes.

IFITM Duplicants in Human

- Total 15 IFITM duplicants in Human
- All are retrogenes –intronless/ PolyA-tailed

| | | | | | | | | | | |
|--------|---|----|----|----|----|----|----|----|----|-----|
| | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
| IFITM2 | | | | | | | | | | |
| IFITM3 | | | | | | | | | | |

| | | |
|-----------------------|-----------|--|
| Intact ORF (2) | IFITM8P2 | MNHTFQTFFTPANSCHPBNYEMLKEEHKVAVVGAPQNPAPPMSTMIHFHSETSMPPDHVVVWSLFNTLFMNSGCLCFTAFVYSMKSRERKKVGDLTGAQAYT |
| | IFITM12P1 | MNHTVQTFFFSPVNSGQPPNYEMLKEEHKVAVLGVPHNPAPPTSTVIHRSKTSVPHHVVWSLFNTLFMNPCCLGFI AFAYSVKSRDRKMVGNDVTGAQAYA |

Deleterious Mutations (13)

| | |
|-----------|---|
| IFITM1P1 | KNHTSKSLFTHANTGLPTIEMFKEKHEVAMLGAPHSSAPLMSSVIHICSETSMPPDHVVVWSLFNTLFMNSGCLCFTAFVYSMKSRERKKVGDLTGAQAYT |
| IFITM1P2 | MNHTAQTFFTPANTGCPLDYQMLIHEVAVLGVPHNSAPPRSTVIHHSSETSVPHHVVWSLFNTLFMNSGCLCFTAFVYSMKSRDRKMVGNDLTGGQACA |
| IFITM2P | MNHAVQTVFTPANTSHPLNYEMLKEEHEVAVLRVPHNPVPTSTVIHHSSETTVPSHVVWSLFNTLFMNSGCLCFTAFVYSMKSRDRKMVGNDLTGAQAYA |
| IFITM3P | SNFRFIFSIFTVSN-LKLKLTPEGNKVALLRVPHSS-PPHPESTTTVIPPCL-TLLSGPCLMCCLSFIALAYSVKSRDRKMVGNDLTGALPCF |
| IFITM4P | MNHTVQPSL-LSTAASPTMRCRRSMRWLCGWPTTLLPRRPSTSATRPCCP-TMLSGPCSTLSSPPAAWASISLTPSLLATLAPRPM |
| IFITM6P1 | MNHTAQTFFTPANSCHPBNYEMLKEEHEVAVLGVPHNPAPPMSTMIHICSETSVSDYVVVWSLSNLFMNPCHCLGFI AFAYSVKSRDRKMVGNDLTGAQAYA |
| IFITM6P2 | MNHTVQTFSSPVNSGHPPQIAQGGARGGCAEGTPQPCSTPTSMHRSSETSVDPDRVSLFNTLFVNPCCLGFI AFAYSVKSRDRKMVGNDLTGAQAYA |
| IFITM7P | MNHTVQTFSSPVNSGQPLNYEMLKEEHEVAVGVPHNPAPP-TASTAARPPCCP-TMLSGPCSTPSSPPAAWASRSPTPSLLATPGPRPMP |
| IFITM8P1 | MNHMVQTFFTPISNTGHPATMRCGRSKRWLCSCPTTLLPQRPSTSAARPLCP-TMSSGPRLTASPPAAWASISPIPSLLVTAGPRPMP |
| IFITM11P | MNHTVQTFSSPVNSGQPPQIAQGGTRRGACAGAPHNPAPPTSTVIHRSSETSVPHHVVWSLFNTLFMNPCHCLGFI AFAYSVKSRDRKMVGNDLTGAQAYA |
| IFITM12P2 | MNHTVQTLFTPANTGRSTNHEMLKEKHEVAVLGVPHNPVPPAFTMIHICSETSVPHHVVWSLFNTLFKNSCCPDFI AFIYYSVKSRGSLLVTLGPRPVS |
| IFITM12P3 | MNHNVTSSLLPTSAIPLTMRCRRRSKRWLCWGHPTTLLPQCPEPSTARTPCP-TMSSGPCSTPSSPPAAWASHWPTPSLLATSGPRPMP |
| IFITM16P | MNYTVQTFFTPANSRSPNYKMLKEEQEVAVLGVPHNPAPLTSTSTSTARPI-TMSSGPCLTSSPPAAWASHLPTPSLLTTTGPRPMP |

| | | | |
|--------|-------------------------|-----|-----|
| | 110 | 120 | 130 |
| IFITM2 | | | |
| IFITM3 | | | |

| | |
|-----------|-----------------------------------|
| IFITM8P2 | STAKCLNIWALIVDIVMTILLIIPVLIFQVY** |
| IFITM12P1 | STTKCLNIWALILGILMTILLIIPVLIFQAH** |

| | |
|-----------|-------------------------------------|
| IFITM1P1 | PPPSISSTSLWSLHSSCFSPACSSKQFL- |
| IFITM1P2 | STAKCLNVWALVLGILMSILLIFIPVLIFQVYQ* |
| IFITM2P | STAKCLNVCTILGIRVTTTLLIILSMIIFQAI SQ |
| IFITM3P | HCKVPHLCPLDRHPCSHCTHDFHTRDNLPSN |
| IFITM4P | PPPSATFWASSFCSSS-SQC*SSKPIN- |
| IFITM6P1 | STAKYLNIALIVCIIMTILLVILVLIFQVCR* |
| IFITM6P2 | STAKCLNIALIVDILMTILLIIRVLIFQAYR* |
| IFITM7P | PPPSATLWASSFCSSS-SQC*SSKPID- |
| IFITM8P1 | PLPSATFCTSI*FCCLSS-SQC*SSKST- |
| IFITM11P | STAKCLNIWALILGILMTILLIIPVLIIQAH** |
| IFITM12P2 | PLPSATLWASSFCSSS-SQC*SSKSLD- |
| IFITM12P3 | LPPSATLCASLRPFCSAS-SQC*SEK SIN- |
| IFITM16P | PPPSATLWASSFCSSS-SRC*PSKSID- |

Evolutionary Conservation of IFITM8P2/12P1

IFITM8P2

(Chr. 8)

| | | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | | | | | | | | | |
|-----------|---------------------|---|----------|--------|---------|---------|----------|---------|---------|-----|-----|-----------|--------|-------|-------|---------|---------|--------|-------|
| | | | | | | | | | | | | | | | | | | | |
| IFITMP8P2 | Human | MNHTFQ | TFFTPANS | GHPRNY | EMLKEEH | KVAVVG | APQNP | APPMST | MIHFHSE | TSM | PDH | VVWSL | FNTLFM | NSGCL | CFTAF | VYSMK | SRE | RKKV | |
| IFITMP8P2 | Bonobo | MNHTFQ | TFFTPANS | GHPRNY | EMLKEEH | EVAVVG | APQNP | ALPMST | MIHFHSE | TSM | PDH | VVWSL | FNTLFM | NSGCL | GFI | AFAYS | SLK | SRE | RKKV |
| IFITMP8P2 | Chimpanzee | MNHTFQ | TFFTPANS | GHPRNY | EMLKEEH | EVAVVG | APQNP | ALPMST | MIHFHSE | TSM | PDH | VVWSL | FNTLFM | NSGCL | GFI | AFAYS | SLK | SRE | RKKV |
| IFITMP8P2 | Gorilla | MNHTFQ | TFFTPANS | GHPRNY | EMLKEEH | EVAVVG | VPQNP | APPMST | MIHFHSE | TSM | PDH | VVWSL | FNTLFM | NSGCL | GFI | AFAYS | SMK | SRE | RKKV |
| IFITMP8P2 | Orangutan | MNHIFQ | TFFTPANS | GHPRNY | EMLKEEH | EVAVVG | APQNP | ALPMST | MIHFHSE | TSM | PDH | VWSL | FNTLFM | NSGCL | GFI | E | FAYS | KKSR | GRKMV |
| IFITMP8P2 | Green Monkey | MNHTFQ | TFFTPANS | SRSRNY | EMLKEEH | EVAVL | GAPHNS | APPMST | MIHFSSE | TSV | PDH | VVWSL | FNTLFM | NSSCL | GFI | AFAYS | MTSR | DRKIV | |
| IFITMP8P2 | Gibbon | MNHTFQ | TFFTPANS | GHPCNY | EMLKEEH | EVAVVG | ARPTT | LLPDV | HHDPL | Q | RDL | MPDH | VIWSL | EDAL | FMNS | GCLC | *SAFAYS | SMKCWT | GRCE |
| IFITMP8P2 | Crab-eating Macaque | MNHTFQ | TFFTPANS | SRPPQL | *DAQGG | A*GGCGG | DTPQLYAS | NVHHDPL | Q | *DL | CAR | PCCLVAVQH | HFH | HELQL | PGLHS | ICLLHDV | *GOEDG | | |
| IFITMP8P2 | Rhesus Macaque | MNHTFQ | TFFTPANS | SRPPQL | *DAQGG | A*GGCGG | DTPQLYAS | NVHHDPL | Q | *DL | CAR | PCCLVAVQH | HFH | HELQL | PGLHS | ICLLHDV | *GOEDG | | |
| IFITMP8P2 | Olive baboon | MNHTFQ | TFFTPANS | RRPPQL | *DAQGG | A*GGCGG | DTPQLYSS | NVHHDPL | Q | *DL | CAR | PCCLVAVQH | PHF | HELQL | PGLHS | ICLLHDV | *GOEDG | | |

IFITMP12P1

(Chr. 12)

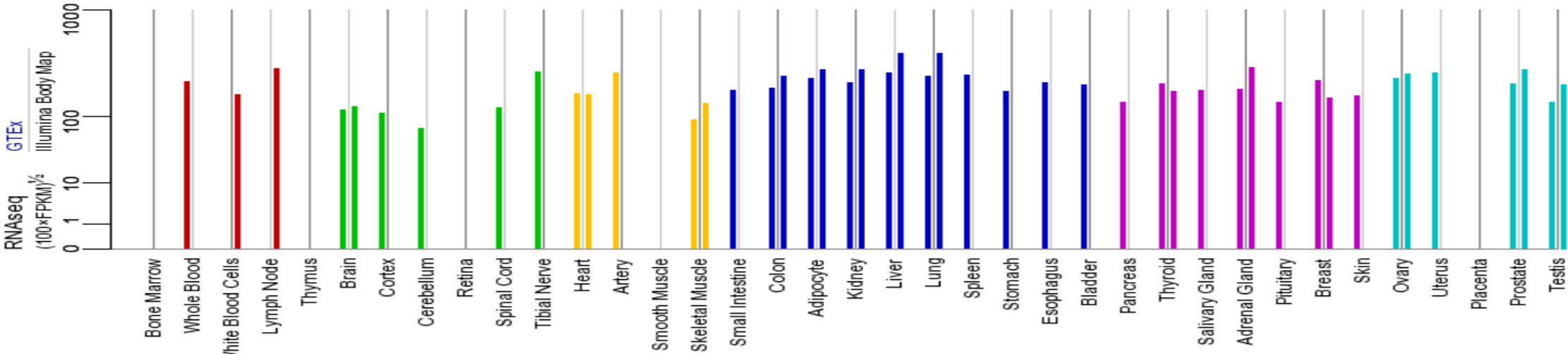
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|------------|------------|--------|----------|--------|---------|-------|-----|------|-------|-----|------|------|-----|-------|-------|------|-------|------|-------|---------|-------|------|------|------|-------|---------|
| IFITMP12P1 | Human | MNHTVQ | TFFSPVNS | GQPPNY | EMLKEEH | KVAVL | GVP | HNPA | PPTST | VIH | IRSK | TSVP | PHH | VVWSL | FNTLF | FMNP | CCLG | FIAF | AYSVK | SRDRKMV | | | | | | |
| IFITMP12P1 | Bonobo | MNHTVQ | TFFSPVNS | GQPPNY | EMLKEEH | KVAVL | GVP | HNPA | PPTST | VIH | IRSK | TSVP | PHC | VVWSL | FNTLF | FMNP | CCLG | FIAF | AYSVK | SRDRKMV | | | | | | |
| IFITMP12P1 | Chimpanzee | MNHTVQ | TFFSPVNS | GQPPNY | EMLKEEH | KVAVL | GVP | HNPA | PPTST | VIH | IRSK | TSVP | PHC | VVWSL | FNTLF | FMNP | CCLG | FIAF | AYSVK | SRDRKMV | | | | | | |
| IFITMP12P1 | Gorilla | MNHTVQ | T-FSPVNS | GQPPNY | EMLKEEH | KVAVL | GVP | HNPA | PPTST | VIH | IRSK | TSVP | PHH | VVWSL | FNTLF | FMNS | CCCLG | FIAF | AYSVK | SRDRKMV | | | | | | |
| IFITMP12P1 | Orangutan | MNHTVQ | TFFSPVNS | GQHPQ | --- | *D | TQ | GGAQ | GGCA | E | GT | PQ | PC | SP | DVH | -R | SET | SVPH | H | VVWSL | FNTLF | FRNP | CCLG | FIAF | AYSVK | SRDRKMV |

| | | 100 | 110 | 120 | 130 | |
|-----------|---------------------|---|------------|----------------|--------------------|--------|
| | | | | | | |
| IFITMP8P2 | Human | GDLTRAQAYTSTA | KCLNIWALIV | DIVMTILLIIIPVL | LIFQVY** | |
| IFITMP8P2 | Bonobo | GDLTRAQAYTSTA | KCLNIWALIV | DIVMTILLIIIPVL | LIFQVY** | |
| IFITMP8P2 | Chimpanzee | GDLTRAQAYTSTA | KCLNIWALIV | DIVMTILLIIIPVL | LIFQVY** | |
| IFITMP8P2 | Gorilla | GDLTRAQAYASTA | KCLNIWALIV | DIVMTILLSIIPVL | LIFQVY** | |
| IFITMP8P2 | Orangutan | GDLTRAQAYASTA | KCLNIWALIV | DIVMTILLIIIPVL | LIFQVY** | |
| IFITMP8P2 | Green Monkey | GDLTRAQAYASTA | KCLNIWALIV | GIVMTVLLIIIPVL | LIFQVY** | |
| IFITMP8P2 | Gibbon | PTCPPLKAISP* | T*DRYTG | P*YSIPS*KFE* | HNMC*SSESID-- | |
| IFITMP8P2 | Crab-eating Macaque | W*PDQGPGLCL | HHQVPEHLG | PD | CGHCHDRSAHHHPGVDLP | SLLI |
| IFITMP8P2 | Rhesus Macaque | W*PDQGPGLCL | HHQVPEHLG | PD | CGHCHDRSAHHHPGVDLP | SLLI |
| IFITMP8P2 | Olive baboon | W*PDQAPMPPPPSA* | TSGP* | LNAL | S*PFCSSSSSRC* | SSKSID |

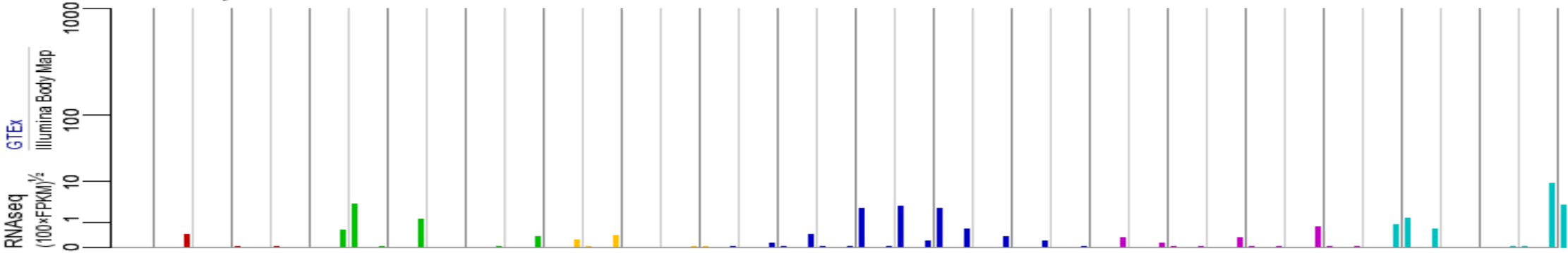
| | | | | | | | | | | | | | | | | | | | | |
|------------|------------|---|-------|--------|------|-------|-------|-------|-------|-------|------|------|--|--|--|--|--|--|--|--|
| IFITMP12P1 | Human | G | NVTGA | QAYAST | TKCL | NIWAL | LILG | ILMT | TILLI | IIPV | LIFQ | AHR* | | | | | | | | |
| IFITMP12P1 | Bonobo | G | NVTGA | QAYAST | TKCL | NIWAL | LILG | ILMT | TILLI | IIPV | LIFQ | AHR* | | | | | | | | |
| IFITMP12P1 | Chimpanzee | G | NVTGA | QAYAST | TKCL | NIWAL | LILG | ILMT | TILLI | IIPV | LIFQ | AHR* | | | | | | | | |
| IFITMP12P1 | Gorilla | G | NVTGA | QAYS | STAK | CLNI | WALIL | GILMT | TILLI | VIPV | LIFQ | AHR* | | | | | | | | |
| IFITMP12P1 | Orangutan | G | DLTGA | QAYAST | TKCL | NIWAL | LILG | ILMT | TILLI | VIPIL | LIFQ | AHR* | | | | | | | | |

mRNA Expression In Normal Human Tissues From Genotype-Tissue Expression (GTEx) Consortium

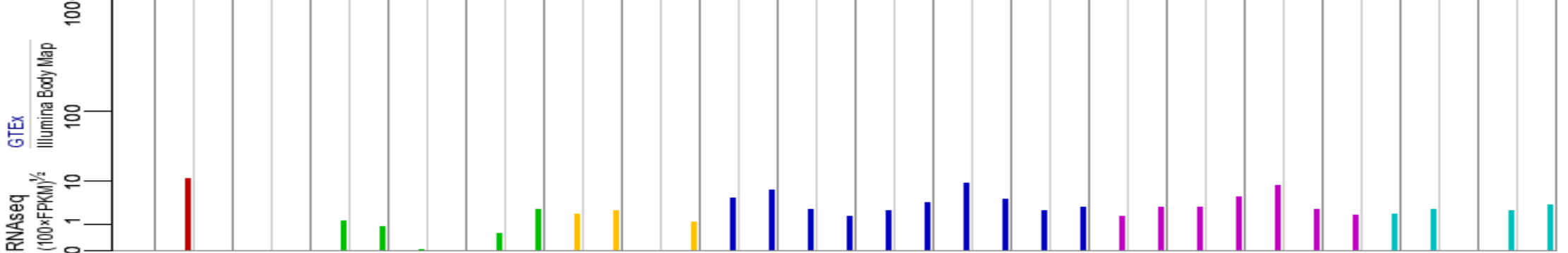
IFITM3



IFITM8P2



IFITMP12P1



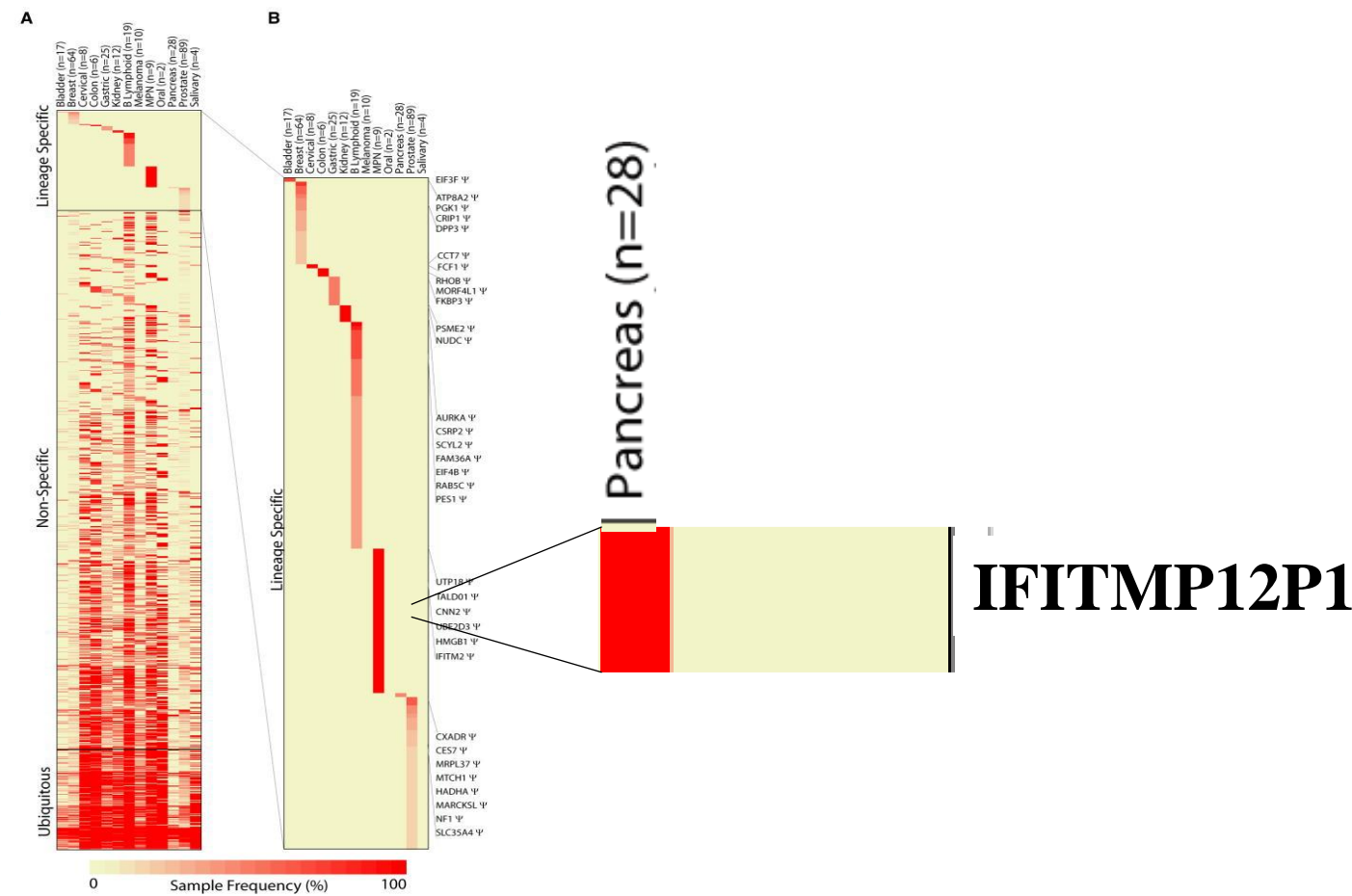
Transcriptomic Evidence of IFITM12P1 Expression

Tissue/Lineage-Specific Pseudogene Expression Profiles



Expressed Pseudogenes in the Transcriptional Landscape of Human Cancers

Cell. 2012 Jun 22; 149(7): 1622–1634.



Proteomics Evidence of IFITM12P1 Expression

| | | | |
|----------------|-----|---|-----------------------------|
| IFITM3 | 1 | MNHTVQTFSPVNSGQPPNYEMLKEEHEVAVLGAPHNPAPPTSTVIHIRSETSVPD | HVV |
| IFITM12P1 | 1 | MNHTVQTFSPVNSGQPPNYEMLKEEHKVAVLGVPHNPAPPTSTVIHIRSKTSVPH | HVV |
| <hr/> | | | |
| IFITM3 | 60 | WSLFNTLFMNPCCLGFI AFAYSVKSRDRKMVG | DTGAQAYASTAKCLNIWALILGILMTI |
| IFITM12P2 | 60 | WSLFNTLFMNPCCLGFI AFAYSVKSRDRKMVG | NTGAQAYASTTKCLNIWALILGILMTI |
| <hr/> | | | |
| Unique peptide | | | |
| IFITM3 | 120 | LLIVIPVLIFQAYG | |
| IFITM12P3 | 120 | LLITIPVLIFQAH | R |

J. Proteomic Research: 2016 Mar 4;15(3):945-55: Myoblastic cell lines

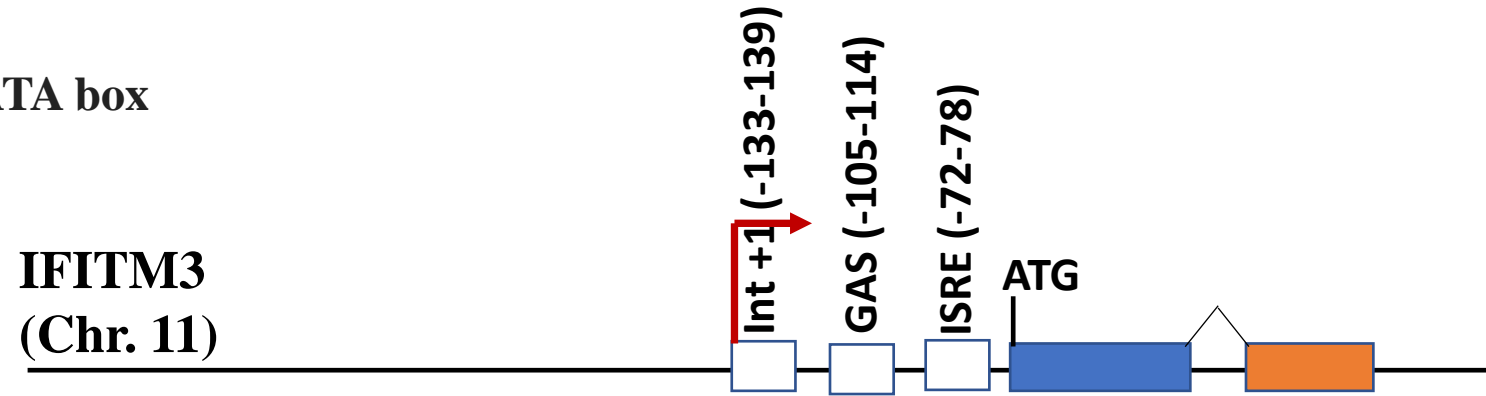
Mol. Cell Proteomics: 2016 Mar;15(3):1072-82: Pancreatic carcinoma

Scientific Data volume2: Article number: 150022 : Colon cancer

Mol. Cell Proteomics: 2012 Mar;11(3) :deep proteomics –in Jurkat cells

Potential Adoption of Regulatory Sequences in IFITM12P1

➤ IFITM3: NO TATA box



ISRE: Interferon-stimulated response element (for IFN-alfa/beta)

GAS: Interferon-Gamma Activated Sequence

Int+1 : Transcription Start Site (TSS)

Aim-1: Is endogenous IFITM12P1 undergoes transcription and if these exapted regulatory sequences are active?

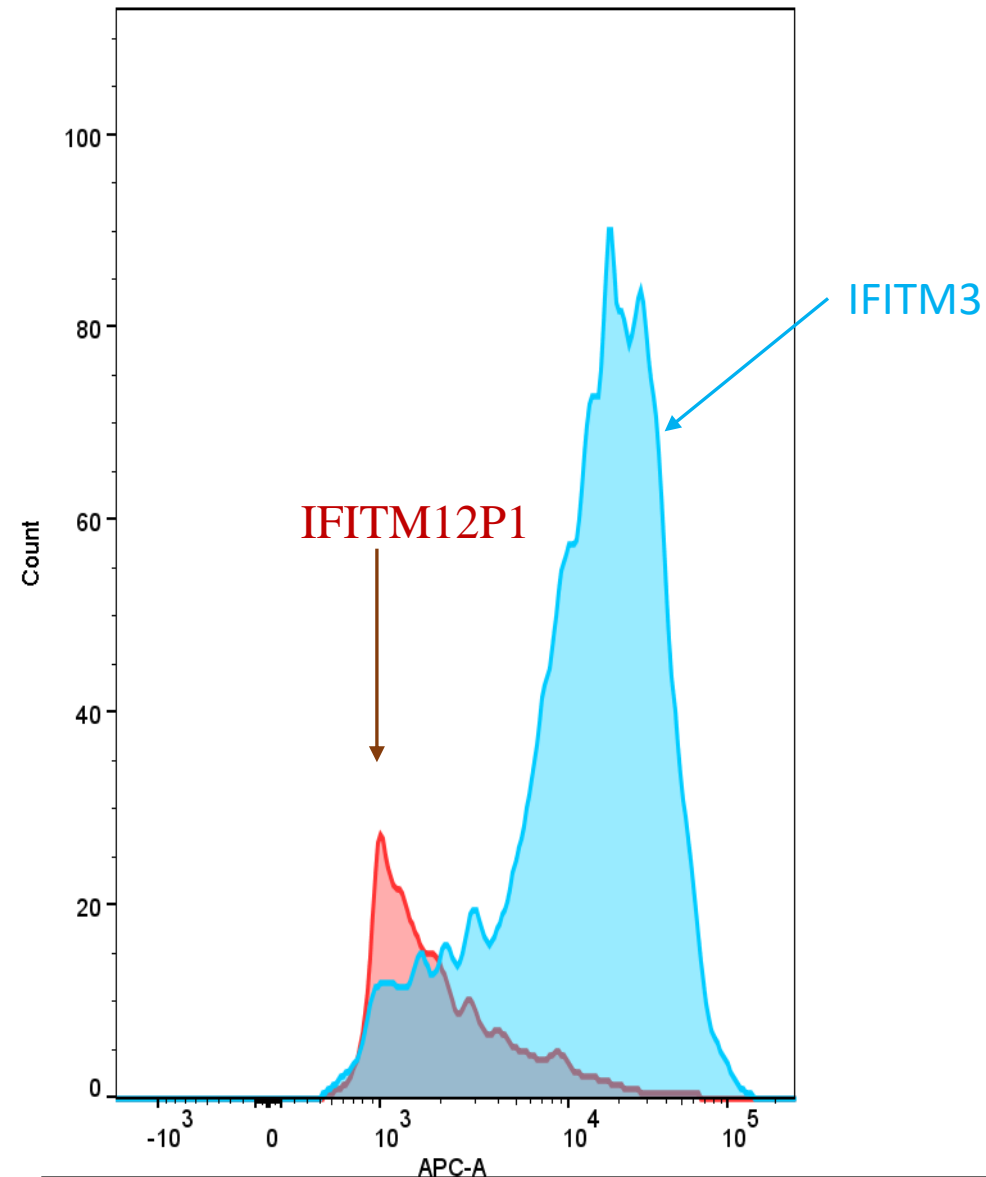
Experiments:

- A.** Determine the level of IFITM12P1 RNA expression in different cells/tissues by RT-PCR
- B.** Endogenous Tagging of IFITM12P1 by CRISPR/Cas9
- C. 5'-RACE** (Rapid amplification of cDNA ends) analysis to check the TSS.

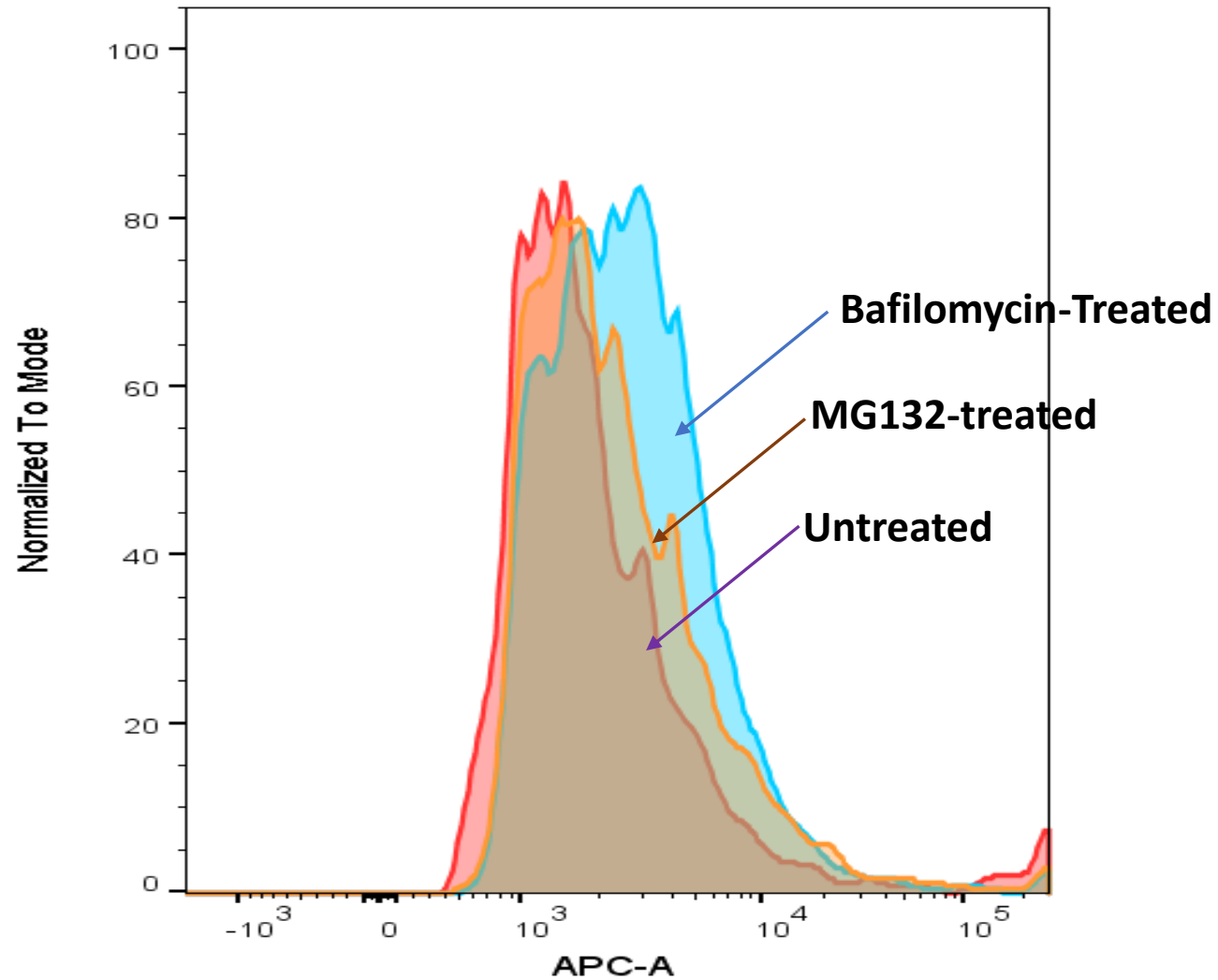
Aim-2: Investigating the role of IFITM12P1

Ectopically Expressed IFITM12P1 Shows Instability

**Stable 293-T cells expressing
FLAG-IFITM12P1/IFITM3**



IFITM12P1 Might Degrade by Lysosome



Why IFIM12P1 shows Instability in comparison to IFITM3?

Sequence Variations among IFITM12P1 and IFITM2/IFITM3

| | | | | | | |
|----------------------|---|--|-------------------------------|--------|------|------|
| IFITM2 | 1 | MNHIVQT-FSPVNSGQPPNYEMLKEEQEVAMLGVPHPNPAPPMSTVIHIRSETSPD | HVVWS | E27K | E50K | D54H |
| IFITM3 | 1 | MNHTVQTF FSPVNSGQPPNYEMLKEEH | EVAVLGAPHNPAPPTSTVIHIRSETSPD | HVVWS | | |
| Human IFITM12P1 | 1 | MNHTVQTF FSPVNSGQPPNYEMLKEEH | KVAVLGVPHPNPAPPTSTVIHIRSKTSPH | HVVWS | | |
| Gorilla- IFITM12P1 | 1 | MNHTVQT-FSPVNSGQPPNYEMLKEEH | KVAVLGVPHPNPAPPTSTVIHIRSKTSPH | HVVWS | | |
| Chimpange- IFITM12P1 | 1 | MNHTVQTF FSPVNSGQPPNYEMLKEEH | KVAVLGVPHPNPAPPTSTVIHIH | SKTSPH | C | VVWS |

| | | | | | |
|----------------------|----|---------------------------------|---------------------------|-------|------|
| IFITM2 | 60 | LFNTLFMNTCCCLGFIAFAYSVKSRDRKMVG | DVTGAQAYASTAKCLNIWALILGIF | MTILL | D91N |
| IFITM3 | 61 | LFNTLFMNPCCCLGFIAFAYSVKSRDRKMVG | DVTGAQAYASTAKCLNIWALILGIL | MTILL | |
| Human- IFITM12P1 | 61 | LFNTLFMNPCCCLGFIAFAYSVKSRDRKMVG | NVTGAQAYASTTKCLNIWALILGIL | MTILL | |
| Gorilla- IFITM12P1 | 60 | LFNTLFMNSCCCLGFIAFAYSVKSRDRKMVG | NVTGAQAYSSTAKCLNIWALILGIL | MTILL | |
| Chimpange- IFITM12P1 | 61 | LFNTLFMNPCCCLGFIAFAYSVKSRDRKMVG | NVTGAQAYASTAKCLNIWALILGIL | MTILL | |

| | | |
|----------------------|-----|--------------|
| IFITM2 | 120 | IIIPVLVVQAQR |
| IFITM3 | 121 | IVIPVLIFQAYG |
| Human- IFITM12P1 | 121 | IIIPVLIFQAHR |
| Gorilla- IFITM12P1 | 120 | IVIPVLIFQAHR |
| Chimpange- IFITM12P1 | 121 | IIIPVLIFQAHR |

Which amino acids are responsible for IFITM12P1 instability?

Aim-2a: Which amino acid(s) render IFITM12P1 unstable?

Experiments:

- A.** SDM of each modified a.a. and testing their impact on expression pattern.
- B.** Determine if that specific mutation is a characteristic pattern of other retrogenes to find for a universal code of retrogene instability

Aim-2b: What is the endogenous role of IFITM12P1 in cellular immunity against virus and/or in other cellular processes?

Experiments:

A. Redundant function of IFITMs?

Knocking out the IFITM12P1 by CRISPR/Cas9 in different cell types and test for mutants against virus challenge.

B. Inhibiting the expression of IFITMs by anti-sense RNA?

Co-expression of IFITM3 and IFITM12P1 and check the impact on IFITM3 expression

C. Enhancing the expression of IFITMs by binding with the IFIMT3 miRNAs?

Expression of IFITM12P1 3'-UTR in Hela cells and explore the expression level of IFITM3

Knocking-out of IFITMP121 by CRISPR attempt

Experimental progress:

-Made stable 293T cells and Hela cells expressing Cas9 protein.



Thanks