Analysis of LC/MS/MS data from latently infected Jurkat cells

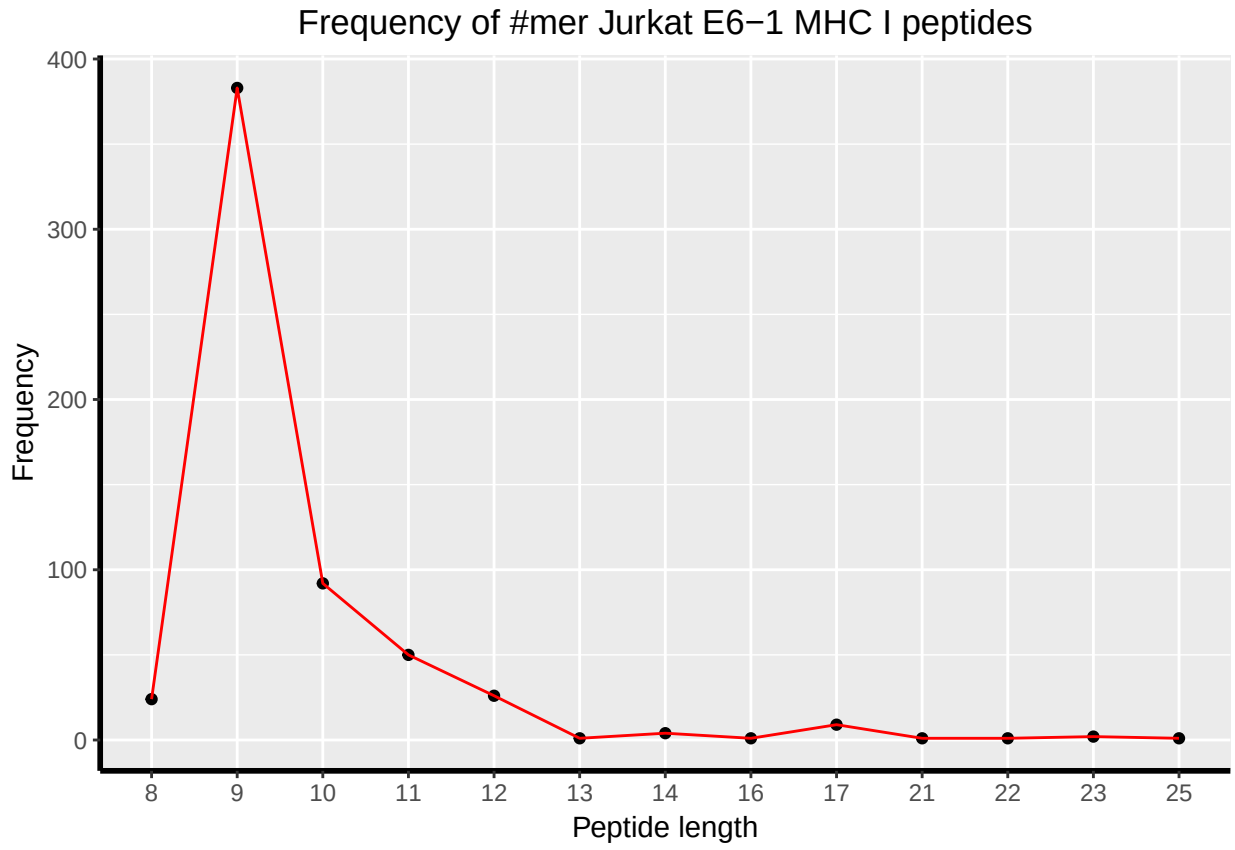
Hem R. Gurung, Ph.D.

October 10, 2018

Approximately 1 gram cell pellet per cell line was processed for HLA-peptide extraction. Peptides were extracted from the HLA molecules, fractionated by RP-HPLC, and ran on nano-LC MS/MS system (TTOF 5600). Raw data from the mass spectrometer was processed within PEAKS software at 1% FDR to generate peptide hits based on peptide spectrum matches.

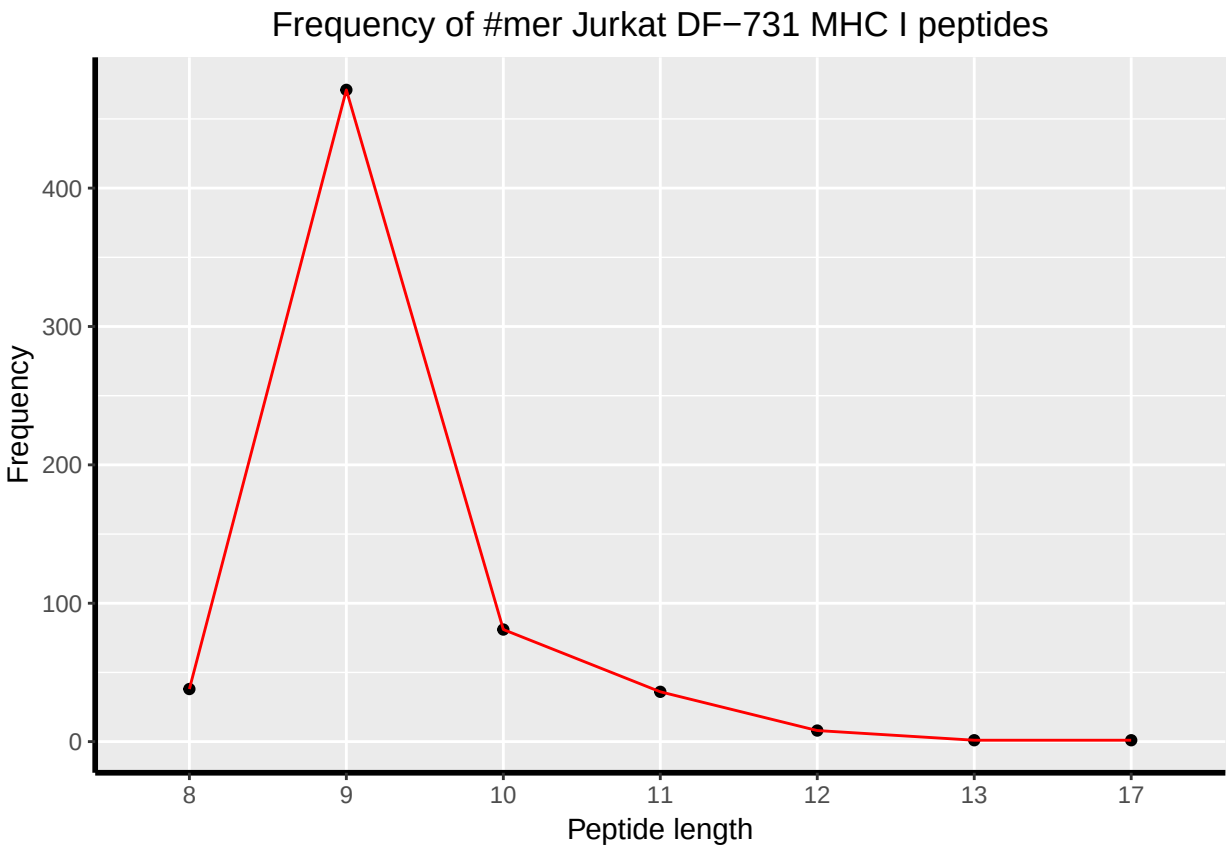
A total of **595** peptides were identified in **uninfected Jurkat E6-1** cell line of which **272** were unique peptides.

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT E61 W632 1g")
data.e61 <- read.csv("protein-peptides.csv", header = T)
table.e61 <- table(data.e61$Length)
table.e61 <- data.frame(table.e61)
library(ggplot2)
ggplot(table.e61, aes(table.e61$Var1, table.e61$Freq, group = 1)) + geom_point() +
  geom_line(color = "red") +
  labs(x = "Peptide length" , y = "Frequency") +
  theme(axis.line = element_line(size = 1),plot.title = element_text(hjust = 0.5)) +
  ggtitle("Frequency of #mer Jurkat E6-1 MHC I peptides")
```



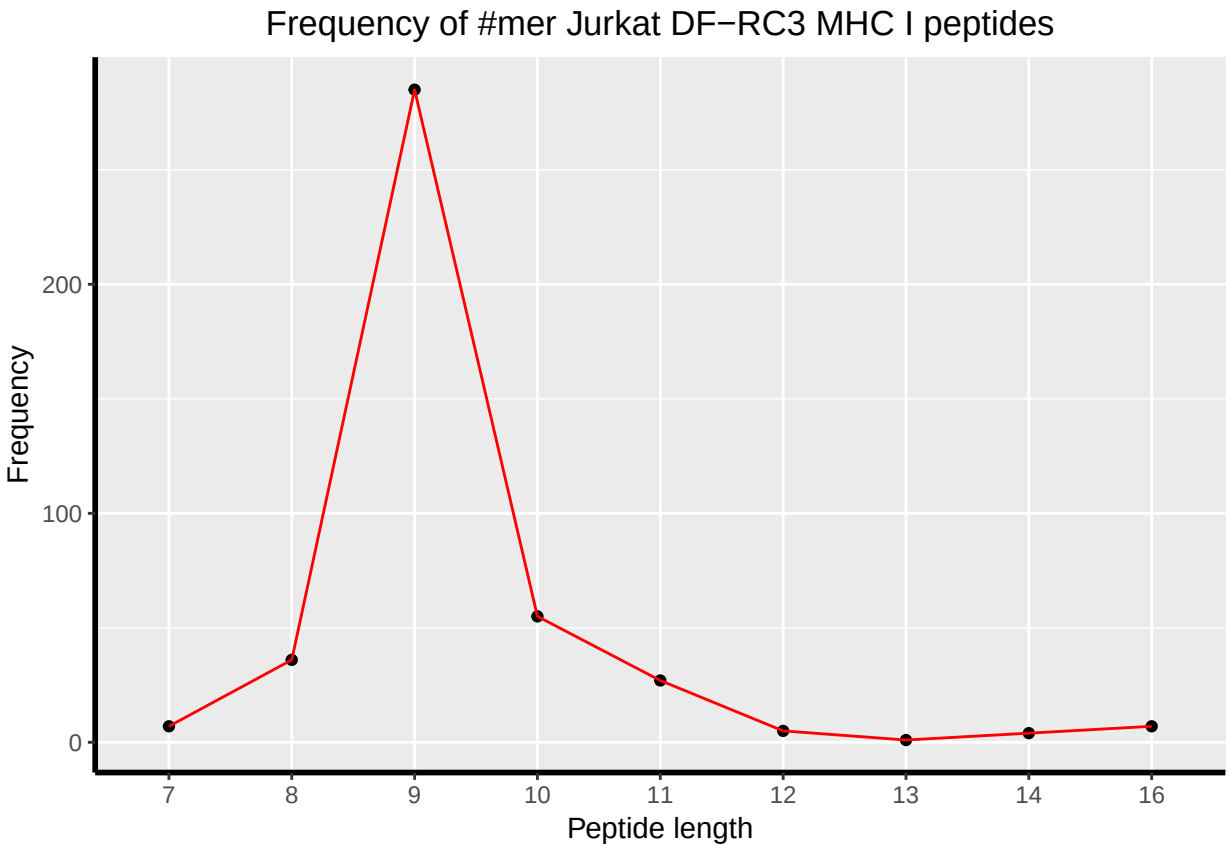
A total of **636** peptides were identified in HIV-1 infected **Jurkat DF731** cell line of which **269** were unique peptides.

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF731 W632 1g")
data.731 <- read.csv("protein-peptides.csv", header = T)
table.731 <- table(data.731$Length)
table.731 <- data.frame(table.731)
library(ggplot2)
ggplot(table.731, aes(table.731$Var1, table.731$Freq, group = 1)) + geom_point() +
  geom_line(color = "red") +
  labs(x = "Peptide length" , y = "Frequency") +
  theme(axis.line = element_line(size = 1),plot.title = element_text(hjust = 0.5)) +
  ggtitle("Frequency of #mer Jurkat DF-731 MHC I peptides")
```



A total of **427** peptides were identified in **HIV-1 infected Jurkat DFRC3** cell line of which **170** were unique peptides.

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF-RC3 W632 1g")
data.RC3 <- read.csv("protein-peptides.csv", header = T)
table.RC3 <- table(data.RC3$Length)
table.RC3 <- data.frame(table.RC3)
library(ggplot2)
ggplot(table.RC3, aes(table.RC3$Var1, table.RC3$Freq, group = 1)) + geom_point() +
  geom_line(color = "red") +
  labs(x = "Peptide length" , y = "Frequency") +
  theme(axis.line = element_line(size = 1),plot.title = element_text(hjust = 0.5)) +
  ggtitle("Frequency of #mer Jurkat DF-RC3 MHC I peptides")
```



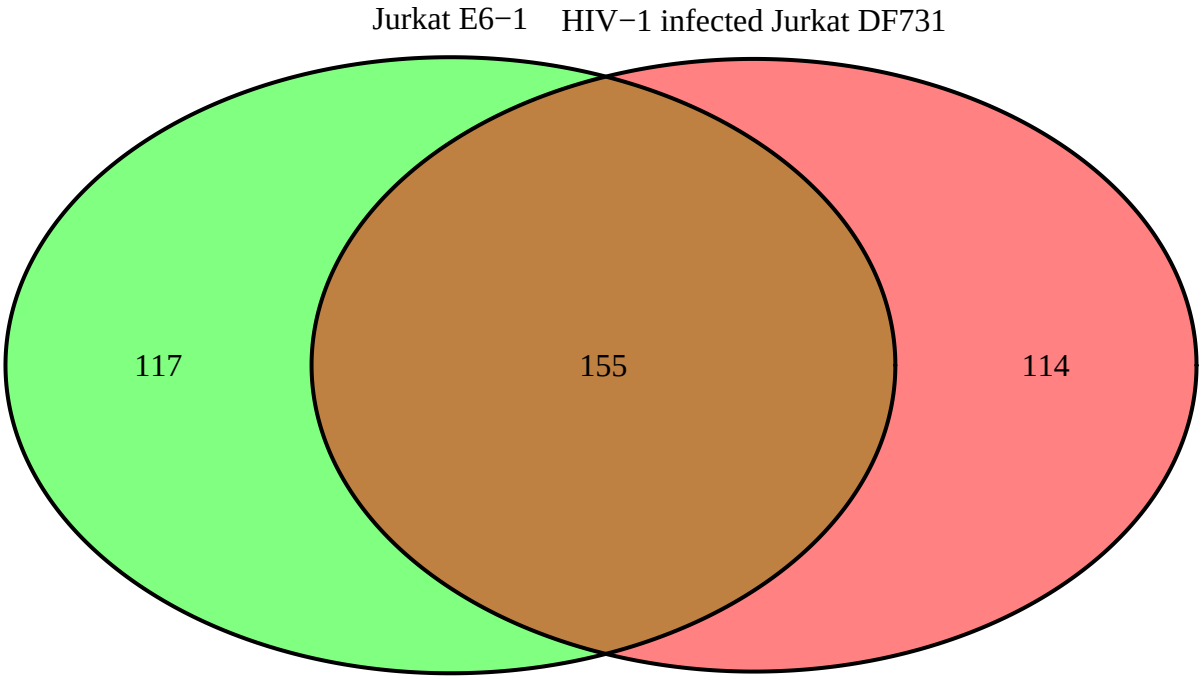
Comparision between uninfected Jurkat E6-1 and HIV-1 infected Jukat DF731 Cell lines

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT E61 W632 1g")
data.e61peptide <- read.csv("peptide.csv", header = T)

setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF731 W632 1g")
data.731peptide <- read.csv("peptide.csv", header = T)

matching.peptides <- data.731peptide$Peptide[data.731peptide$Peptide %in% data.e61peptide$Peptide]
length(matching.peptides)
```

```
## [1] 155
```

Between uninfected Jurkat E6-1 and infected Jurkat DF731, **155** were overlapping peptides.

```
library(VennDiagram)
draw.pairwise.venn(272, 269, 155, category = c("Jurkat E6-1", "HIV-1 infected Jurkat DF731"),
                   fill = c("green", "red"), alpha = rep(0.5,2), cat.pos = c(0,0),
                   cat.dist = rep(0.025,2))
```




```
## (polygon[GRID.polygon.139], polygon[GRID.polygon.140], polygon[GRID.polygon.141], polygon[GRID.polyg
```

That means 117 peptides were *unique* to Jurkat E6-1 and 114 peptides were *unique* to HIV-1 infected Jurkat DF731. Performing database match using Anthony Purcell's Latent HIV-1 database (upregulated genes/proteins) http://hivlatency.erc.monash.edu/searchDBwithExpression?SearchDataset=Mo-HIVvsMockinLatency&SearchEffect=up&SubmitExpression=Submit, 45 peptides in Jurkat E6-1 matched the database whereas 57 peptides in Jurkat DF731 matched the database. Further analysis showed only 2 overlapping genes/proteins between the Jurkat E6-1 and Jurkat DF731 which means 43 peptides were *unique* to Jurkat E6-1 and 55 peptides were *unique* to HIV-1 infected Jurkat DF731.

The reason why the 2 overlapping genes/proteins were observed between the Jurkat E6-1 and Jurkat DF731 was because the two genes/proteins (TNKS2 and TRIP12) were associated with different peptide sequences within each cell line. In other words, different sequences were sampled from the same proteins between the two cell lines.

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT E61 W632 1g")
data.vlookupUniqueE61 <- read.csv("vlookupUniqueE61vsDF731.csv", header = T)
data.vlookupUniqueE61 <- data.frame(lapply(data.vlookupUniqueE61, as.character),
                                    stringsAsFactors = F)
dt <- data.frame(matrix(unlist(strsplit(data.vlookupUniqueE61$Proteins, split = " "))))
colnames(dt) <- "strings"
dt <- as.data.frame(dt[(grep("^GN=", dt$strings)),])
colnames(dt) <- "Genes"
dt <- data.frame(substr(dt$Genes, 4, length(dt$Genes)))
colnames(dt) <- "Genes"

setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF731 W632 1g")
data.vlookupUniqueDF731 <- read.csv("vlookupUniqueDF731vsE61.csv", header = T)
```

```
data.vlookupUniqueDF731 <- data.frame(lapply(data.vlookupUniqueDF731, as.character),
                                      stringsAsFactors = F)
dt.731 <- data.frame(matrix(unlist(strsplit(data.vlookupUniqueDF731$PROTEINS, split = " "))))
colnames(dt.731) <- "strings"
dt.731 <- as.data.frame(dt.731[(grep("^GN=", dt.731$strings)),])
colnames(dt.731) <- "Genes"
dt.731 <- data.frame(substr(dt.731$Genes, 4, length(dt.731$Genes)))
colnames(dt.731) <- "Genes"

# Read hiv-1 latency associated genes
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan")
latency <- read.csv("HIV-1 latency upregulated peptides.csv",  header = T)

# compare to E61 genes
matching.E61.genes <- dt$Genes[dt$Genes %in% latency$Gene.Name]
length(matching.E61.genes) # Number of genes that matched the HIV-1 latency database in Jurkat E6-1
```

```
## [1] 45
```

```
matching.E61.genes <- data.frame(matching.E61.genes)
colnames(matching.E61.genes) <- "Genes"

# compare to DF731 genes
matching.DF731.genes <- dt.731$Genes[dt.731$Genes %in% latency$Gene.Name]
length(matching.DF731.genes)# Number of genes that matched the HIV-1 latency database in
```

```
## [1] 57
```

```
                            # HIV-1 infected Jurkat DF731
matching.DF731.genes <- data.frame(matching.DF731.genes)
colnames(matching.DF731.genes) <- "Genes"

# next question.... are those that matched in each group the same peptides...what are unique to df731?

matching.genes.in.both.samples <- matching.E61.genes$Genes[matching.E61.genes$Genes %in% matching.DF731
length(matching.genes.in.both.samples) # only 2 HIV-1 associated genes were overlapping
```

```
## [1] 2
```

```
                                    # between E6-1 and DF731
matching.genes.in.both.samples <- data.frame(matching.genes.in.both.samples)
colnames(matching.genes.in.both.samples) <- "Genes"

unique.matching.E61.genes <- matching.E61.genes[!(matching.E61.genes$Genes %in% matching.genes.in.both.s
length(unique.matching.E61.genes) # 43 genes in E61 uniquely match the HIV latency database
```

```
## [1] 43
```

```
unique.matching.E61.genes <- data.frame(unique.matching.E61.genes)
colnames(unique.matching.E61.genes) <- "Genes"

unique.matching.DF731.genes <- matching.DF731.genes[!(matching.DF731.genes$Genes %in% matching.genes.in
length(unique.matching.DF731.genes) # 55 genes in DF731 sample uniquely match the HIV latency database
```

```
## [1] 55
```

```
unique.matching.DF731.genes <- data.frame(unique.matching.DF731.genes)
colnames(unique.matching.DF731.genes) <- "Genes"
```

The latent HIV-1 associated host peptides that were *unique* to each group are presented in tables below for each cell line.

Latent HIV-1 associated host peptides in Jurkat E6-1 as compared to Jurkat DF731

```
library(DT)
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT E61 W632 1g")
unique.E61.peptides <- read.csv("Unique-E61-peptides-vs-DF731.csv", header = T)
unique.matching.E61.proteins <- unique.E61.peptides[which(unique.E61.peptides$Genes %in% unique.matchin
datatable(unique.matching.E61.proteins[ , c(1,2,3,4,5,6,7,10,15,16)], filter = "top")
```

Show 10 entries Search:

	Peptide	X.10lgP	Mass	Length	ppm	m.z	RT	Scan	Proteins	Genes
	All	All	All	All	All	All	A	All	All	All
2	GKVKVGVNGFGRIG	68.44	1386.8096	14	0.9	463.2775	27.83	4014	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens OX=9606 GN=GAPDH PE=1 SV=3	GAPDH
4	A(+42.01)SSDIQVKELEKRASGQ	63.72	1886.9697	17	-2.5	629.9956	36.33	5537	Stathmin OS=Homo sapiens OX=9606 GN=STMN1 PE=1 SV=3	STMN1
6	IQRTPKIQVYSRHPAEN	62.11	2036.0916	17	-8.2	510.026	22.03	3288	Beta-2-microglobulin OS=Homo sapiens OX=9606 GN=B2M PE=1 SV=1	B2M
8	S(+42.01)DAAVDTSSEITTK	56.69	1465.6783	14	-4	733.8435	29.64	4463	Isoform 2 of Prothymosin alpha OS=Homo sapiens OX=9606 GN=PTMA	PTMA
14	KLFDHAVSKF	52.69	1190.6448	10	1.4	397.8894	29.47	4399	Isoform Short of Long-chain-fatty-acid--CoA ligase 4 OS=Homo sapiens OX=9606 GN=ACSL4	ACSL4
15	APRLNSADRPM	52.3	1226.6189	11	-0.5	409.88	22.62	3348	Isoform 2 of Proline-rich nuclear receptor coactivator 2 OS=Homo sapiens OX=9606 GN=PNRC2	PNRC2
17	KLISEEDLLRK	50.81	1342.782	11	0	448.6013	29.69	4425	Myc proto-oncogene protein OS=Homo sapiens OX=9606 GN=MYC PE=1 SV=1	MYC
24	Q(-17.03)YDEAVAQF	48.81	1052.4451	9	-4.4	527.2275	50.37	8427	Isoform 4 of Nuclear autoantigenic sperm protein OS=Homo sapiens OX=9606 GN=NASP	NASP
31	APRAIPTTGRGSSGVGL	44.94	1595.8743	17	-2.8	532.9639	27.64	3991	DNA replication licensing factor MCM3 OS=Homo sapiens OX=9606 GN=MCM3 PE=1 SV=3	MCM3
36	VSFPIGIYK	42.7	1022.58	9	2.3	512.2985	44.71	7239	Isoform 3 of Leucine-rich repeat-containing protein 20 OS=Homo sapiens OX=9606 GN=LRRC20	LRRC20

Showing 1 to 10 of 43 entries Previous 1 2 3 4 5 Next

Latent HIV-1 associated host peptides in HIV-1 infected Jurkat DF731 as compared to Jurkat E6-1

```
library(DT)
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF731 W632 1g")
unique.DF731.peptides <- read.csv("Unique-DF731-peptides-vs-E61.csv", header = T)
unique.matching.DF731.proteins <- unique.DF731.peptides[which(unique.DF731.peptides$Genes %in% unique.ma
datatable(unique.matching.DF731.proteins[ , c(1,2,3,4,5,6,7,10,15,16)], filter = "top")
```

Show 10 entries Search:

	Peptide	X.10lgP	Mass	Length	ppm	m.z	RT	Scan	PROTEINS	Genes
	All	All	All	All	All	All	A	All	All	All
3	IYDFIGEFM(+15.99)	56.42	1149.5052	9	1.9	575.761	57.77	8559	Isoform 2 of Transmembrane protein 147 OS=Homo sapiens OX=9606 GN=TMEM147	TMEM147
6	KPSLPFTSL	48.65	988.5593	9	7.8	495.2908	44.8	6330	Cytokine receptor common subunit gamma OS=Homo sapiens OX=9606 GN=IL2RG PE=1 SV=1	IL2RG
7	SFDETVTHF	48.28	1081.4717	9	5.7	541.7462	37.88	5382	DNA topoisomerase 2-binding protein 1 OS=Homo sapiens OX=9606 GN=TOPBP1 PE=1 SV=3	TOPBP1
9	VFDPALNLF	46.99	1034.5437	9	0.7	518.2795	61.62	8427	Ubiquitin-protein ligase E3B OS=Homo sapiens OX=9606 GN=UBE3B PE=1 SV=3	UBE3B
11	IRSSYIRVL	44.8	1105.6608	9	0.8	369.5612	31.34	4773	DNA replication licensing factor MCM5 OS=Homo sapiens OX=9606 GN=MCM5 PE=1 SV=5	MCM5
14	LYDPVISKL	41.18	1046.6012	9	6.4	524.3112	42.74	6109	Isoform 3 of Microspherule protein 1 OS=Homo sapiens OX=9606 GN=MCRS1	MCRS1
17	SFDGIIAM(+15.99)M(+15.99)	40.46	1015.4354	9	6.6	508.7283	39.85	5815	Transcription elongation factor SPT4 OS=Homo sapiens OX=9606 GN=SUPT4H1 PE=1 SV=1	SUPT4H1
19	IARLPSSTL	39.56	956.5654	9	5.5	479.2926	30.97	4723	Isoform 2 of Kinesin-like protein KIF18B OS=Homo sapiens OX=9606 GN=KIF18B	KIF18B
20	RLPSSTLKR	38.94	1056.6404	9	0.6	353.2209	21.81	3546	Isoform 2 of Kinesin-like protein KIF18B OS=Homo sapiens OX=9606 GN=KIF18B	KIF18B
22	VRPPPGPEL	38.29	960.5392	9	-3.8	481.275	28.29	4248	Isoform 3 of Helicase SRCAP OS=Homo sapiens OX=9606 GN=SRCAP	SRCAP

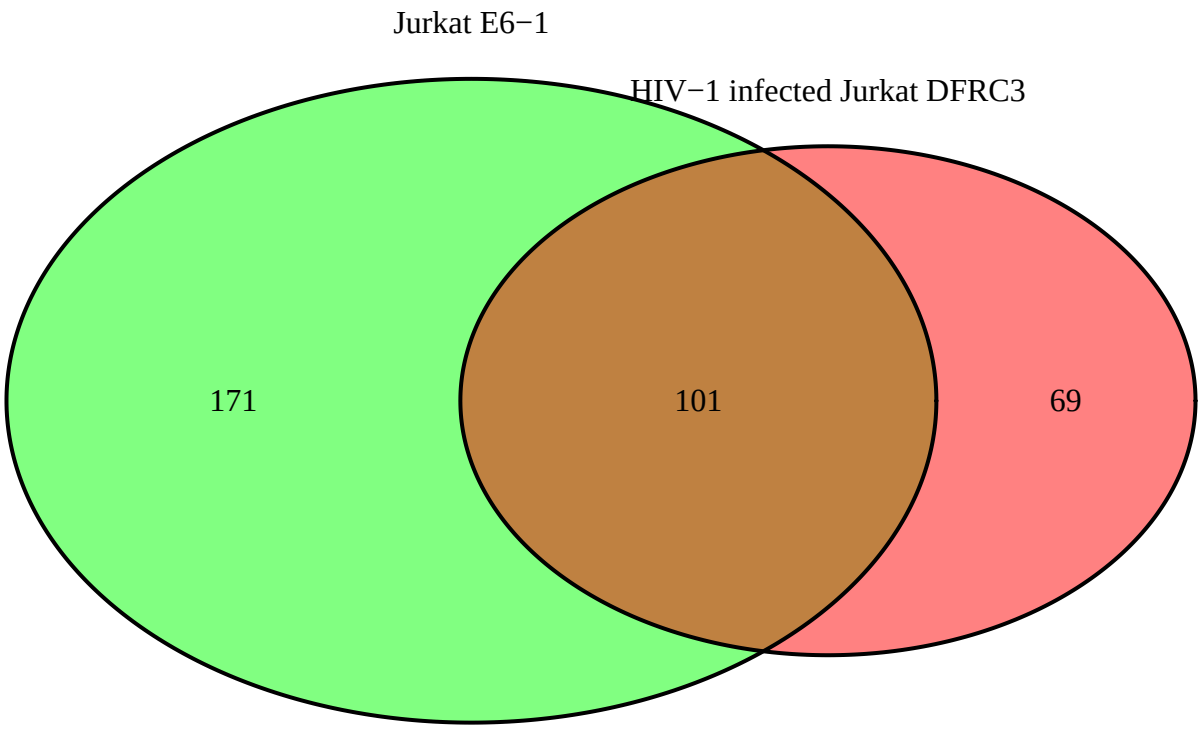
Showing 1 to 10 of 55 entries Previous 1 2 3 4 5 6 Next

Similarly, comparing the uninfected Jurkat E6-1 and HIV-1 infected Jukat DFRC3 Cell lines

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT E61 W632 1g")
data.e61peptide <- read.csv("peptide.csv", header = T)

setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF-RC3 W632 1g")
data.RC3peptide <- read.csv("peptide.csv",  header = T)

matching.peptides <- data.RC3peptide$Peptide[data.RC3peptide$Peptide %in% data.e61peptide$Peptide]
```

Between uninfected Jurkat E6-1 and infected Jurkat DFRC3, **101** were overlapping peptides.

```
library(VennDiagram)
draw.pairwise.venn(272, 170, 101, category = c("Jurkat E6-1", "HIV-1 infected Jurkat DFRC3"),
                   fill = c("green", "red"), alpha = rep(0.5,2), cat.pos = c(0,0),
                   cat.dist = rep(0.035,2))
```




```
## (polygon[GRID.polygon.148], polygon[GRID.polygon.149], polygon[GRID.polygon.150], polygon[GRID.polyg
```

That means 171 peptides were *unique* to Jurkat E6-1 and 69 peptides were *unique* to HIV-1 infected Jurkat DFRC3. Performing database match using Anthony Purcell's Latent HIV-1 database (up-regulated genes/proteins) http://hivlatency.erc.monash.edu/searchDBwithExpression?SearchDataset=Mo-HIVvsMockinLatency&SearchEffect=up&SubmitExpression=Submit, 66 peptides in Jurkat E6-1 matched the database whereas 28 peptides in Jurkat DF731 matched the database. Further analysis showed only 1 overlapping genes/proteins between the Jurkat E6-1 and Jurkat DF731 which means 65 peptides were *unique* to Jurkat E6-1 and 27 peptides were *unique* to HIV-1 infected Jurkat DFRC3.

Again, the reason why the 1 overlapping genes/protein was observed between the Jurkat E6-1 and Jurkat DFRC3 was because the gene/protein (TNKS2) was associated with different peptide sequences within each cell line. In other words, different sequences were sampled from the same protein between the two cell lines.

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT E61 W632 1g")
data.vlookupUniqueE61 <- read.csv("vlookupUniqueE61vsRC3.csv", header = T)
data.vlookupUniqueE61 <- data.frame(lapply(data.vlookupUniqueE61, as.character), stringsAsFactors = F)

dt <- data.frame(matrix(unlist(strsplit(data.vlookupUniqueE61$Proteins, split = " "))))
colnames(dt) <- "strings"
dt <- as.data.frame(dt[(grep("^GN=", dt$strings)),])
colnames(dt) <- "Genes"
dt <- data.frame(substr(dt$Genes, 4, length(dt$Genes)))
colnames(dt) <- "Genes"

###
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF-RC3 W632 1g")
data.vlookupUniqueRC3 <- read.csv("vlookupUniqueRC3vsE61.csv", header = T)
data.vlookupUniqueRC3 <- data.frame(lapply(data.vlookupUniqueRC3, as.character), stringsAsFactors = F)

dt.RC3 <- data.frame(matrix(unlist(strsplit(data.vlookupUniqueRC3$Proteins, split = " "))))
colnames(dt.RC3) <- "strings"
dt.RC3 <- as.data.frame(dt.RC3[(grep("^GN=", dt.RC3$strings)),])
colnames(dt.RC3) <- "Genes"

dt.RC3 <- data.frame(substr(dt.RC3$Genes, 4, length(dt.RC3$Genes)))
colnames(dt.RC3) <- "Genes"

# Read hiv-1 latency associated genes
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan")
latency <- read.csv("HIV-1 latency upregulated peptides.csv",  header = T)

# compare to E61 genes
matching.E61.genes <- dt$Genes[dt$Genes %in% latency$Gene.Name]
length(matching.E61.genes) # 66 peptides from E6-1 matched the HIV-1 latency database
```

```
## [1] 66
```

```
matching.E61.genes <- data.frame(matching.E61.genes)
colnames(matching.E61.genes) <- "Genes"

# compare to DFRC3 genes
matching.DFRC3.genes <- dt.RC3$Genes[dt.RC3$Genes %in% latency$Gene.Name]
length(matching.DFRC3.genes) # 28 peptides from RC3 matched the HIV-1 latency database
```

```
## [1] 28
```

```
matching.DFRC3.genes <- data.frame(matching.DFRC3.genes)
colnames(matching.DFRC3.genes) <- "Genes"

# next question.... are those that matched in each group the same peptides...what are unique to dfRC3?

matching.genes.in.both.samples <- matching.E61.genes$Genes[matching.E61.genes$Genes %in% matching.DFRC3
length(matching.genes.in.both.samples) # Only 1 HIV-1 latency associated gene/protein overlapped
```

```
## [1] 1
```

```
matching.genes.in.both.samples <- data.frame(matching.genes.in.both.samples)
colnames(matching.genes.in.both.samples) <- "Genes"

# only 1 gene (TNKS2) is common to both uninfected and infected jurkats that matched
# the HIV latency associated genes

unique.matching.E61.genes <- matching.E61.genes[!(matching.E61.genes$Genes %in% matching.genes.in.both.s
length(unique.matching.E61.genes) # 65
```

```
## [1] 65
```

```
unique.matching.E61.genes <- data.frame(unique.matching.E61.genes)
colnames(unique.matching.E61.genes) <- "Genes"

unique.matching.DFRC3.genes <- matching.DFRC3.genes[!(matching.DFRC3.genes$Genes %in% matching.genes.in
length(unique.matching.DFRC3.genes) #27
```

```
## [1] 27
```

```
unique.matching.DFRC3.genes <- data.frame(unique.matching.DFRC3.genes)
colnames(unique.matching.DFRC3.genes) <- "Genes"
```

The latent HIV-1 associated host peptides that were *unique* to each group are presented in tables below for each cell line.

Latent HIV-1 associated host peptides in Jurkat E6-1 as compared to Jurkat DFRC3

```
library(DT)
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT E61 W632 1g")
unique.E61.peptides <- read.csv("Unique-E61-peptides-vs-RC3.csv", header = T)
unique.matching.E61.proteins <- unique.E61.peptides[which(unique.E61.peptides$Genes %in% unique.matching
datatable(unique.matching.E61.proteins[ , c(1,2,3,4,5,6,7,10,15,16)], filter = "top")
```

Show 10 entries Search:

	Peptide	X.10lgP	Mass	Length	ppm	m.z	RT	Scan	Proteins	Genes
	All	All	All	All	All	All	All	All	All	All
2	GKVKVGVNGFGRIG	68.44	1386.8096	14	0.9	463.2775	27.83	4014	Glyceraldehyde-3-phosphate dehydrogenase OS=Homo sapiens OX=9606 GN=GAPDH PE=1 SV=3	GAPDH
4	A(+42.01)SSDIQVKELEKRASGQ	63.72	1886.9697	17	-2.5	629.9956	36.33	5537	Stathmin OS=Homo sapiens OX=9606 GN=STMN1 PE=1 SV=3	STMN1
6	IQRTPKIQVYSRHPAEN	62.11	2036.0916	17	-8.2	510.026	22.03	3288	Beta-2-microglobulin OS=Homo sapiens OX=9606 GN=B2M PE=1 SV=1	B2M
15	KLFDHAVSKF	52.69	1190.6448	10	1.4	397.8894	29.47	4399	Isoform Short of Long-chain-fatty-acid--CoA ligase 4 OS=Homo sapiens OX=9606 GN=ACSL4	ACSL4
16	SVYYNEATGGK	52.67	1187.5459	11	-4.9	594.7773	24.8	3680	Tubulin beta chain OS=Homo sapiens OX=9606 GN=TUBB PE=1 SV=2	TUBB
17	APRLNSADRPM	52.3	1226.6189	11	-0.5	409.88	22.62	3348	Isoform 2 of Proline-rich nuclear receptor coactivator 2 OS=Homo sapiens OX=9606 GN=PNRC2	PNRC2
20	KLISEEDLLRK	50.81	1342.782	11	0	448.6013	29.69	4425	Myc proto-oncogene protein OS=Homo sapiens OX=9606 GN=MYC PE=1 SV=1	MYC
28	Q(-17.03)YDEAVAQF	48.81	1052.4451	9	-4.4	527.2275	50.37	8427	Isoform 4 of Nuclear autoantigenic sperm protein OS=Homo sapiens OX=9606 GN=NASP	NASP
29	VLAPEGSVPNK	48.68	1109.608	11	-9.9	555.8058	24.86	3684	Isoform B of Ras GTPase-activating protein-binding protein 2 OS=Homo sapiens OX=9606 GN=G3BP2	G3BP2
36	LPRAVGTQTL	45.91	1054.6135	10	-7.4	528.3101	29.78	4370	Isoform 1A of Sorting nexin-1 OS=Homo sapiens OX=9606 GN=SNX1	SNX1

Showing 1 to 10 of 65 entries Previous 1 2 3 4 5 6 7 Next

Latent HIV-1 associated host peptides in HIV-1 infected Jurkat DFRC3 as compared to Jurkat E6-1

```
library(DT)
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF-RC3 W632 1g")
unique.DFRC3.peptides <- read.csv("Unique-DFRC3-peptides-vs-E61.csv", header = T)
unique.matching.DFRC3.proteins <- unique.DFRC3.peptides[which(unique.DFRC3.peptides$Genes %in% unique.ma
datatable(unique.matching.DFRC3.proteins[ , c(1,2,3,4,5,6,7,10,15,16)], filter = "top")
```

Show 10 entries Search:

	Peptide	X.10lgP	Mass	Length	ppm	m.z	RT	Scan	Proteins	Genes
	All	All	All	All	All	All	A	All	All	All
7	IVAGSLITK	46.51	900.5644	9	14.3	451.2959	28.13	4242	Isoform 2 of Pre-mRNA-processing factor 40 homolog A OS=Homo sapiens OX=9606 GN=PRPF40A	PRPF40A
8	IRSSYIRVL	44.85	1105.6608	9	8.8	369.5641	31.49	4776	DNA replication licensing factor MCM5 OS=Homo sapiens OX=9606 GN=MCM5 PE=1 SV=5	MCM5
13	VRPPPGPEL	39.37	960.5392	9	3.2	481.2784	28.48	4287	Isoform 3 of Helicase SRCAP OS=Homo sapiens OX=9606 GN=SRCAP	SRCAP
14	IYDFIGEFM(+15.99)	38.48	1149.5052	9	4.5	575.7625	57.7	7612	Isoform 2 of Transmembrane protein 147 OS=Homo sapiens OX=9606 GN=TMEM147	TMEM147
15	AYRQGHFTY	37.4	1141.5305	9	6.3	381.5198	24.95	3851	Structural maintenance of chromosomes protein 6 OS=Homo sapiens OX=9606 GN=SMC6 PE=1 SV=2	SMC6
16	IARLPSSTL	37.39	956.5654	9	5.8	479.2928	31.34	4637	Isoform 2 of Kinesin-like protein KIF18B OS=Homo sapiens OX=9606 GN=KIF18B	KIF18B
17	IYDKFIAQL	36.99	1109.6121	9	7.9	555.8177	46.34	6625	Isoform 4 of Polyamine-modulated factor 1 OS=Homo sapiens OX=9606 GN=PMF1	PMF1
23	VPEEARPAL	33.4	980.5291	9	1.2	491.2724	26.16	3948	DDB1- and CUL4-associated factor 15 OS=Homo sapiens OX=9606 GN=DCAF15 PE=1 SV=1	DCAF15
24	KPSLPFTSL	32.55	988.5593	9	7.3	495.2906	44.96	6246	Cytokine receptor common subunit gamma OS=Homo sapiens OX=9606 GN=IL2RG PE=1 SV=1	IL2RG
25	VLYNGFTGR	32.26	1025.5294	9	7.5	513.7758	31.34	4488	DNA-directed RNA polymerase II subunit RPB2 OS=Homo sapiens OX=9606 GN=POLR2B PE=1 SV=1	POLR2B

Showing 1 to 10 of 27 entries Previous 1 2 3 Next

How many genes overlap between the Jurkat DF731 and Jurkat DFRC3?

20 genes (18 peptides) overlapped the HIV-1 infected Jurkat DF731 and Jurkat DFRC3

```
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF731 W632 1g")
data.jrkt731 <- read.csv("Unique-HIV1-related-DF731-peptides.csv", header = T)
setwd("C:/Users/HGURUNG1/Desktop/HG/Projects/West and Belshan/JRKT DF-RC3 W632 1g")
data.jrktRC3 <- read.csv("Unique-HIV1-related-DFRC3-peptides.csv", header = T)

# How many genes overlap ?

overlap.genes <- data.jrkt731$Genes[data.jrkt731$Genes %in% data.jrktRC3$Genes]
length(unique(overlap.genes))
```

```
## [1] 18
```

```
overlap.peptides <- data.jrkt731$Peptide[data.jrkt731$Peptide %in% data.jrktRC3$Peptide]
length(unique(overlap.peptides))
```

```
## [1] 18
```

```
# identify the two peptides that do not overlap

# overlapping.peptides.from.genes <- data.jrkt731[data.jrkt731$Genes %in% overlap.genes, ]
# two.nonoverlapping.peptides <- overlapping.peptides.from.genes[!overlapping.peptides.from.genes$Pepti
# two.nonoverlapping.peptides$Peptide

# The two peptides were from KIF18B gene that were not present in RC3
```

The table below contains the final list of 18 unique peptides/genes/proteins that overlapped HIV-1 infected Jurkats DF731 and Jurkat DFRC3

```
library(DT)
datatable(data.jrkt731[data.jrkt731$Peptide %in% overlap.peptides, c(1,2,3,4,5,6,7,10,15,16)],
          filter = "top")
```

Show 10 entries

Search:

	Peptide	X.10lgP	Mass	Length	ppm	m.z	RT	Scan	PROTEINS	Genes
	All	All	All	All	All	All	A	All	All	All
1	IYDFIGEFM(+15.99)	56.42	1149.5052	9	1.9	575.761	57.77	8559	Isoform 2 of Transmembrane protein 147 OS=Homo sapiens OX=9606 GN=TMEM147	TMEM147
2	KPSLPFTSL	48.65	988.5593	9	7.8	495.2908	44.8	6330	Cytokine receptor common subunit gamma OS=Homo sapiens OX=9606 GN=IL2RG PE=1 SV=1	IL2RG
3	SFDETVTHF	48.28	1081.4717	9	5.7	541.7462	37.88	5382	DNA topoisomerase 2-binding protein 1 OS=Homo sapiens OX=9606 GN=TOPBP1 PE=1 SV=3	TOPBP1
5	IRSSYIRVL	44.8	1105.6608	9	0.8	369.5612	31.34	4773	DNA replication licensing factor MCM5 OS=Homo sapiens OX=9606 GN=MCM5 PE=1 SV=5	MCM5
8	IARLPSSTL	39.56	956.5654	9	5.5	479.2926	30.97	4723	Isoform 2 of Kinesin-like protein KIF18B OS=Homo sapiens OX=9606 GN=KIF18B	KIF18B
10	VRPPPGPEL	38.29	960.5392	9	-3.8	481.275	28.29	4248	Isoform 3 of Helicase SRCAP OS=Homo sapiens OX=9606 GN=SRCAP	SRCAP
12	AIYKTPPGIQK	36.86	1214.7023	11	3.1	405.9093	23.43	3576	Isoform 3 of BCL-6 corepressor OS=Homo sapiens OX=9606 GN=BCOR	BCOR
13	IYDKFIAQL	36.75	1109.6121	9	11.9	555.8199	46.6	6708	Isoform 4 of Polyamine-modulated factor 1 OS=Homo sapiens OX=9606 GN=PMF1	PMF1
17	VM(+15.99)APRTLLL	35.49	1028.6052	9	8	515.314	40.5	5857	HLA class I histocompatibility antigen A-33 alpha chain OS=Homo sapiens OX=9606 GN=HLA-A PE=1 SV=3	HLA-A
20	APSGGRLLSA	33.27	927.5137	10	13.7	464.7705	25.74	3934	Isoform 2 of Aladin OS=Homo sapiens OX=9606 GN=AAAS	AAAS

Showing 1 to 10 of 18 entries

Previous 1 2 Next