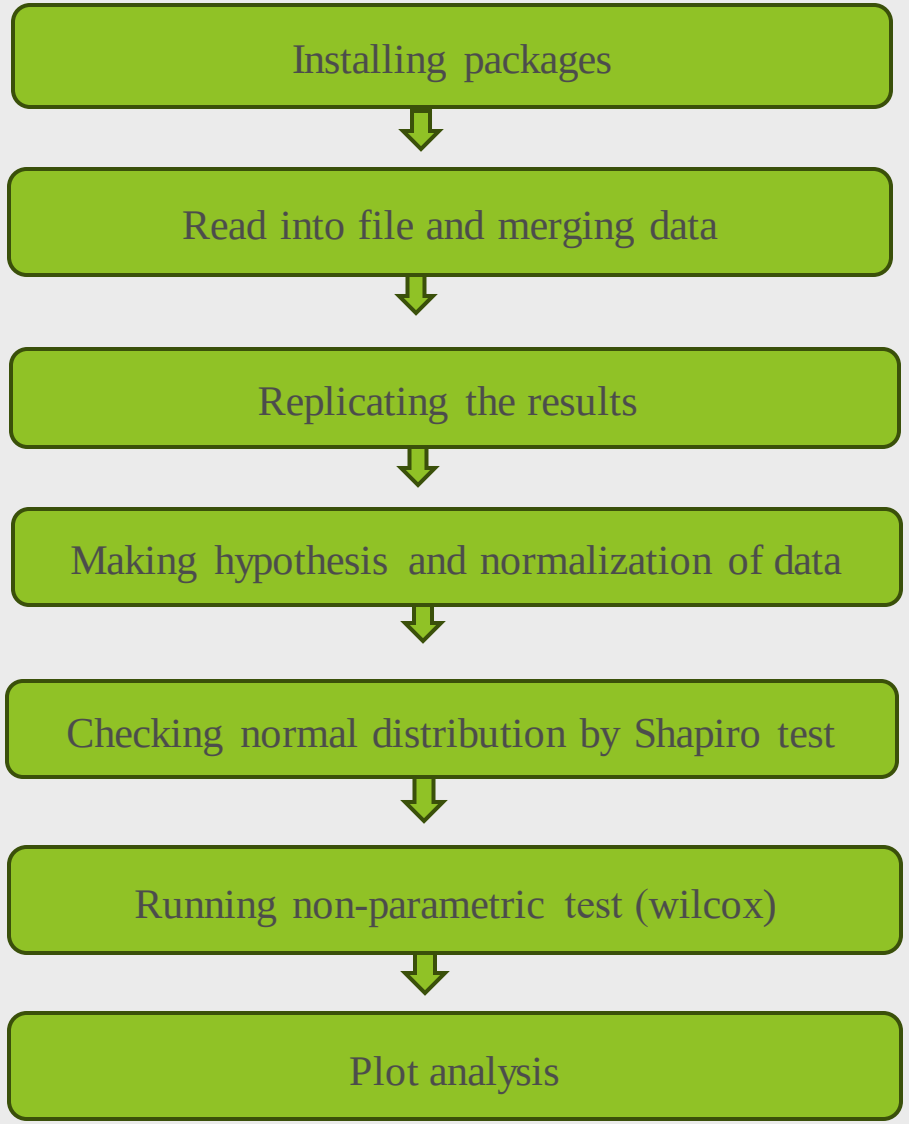
BIOSTATISTICS COURSE PROJECT

Group members -

- Aditya Fulluke - 20018
- Richard David - 20225
- Vaishnavi Agarwal - 20301
- Vaishnavi Khatri - 20302
- Mehar Khurana - 20322

Group number 20

WORKING SCHEME



20301, 20302

ARTICLE

doi:10.1038/nature15399

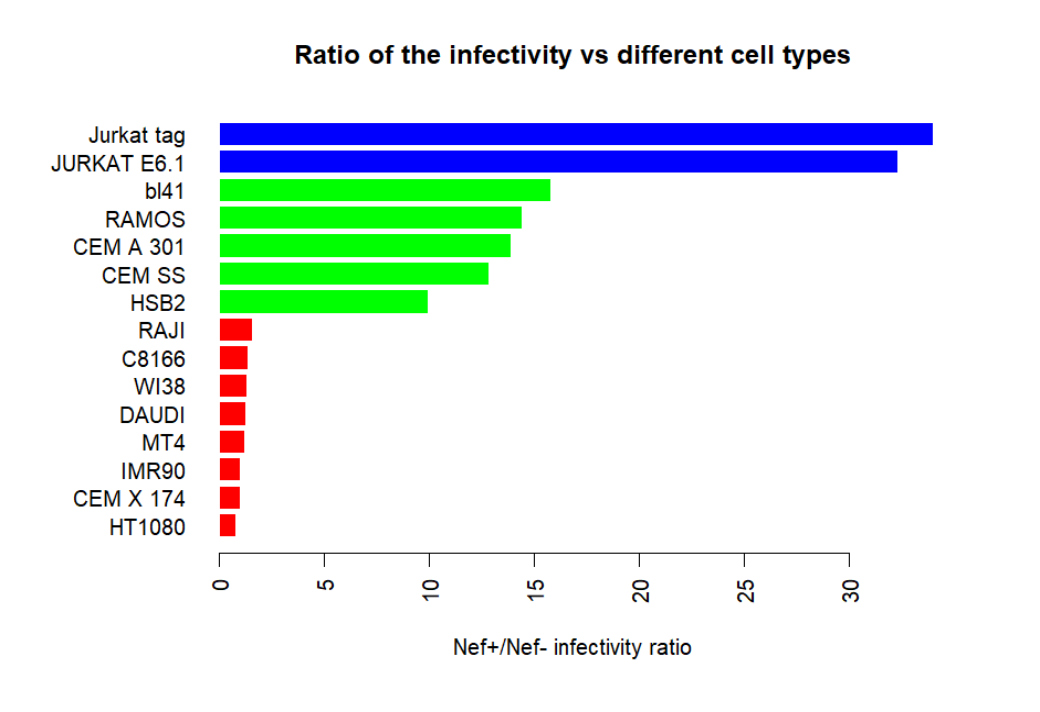
HIV–1 Nef promotes infection by excluding SERINC5 from virion incorporation

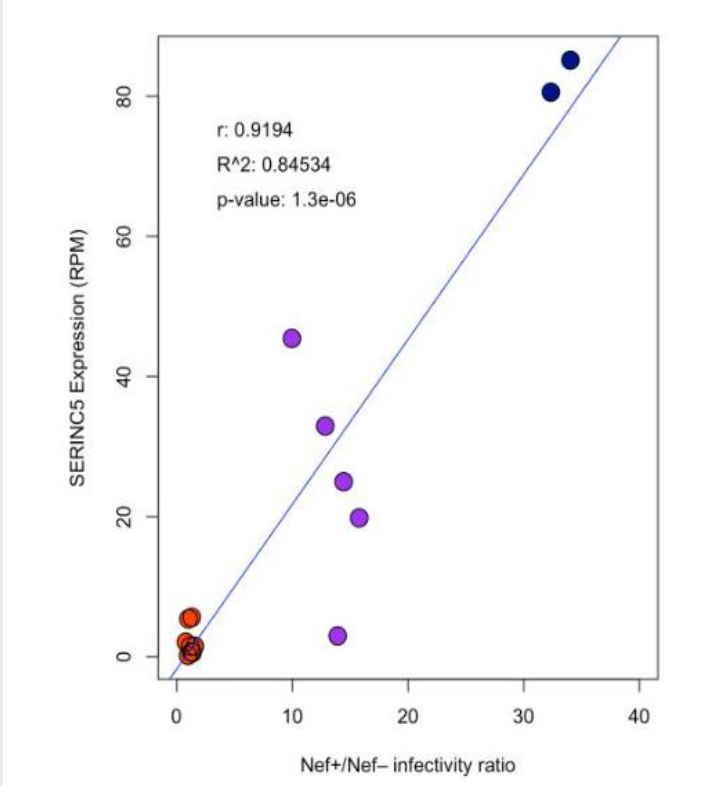
Annachiara Rosa[1]*, Ajit Chande[1]*, Serena Ziglio[1]*, Veronica De Sanctis[2], Roberto Bertorelli[2], Shih Lin Goh[3], Sean M. McCauley[3], Anetta Nowosielska[3], Stylianos E. Antonarakis[4,5], Jeremy Luban[3], Federico Andrea Santoni[4] & Massimo Pizzato[1]

HIV-1 Nef, a protein important for the development of AIDS, has well-characterized effects on host membrane trafficking and receptor downregulation. By an unidentified mechanism, Nef increases the intrinsic infectivity of HIV-1 virions in a host-cell-dependent manner. Here we identify the host transmembrane protein SERINC5, and to a lesser extent SERINC3, as a potent inhibitor of HIV-1 particle infectivity that is counteracted by Nef. SERINC5 localizes to the plasma membrane, where it is efficiently incorporated into budding HIV-1 virions and impairs subsequent virion penetration of susceptible target cells. Nef redirects SERINC5 to a Rab7-positive endosomal compartment and thereby excludes it from HIV-1 particles. The ability to counteract SERINC5 was conserved in Nef encoded by diverse primate immunodeficiency viruses, as well as in the structurally unrelated glycosylated Gag from murine leukaemia virus. These examples of functional conservation and convergent evolution emphasize the fundamental importance of SERINC5 as a potent anti-retroviral factor.

20301, 20302

Replication of results





RESULTS

- SERINC5 Inhibits HIV – 1 and MLV
- NEF acts against SERINC5
- SERINIC5 acts as anti – HIV – 1

```
#REPLICATION OF RESULTS
# ------------Code for Ratio of the infectivity vs different cell types plot--------
setwd("C:/Users/ABHIS/Downloads/biostats/datasets")
infect_data <- read.delim(file = "infect.txt", header = FALSE)
infect_data <- infect_data[order(infect_data$V2),]
infect_ratio <- infect_data$V2
names(infect_ratio) <- infect_data$V3
color_given <- function(value){
  if (value < 5) return("red")
  else if ( value < 25) return("green")
  else return("blue")
}
par(mar = c(6, 8, 4, 4))
color_vector <- unlist(lapply(infect_data$V2, color_given))
infect_data$V4 <- color_vector
barplot(infect_ratio, cex.names=0.6, horiz = TRUE,las=1, xlab = "Nef+/Nef- infectivity ratio", col=color_vector,
        main = "Ratio of the infectivity vs different cell types", border=0)
```

HYPOTHESIS 1

- All the 18 ITGAs play equal significant role in promoting viral infection.

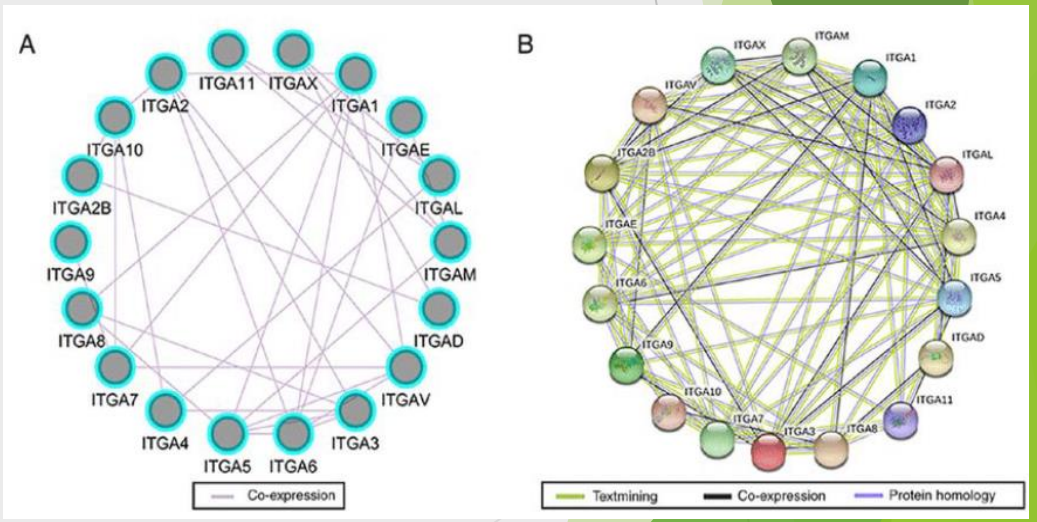
Diagnostic and prognostic values of integrin α subfamily mRNA expression in colon adenocarcinoma

YI-ZHEN GONG[1*], GUO-TIAN RUAN[1*], XI-WEN LIAO[2], XIANG-KUN WANG[2], CUN LIAO[1], SHUAI WANG[1] and FENG GAO[1]

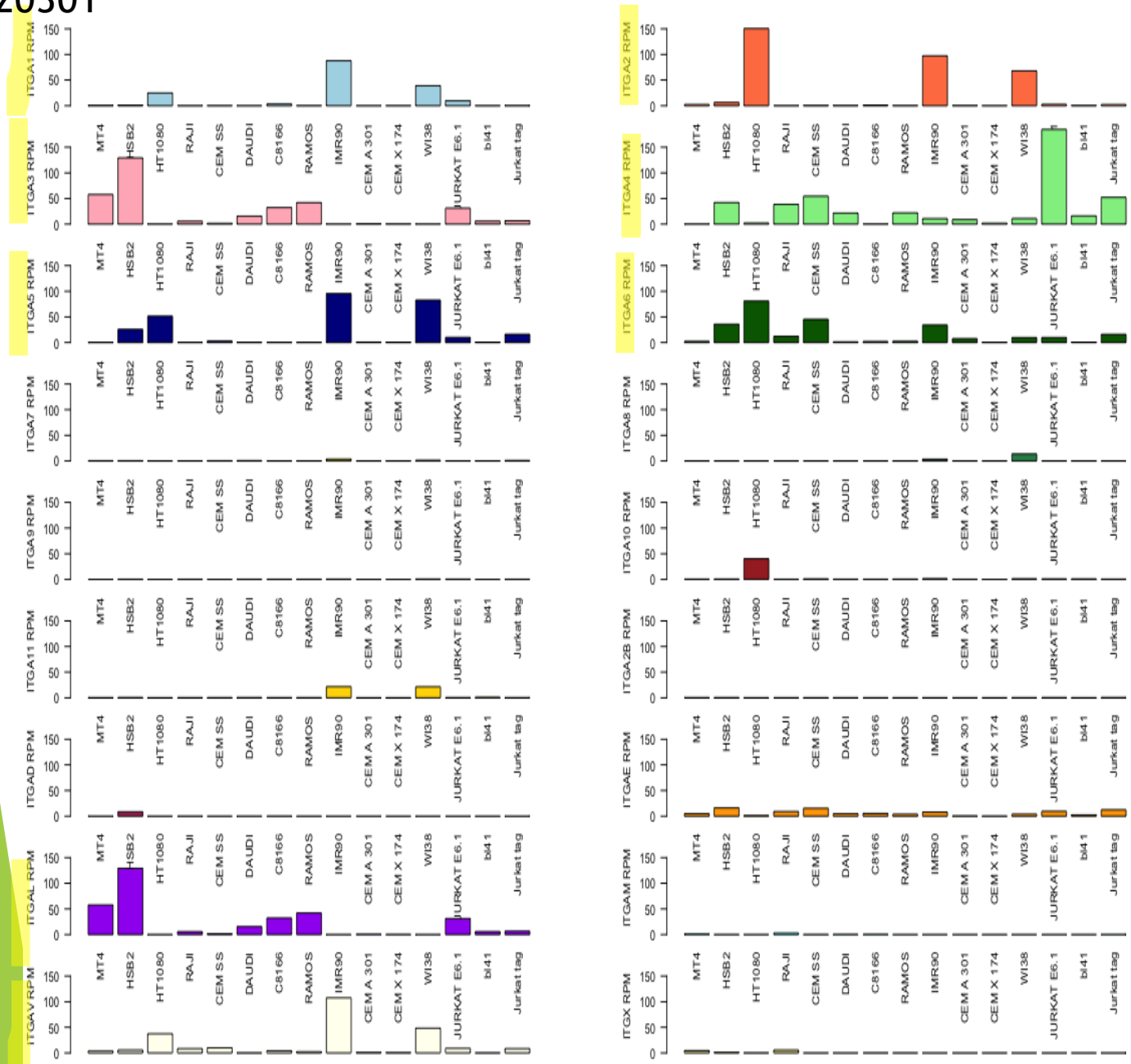
Departments of [1]Colorectal and Anal Surgery and [2]Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi Zhuang Autonomous Region 530021, P.R. China

Received February 14, 2019; Accepted June 26, 2019

DOI: 10.3892/or.2019.7216



20301



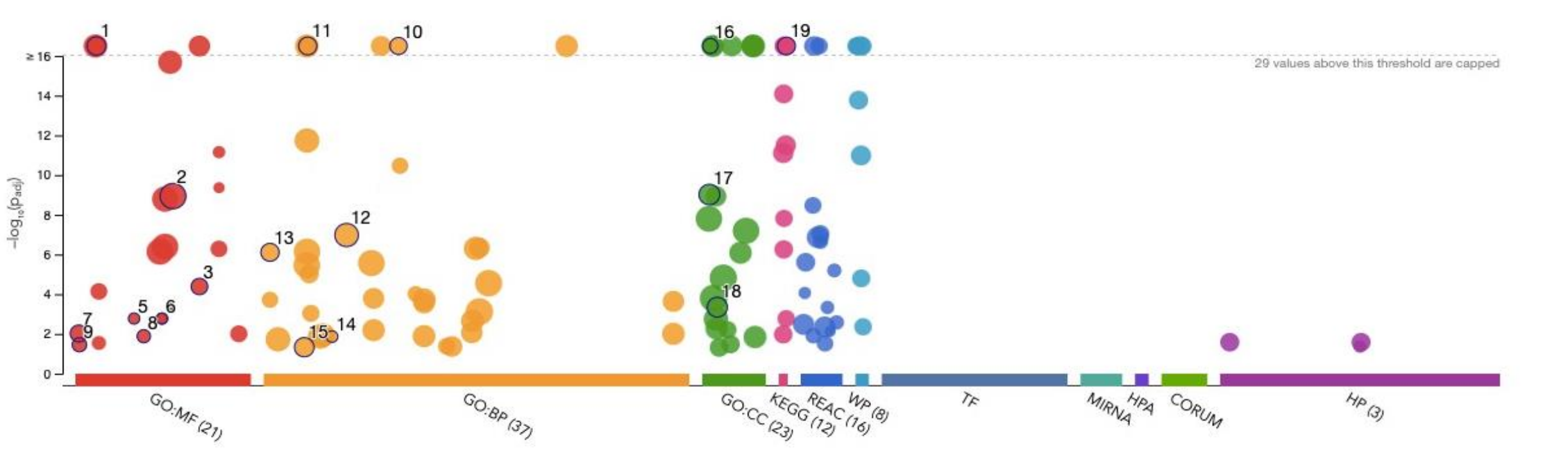
Plotting RPM value for 18 ITGAs and all the cell lines

```
#HYPOTHESIS 2
# Plot fewer plots in each row and column to increase the size of individual plots
par(mfrow = c(9, 2))  # Arrange plots in a 9x2 grid
# Find the maximum RPM value across all plots to set the same y-axis limit
max_rpm <- max(sapply(rpm_values, max))
for (j in 1:length(protein_names)) {
  protein <- protein_names[j]
  gene_rpm_values <- sapply(rpm_values, `[`, j)
  # Plot each barplot with the same y-axis limits and title on y-axis
  barplot(gene_rpm_values, names.arg = cell_lines, las = 2, col = colors[j],
          ylab = paste(protein, "RPM"), xlab = "", ylim = c(0, max_rpm))
}
```

20301,20225

BIOLOGICAL SIGNIFICANCE OF ITGAs

- ITGA has significant role in tumor and viral infection.
- G-PROFILER – To identify the biological importance of all ITGAs



20301,20225

ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	GO:MF	GO:0005178	integrin binding	1.583×10^{-34}
2	GO:MF	GO:0046872	metal ion binding	1.141×10^{-9}
3	GO:MF	GO:0050840	extracellular matrix binding	4.096×10^{-5}
4	GO:MF	GO:0038132	neuregulin binding	1.679×10^{-3}
5	GO:MF	GO:0019960	C-X3-C chemokine binding	1.679×10^{-3}
6	GO:MF	GO:0038064	collagen receptor activity	1.679×10^{-3}
7	GO:MF	GO:0001618	virus receptor activity	9.487×10^{-3}
8	GO:MF	GO:0031994	insulin-like growth factor I binding	1.304×10^{-2}
9	GO:MF	GO:0001846	opsonin binding	3.498×10^{-2}
10	GO:BP	GO:0033627	cell adhesion mediated by integrin	2.769×10^{-39}
11	GO:BP	GO:0007229	integrin-mediated signaling pathway	2.044×10^{-36}
12	GO:BP	GO:0016477	cell migration	1.047×10^{-7}
13	GO:BP	GO:0001704	formation of primary germ layer	7.781×10^{-7}
14	GO:BP	GO:0010668	ectodermal cell differentiation	1.354×10^{-2}
15	GO:BP	GO:0006909	phagocytosis	4.666×10^{-2}
16	GO:CC	GO:0008305	integrin complex	4.098×10^{-49}
17	GO:CC	GO:0005925	focal adhesion	9.527×10^{-10}
18	GO:CC	GO:0030667	secretory granule membrane	4.500×10^{-4}

Results:

- From the graph and after calculating p-value, result found was that all ITAGs functions with different intensities while viral replication.
- Mostly ITGAs upregulates viral infection.
- ITGA1, ITGA3 and ITGA4 are the most significant genes during HIV-1 infection.

	p value	
ITGAM	4.098*10-49	most significant
ITGA10	2.769*10-39	
ITGA11	2.044*10-36	
ITGA1	1.583*10-34	
ITGAV	9.527*10-10	
ITGA2	1.141*10-9	
ITGA2B	1.047*10-7	
ITGAD	7.781*10-7	
ITGA3	4.096*10-5	
ITGAX	4.500*10-4	
ITGA4	1679*10-3	
ITGA5	1.679*10-3	
ITGA6	1.679*10-3	
ITGA7	9.487*10-3	
ITGA8	1.304*10-2	
ITGAE	1.364*10-2	
ITGA9	3.498*10-2	
ITGAL	4.668*10-2	least significant

HYPOTHESIS 2

- ITGA3 genes upregulates post HIV-1 infection resulting in aiding in viral production.

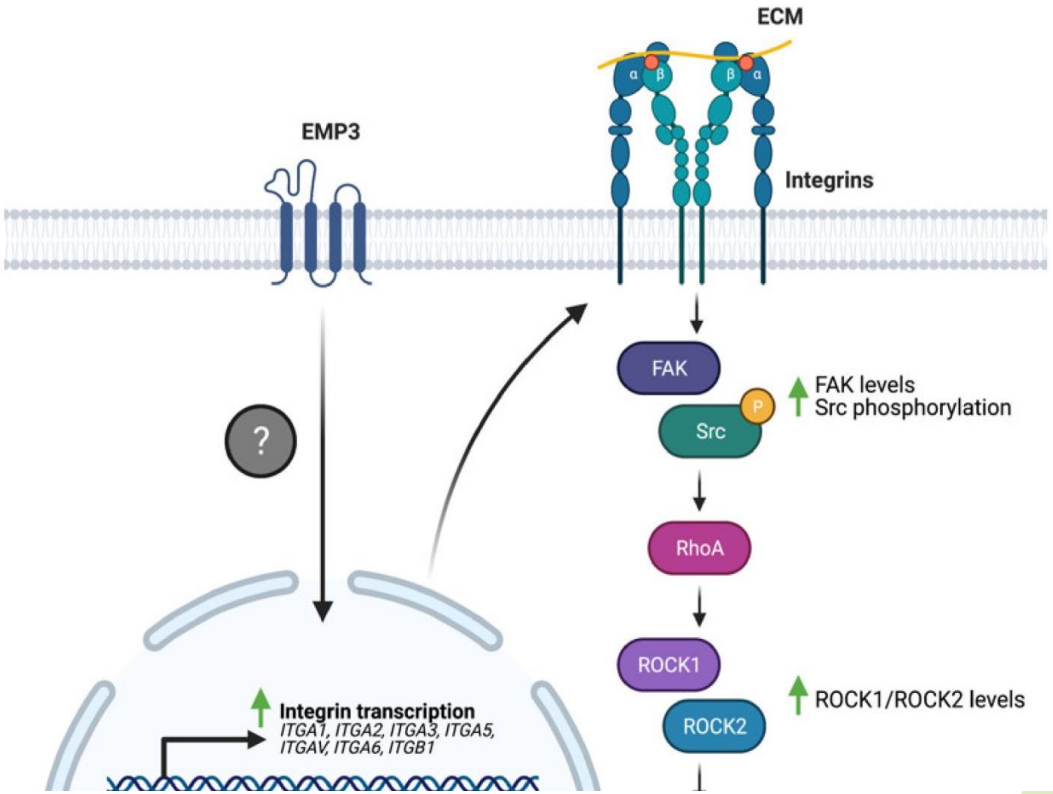
CD9 and ITGA3 are regulated during HIV−1 infection in macrophages to support viral replication

Zita Kruize [1], Viviana Cobos Jiménez [1], Fernando O Martinez [2], Riccardo Di Vincenzo [1], Karel A van Dort [1], Ad C van Nuenen [1], Thijs Booiman [1], Neeltje A Kootstra [3]

Affiliations + expand

PMID: 34242748 DOI: 10.1016/j.virol.2021.07.002

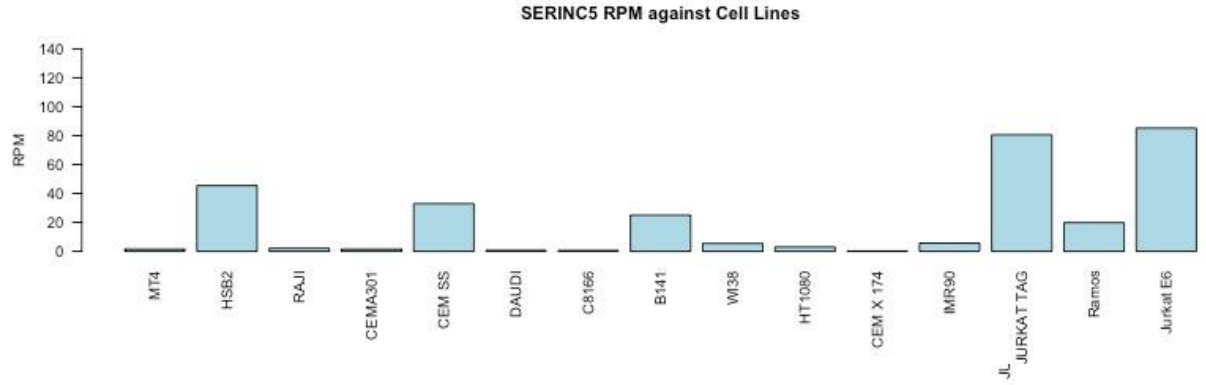
20018

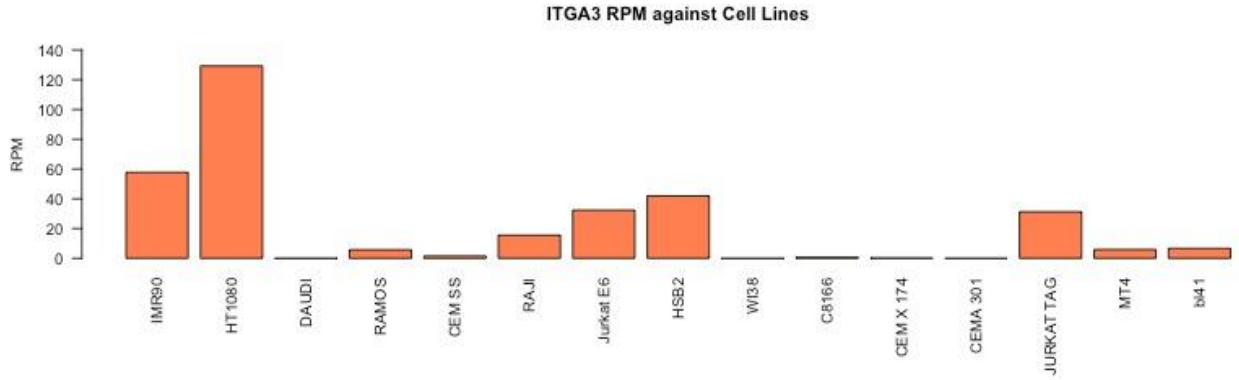


Biological Significance

- ITGA3 is a cell surface receptor protein belonging to the integrin family,
- ITGA3 upregulation following HIV-1 infection aids in viral production by promoting cell-to-cell interactions, modulating cellular signaling, and altering the cellular microenvironment to favor viral replication and spread.

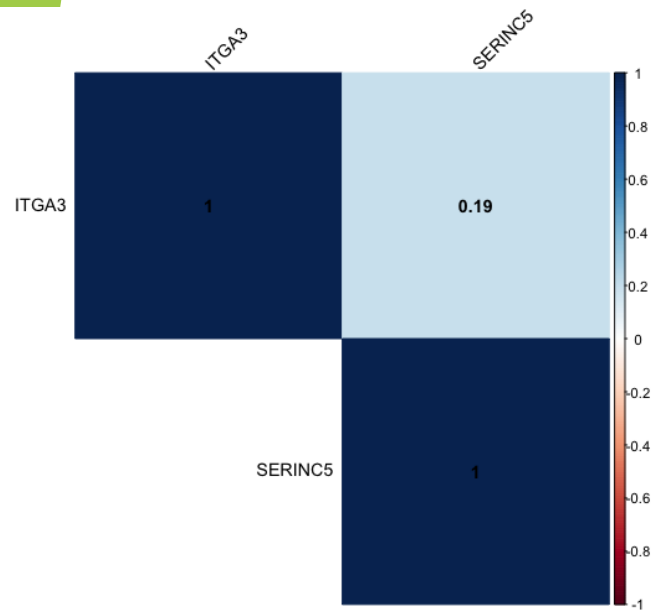
20018





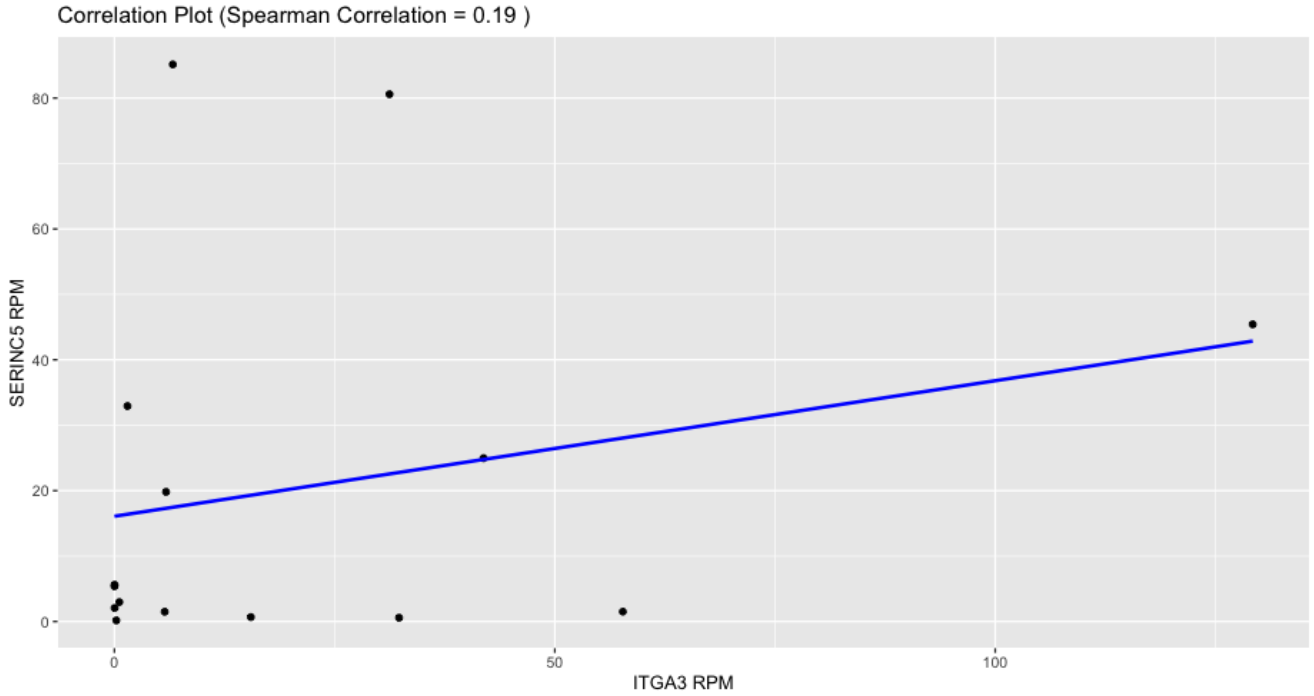

```
84  # Running Shapiro-Wilk test
85  shapiro_test_result <- shapiro.test(infect_data$V2)
86
87  # Printing the result
88  print(shapiro_test_result)
89
90  Shapiro-Wilk normality test
91
92  data:  infect_data$V2
93  W = 0.76301, p-value = 0.00128
```

Expression of SERINC5 and ITGA3 across different cell lines



Correlation plot between ITGA3 and SERINC5

```
        Spearman's rank correlation rho

data:  serinc5 and itga3_rpm
S = 452, p-value = 0.4901
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1928571
```




```r
#Correlation plot
# Load necessary libraries
library(ggplot2)

# Create a data frame with RPM values for ITGA3 and SERINC5
rpm_data <- data.frame(ITGA3 = itga3_rpm, SERINC5 = serinc5_rpm)

# Calculate correlation coefficient
correlation <- cor(itga3_rpm, serinc5_rpm, method = "spearman")

# Plot correlation plot
ggplot(rpm_data, aes(x = ITGA3, y = SERINC5)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = paste("Correlation Plot (Spearman Correlation =", round(correlation, 2), ")"),
       x = "ITGA3 RPM", y = "SERINC5 RPM")
#------
# Load necessary libraries
library(corrplot)

# Create a data frame with RPM values for ITGA3 and SERINC5
rpm_data <- data.frame(ITGA3 = itga3_rpm, SERINC5 = serinc5_rpm)

# Calculate the correlation matrix
correlation_matrix <- cor(rpm_data, method = "spearman")

# Plot correlation matrix heatmap
corrplot(correlation_matrix, method = "color", type = "upper",
         addCoef.col = "black", tl.col = "black", tl.srt = 45)
```

Results

- ITGA3 upregulates HIV-1infection and facilitates its entry in the host cell.
- A weakly positive relationship between ITGA3 and SERINC5, meaning that alterations in SERINC5 expression will impact the upregulation of ITGA3 in various cell lines when SERINC5 is present.
-

HYPOTHESIS 3

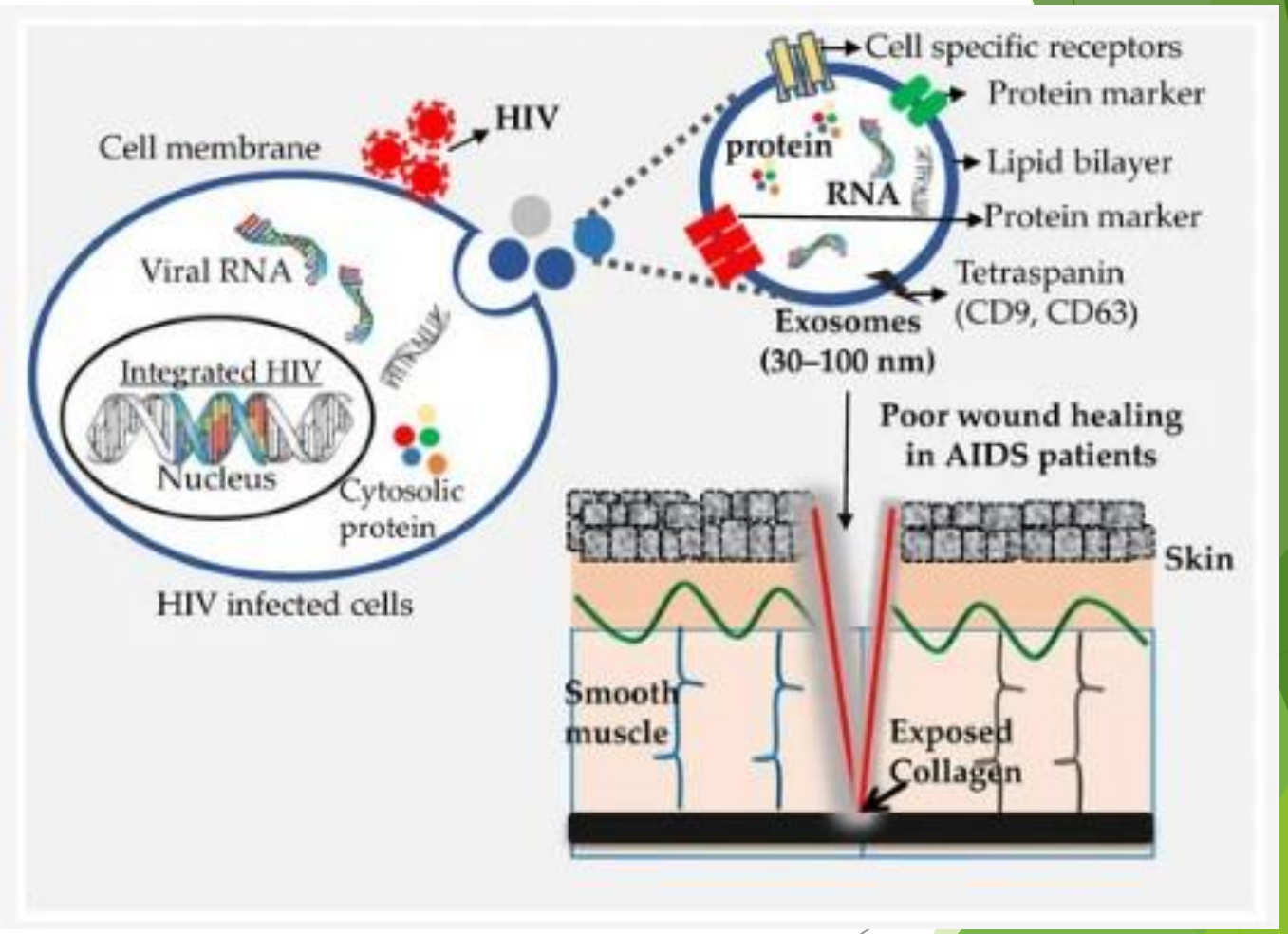
CD9 upregulates HIV infection by correlating negatively with SERINC5.

Tetraspanins CD9 and CD81 modulate HIV-1-induced membrane fusion

Mónica Gordón-Alonso [1], María Yañez-Mó, Olga Barreiro, Susana Alvarez, M Angeles Muñoz-Fernández, Agustín Valenzuela-Fernández, Francisco Sánchez-Madrid

Affiliations + expand

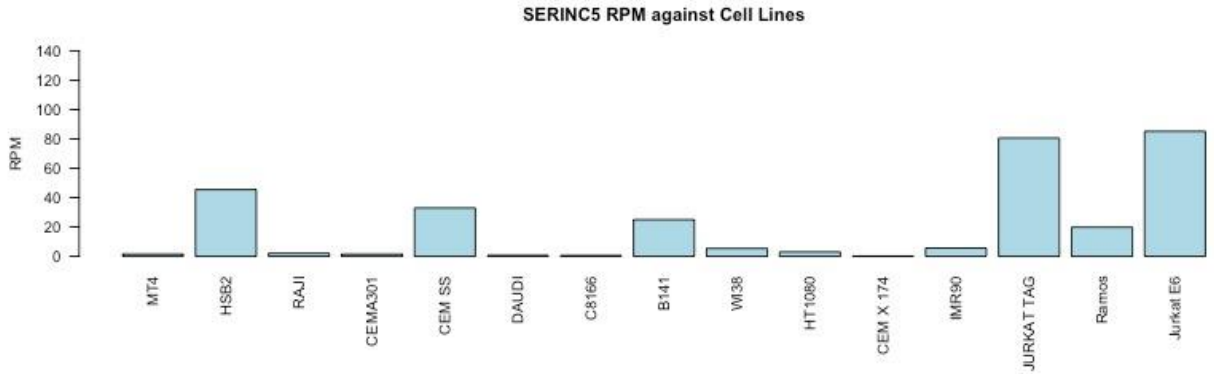
PMID: 17015697 DOI: 10.4049/jimmunol.177.8.5129

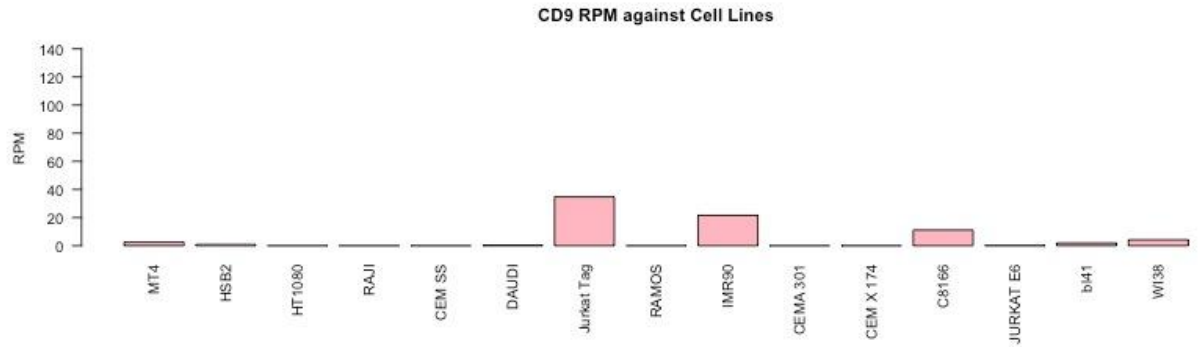


Biological Significance

- CD9 is a member of the tetraspanin family of cell surface proteins.
- CD9 can interact with HIV-1 particles and facilitate their entry into target cells.
- This interaction may involve CD9 serving as a co-receptor or facilitating viral fusion with the host cell membrane, thus promoting HIV-1 infection.

CD9 expression in different cell lines





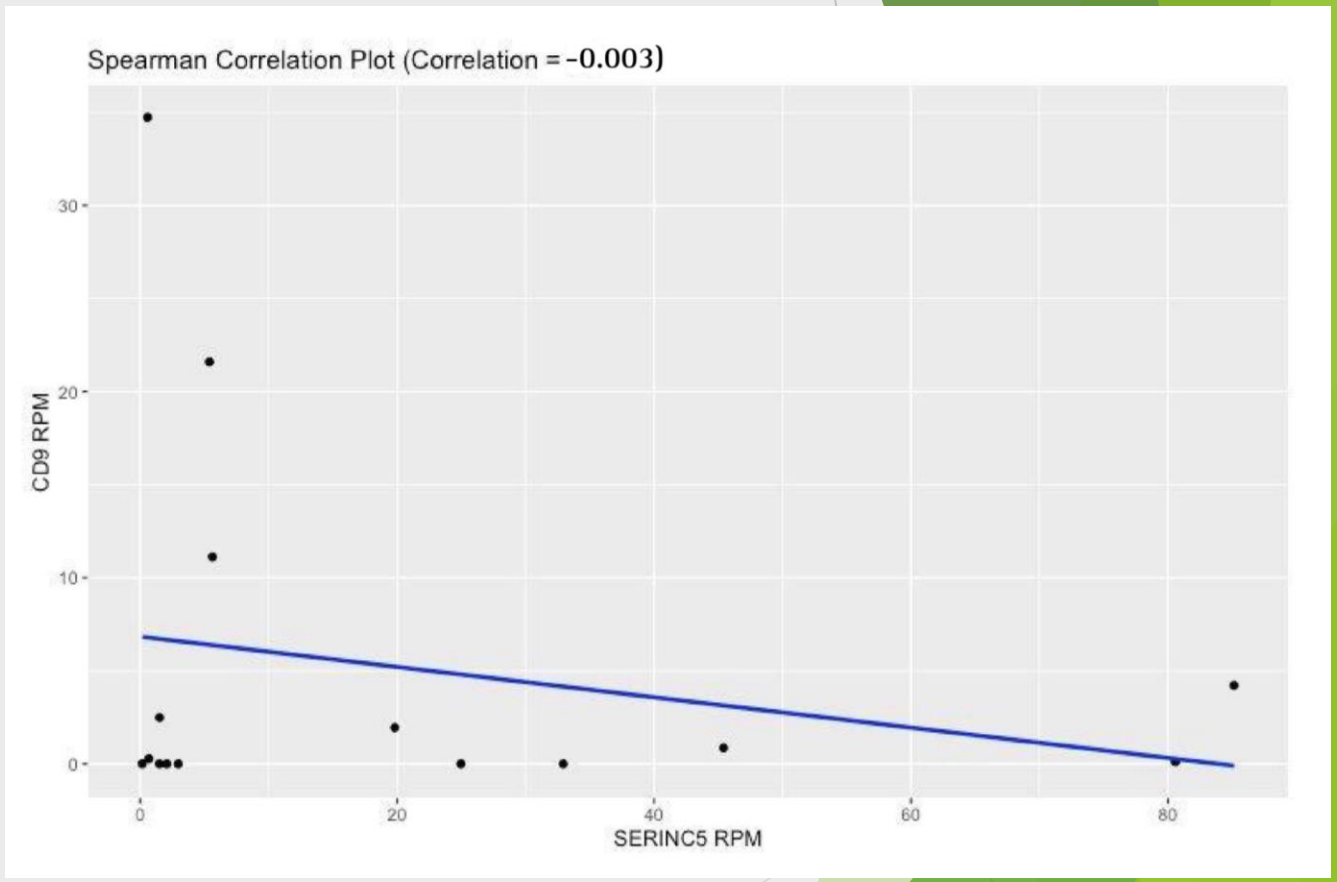

```
> shapiro_df
     Protein W.W                  p_value
[1,] "CD9"   "0.595405155906628" "2.31148599601429e-05"
```

```r
# Define colors for each graph
colors <- c("lightpink", "red")

# Set custom margin
par(mar = c(5, 4, 4, 2))  # margin: bottom, left, top, right

# Increase the plot region size
par(plt = c(0.1, 0.9, 0.1, 0.9))  # plot region: xmin, xmax, ymin, ymax

# Plot fewer plots in each row and column to increase the size of individual plots
par(mfrow = c(9, 2))  # Arrange plots in a 9x2 grid

# Find the maximum RPM value across all plots to set the same y-axis limit
max_rpm <- max(sapply(rpm_values, max))

for (j in 1:length(protein_names)) {
  protein <- protein_names[j]
  gene_rpm_values <- sapply(rpm_values, '[', j)

  # Plot each barplot with the same y-axis limits and title on y-axis
  barplot(gene_rpm_values, names.arg = cell_lines, las = 2, col = colors[j],
          ylab = paste(protein, "RPM"), xlab = "", ylim = c(0,max_rpm))
}
```

Data is not normally distributed

Correlation plot between CD9 and SERINC5

The estimated Spearman correlation coefficient (rho) is close to 0 (-0.0036), suggesting that there is little to no monotonic relationship between the RPM values of SERINC5 and CD9.




```
        Spearman's rank correlation rho

data:  serinc5_rpm and cd9_rpm
S = 562.04, p-value = 0.9897
alternative hypothesis: true rho is not equal to 0
sample estimates:
         rho
-0.003636965
```

RESULT:

- CD9 upregulates HIV-1 infection and facilitates its entry into the host cell.

- Rho value and p-value suggest that there may not be a significant correlation between CD9 and SERINC5 and thus changes in SERINC5 expression will not affect the expression of CD9 in different cell lines present in the data.

HYPOTHESIS 4

- CD9 promotes HIV-1 infection, while CD4 counters HIV-1 Nef, suggesting opposing roles in viral pathogenesis and host defense.

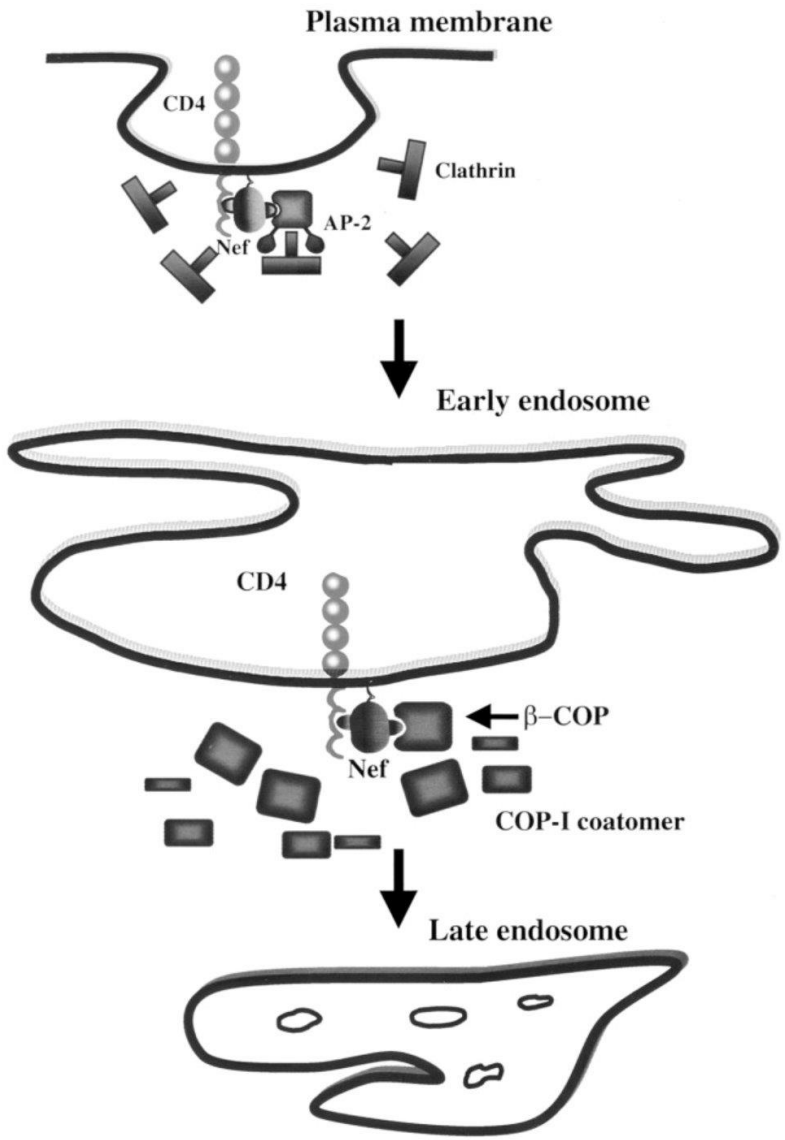
Cell–Extrinsic Priming Increases Permissiveness of CD4+ T Cells to Human Immunodeficiency Virus Infection by Increasing C–C Chemokine Receptor Type 5 Co–receptor Expression and Cellular Activation Status

Jesper G Pedersen [1], Johanne H Egedal [1] [2], Thomas A Packard [2], Karthiga Thavachelvam [1], Guorui Xie [2] [3], Renée Marije van der Sluis [1] [4], Warner C Greene [2], Nadia R Roan [2] [3], Martin R Jakobsen [1]

Affiliations + expand

PMID: 34899645 PMCID: PMC8661899 DOI: 10.3389/fmicb.2021.763030

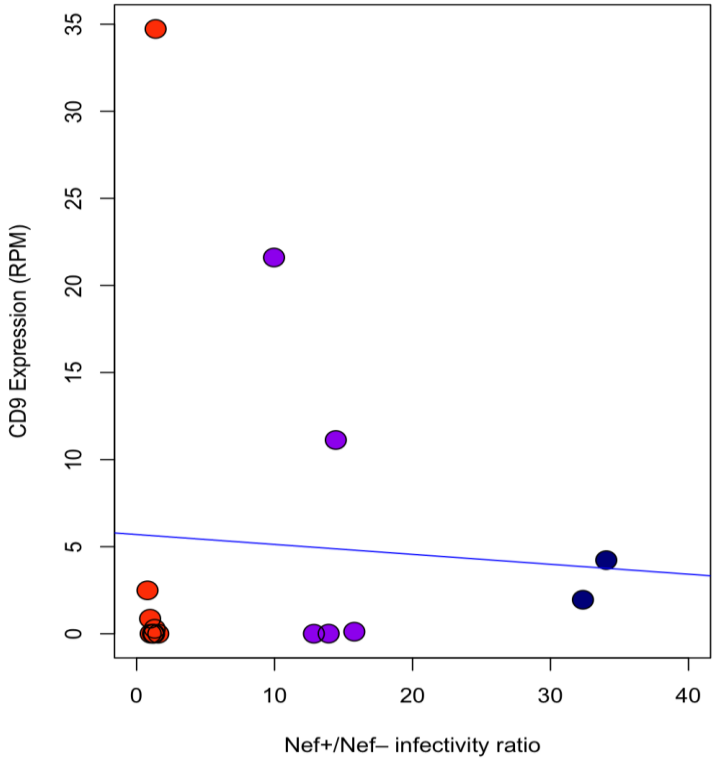
20302

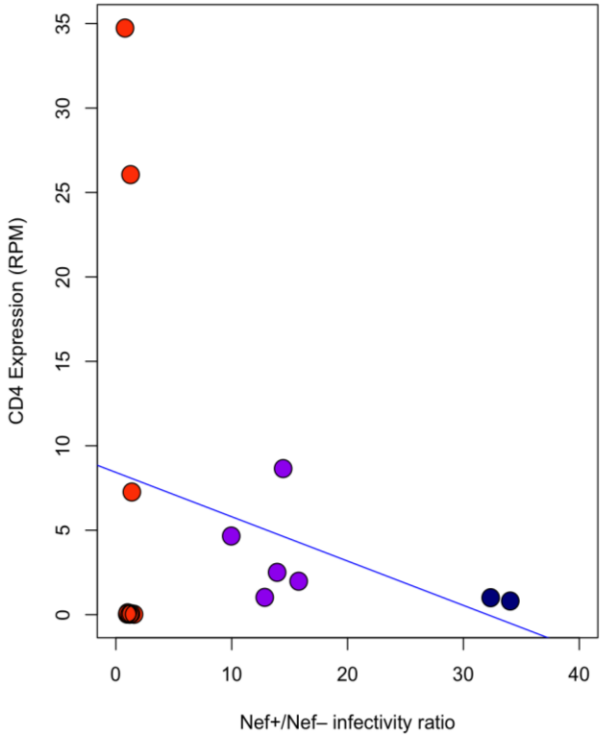


Biological Significance

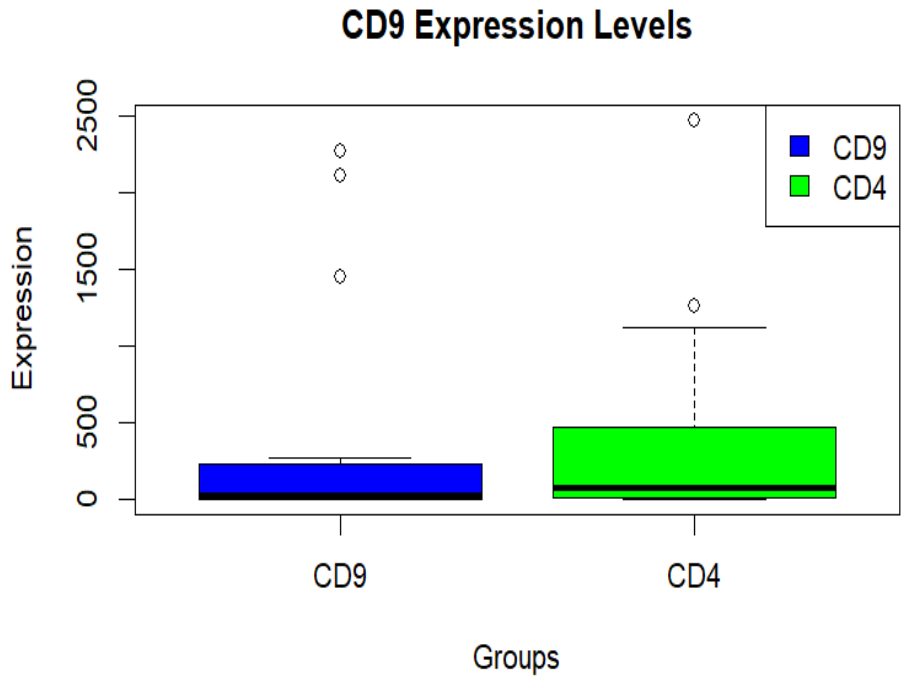
- CD4 is the primary receptor for HIV-1 entry into host cells, particularly CD4+ T lymphocytes.
- The interaction between CD4 and Nef is crucial in HIV-1 pathogenesis.
- CD4 counters Nef's activity by competing for binding or interfering with Nef-mediated downregulation, thereby maintaining CD4 expression on the cell surface.
- CD9 serves as a co-receptor and facilitates viral fusion with the host cell membrane, thus promoting HIV-1 infection

Expression of CD9 and CD4 against Nef+/Nef- infectivity ratio



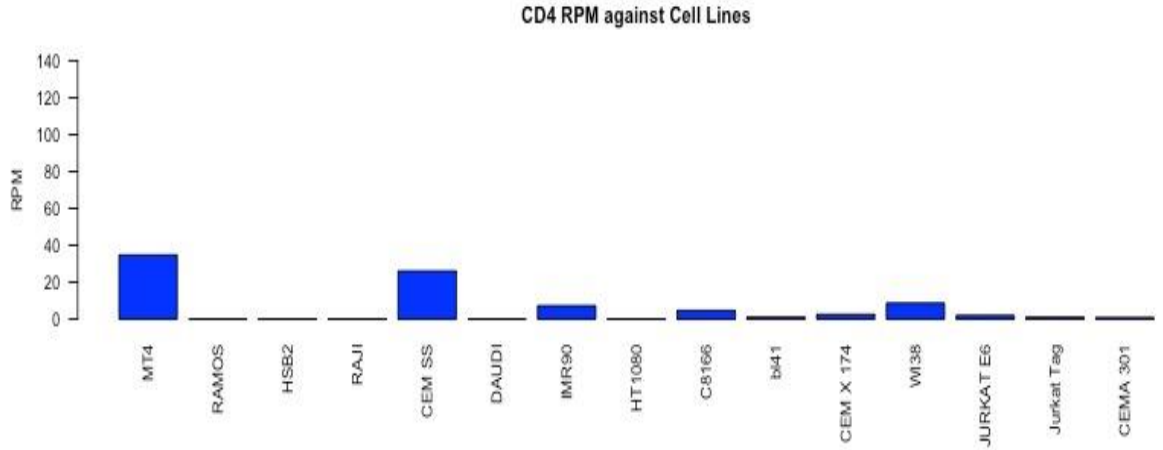


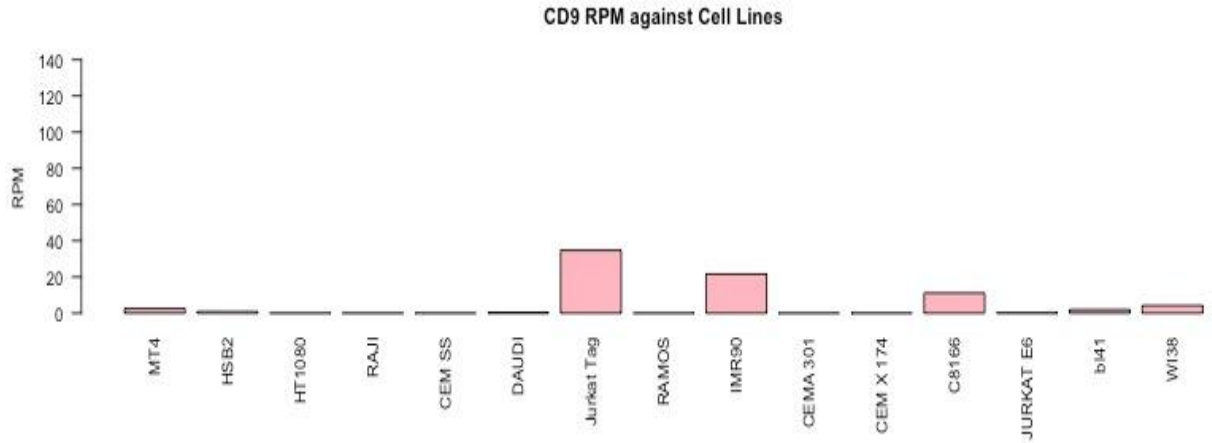
Scatter plot



Boxplot

20302





Expression of CD4 and CD9 across different cell lines

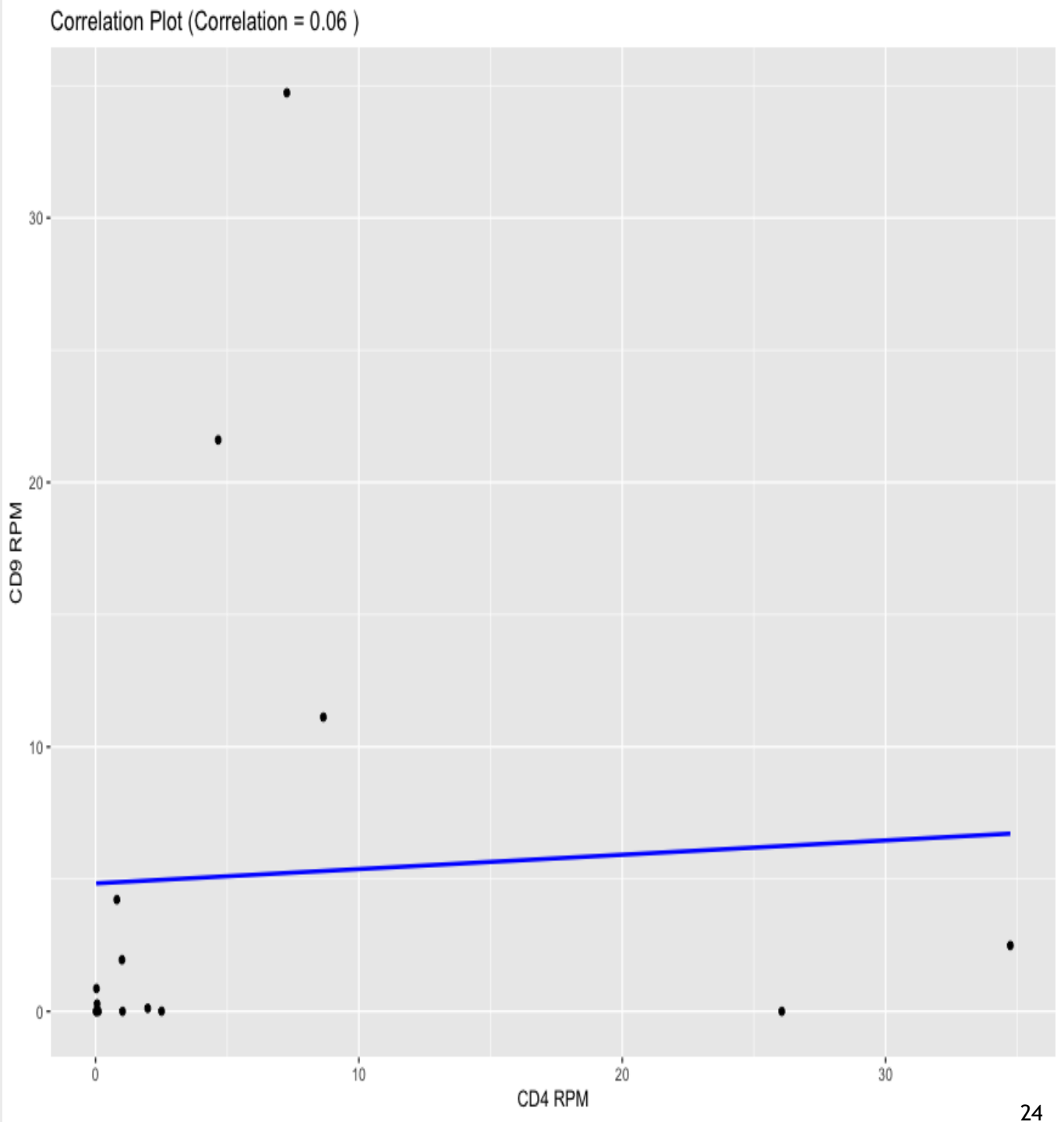
```
> shapiro_df
     Protein W.W                  p_value
[1,] "CD9"   "0.595405155906628" "2.31148599601429e-05"
[2,] "CD4"   "0.62383214785954"  "4.25027229001761e-05"
```

Data is not normally distributed

Correlation Plot between CD9 and CD4

Spearman Correlation test

```
        Spearman's rank correlation rho

data:  cd4_rpm and cd9_rpm
S = 311.52, p-value = 0.09757
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.06
```



Results:

- CD9 promotes HIV-1 infection, while CD4 counters HIV-1 Nef.
- CD4 and CD9 works oppositely independently without affecting expression of each other.
- Rho value and p-value suggest that there is no significant correlation between CD4 and CD9 and thus changes in CD4 expression will not affect the expression of CD9 in different cell lines or vice versa.

HYPOTHESIS 5

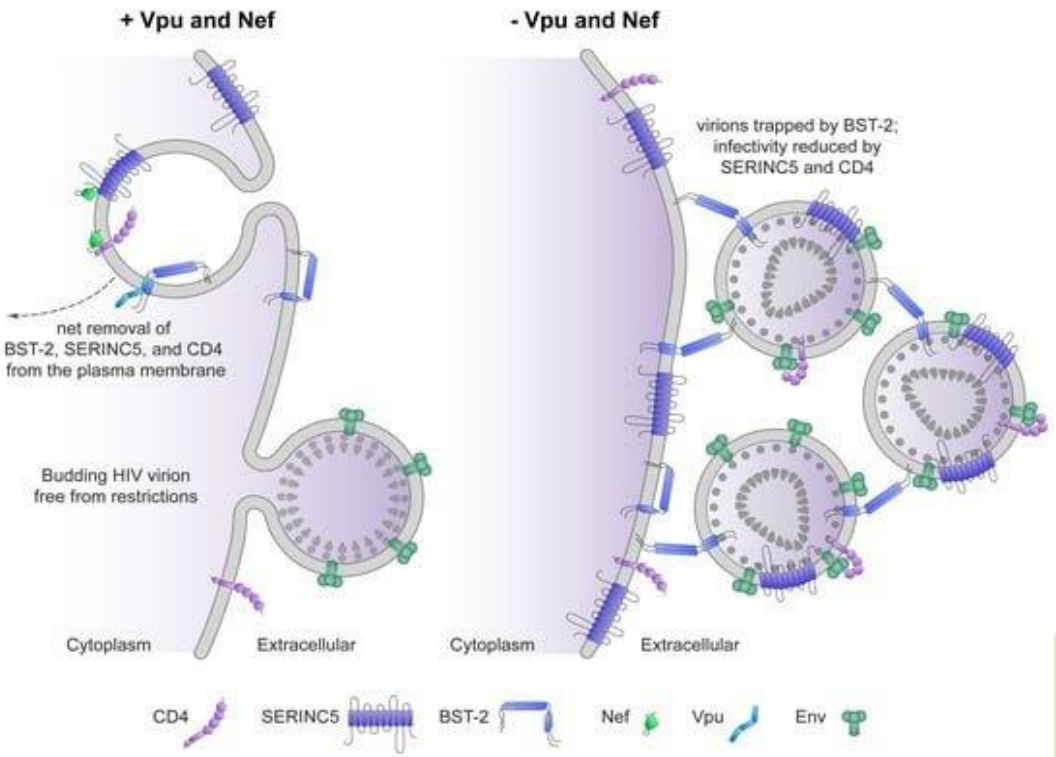
BST –2 counters HIV infection by correlating with SERINC5

The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein

Nanette Van Damme [1], Daniel Goff, Chris Katsura, Rebecca L Jorgenson, Richard Mitchell, Marc C Johnson, Edward B Stephens, John Guatelli

Affiliations + expand

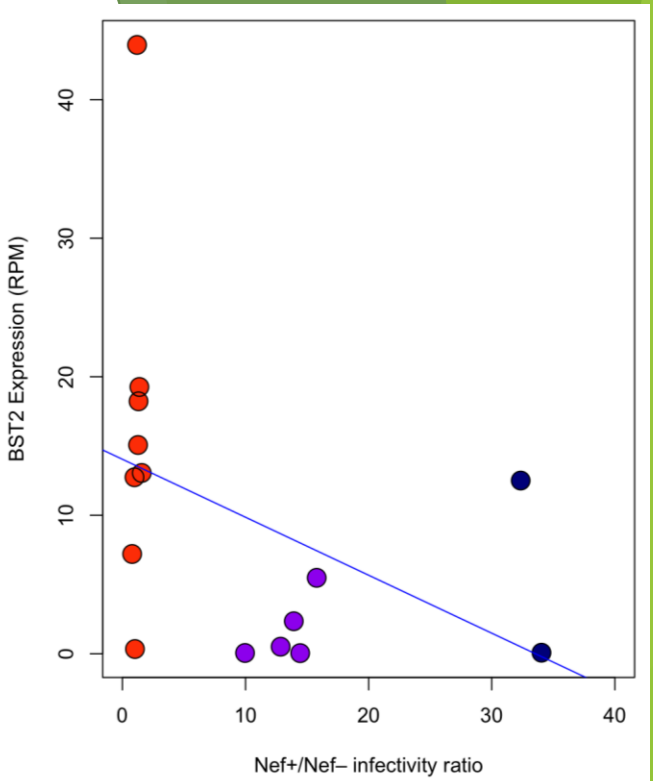
PMID: 18342597 PMCID: PMC2474773 DOI: 10.1016/j.chom.2008.03.001



Biological Significance

- BST-2 is a host protein known for its ability to restrict the release of HIV-1 virions from infected cells.
- It functions by tethering newly formed viral particles to the plasma membrane, preventing their release and thereby limiting viral spread.

Expression of BST2 against Nef+/Nef- infectivity ratio



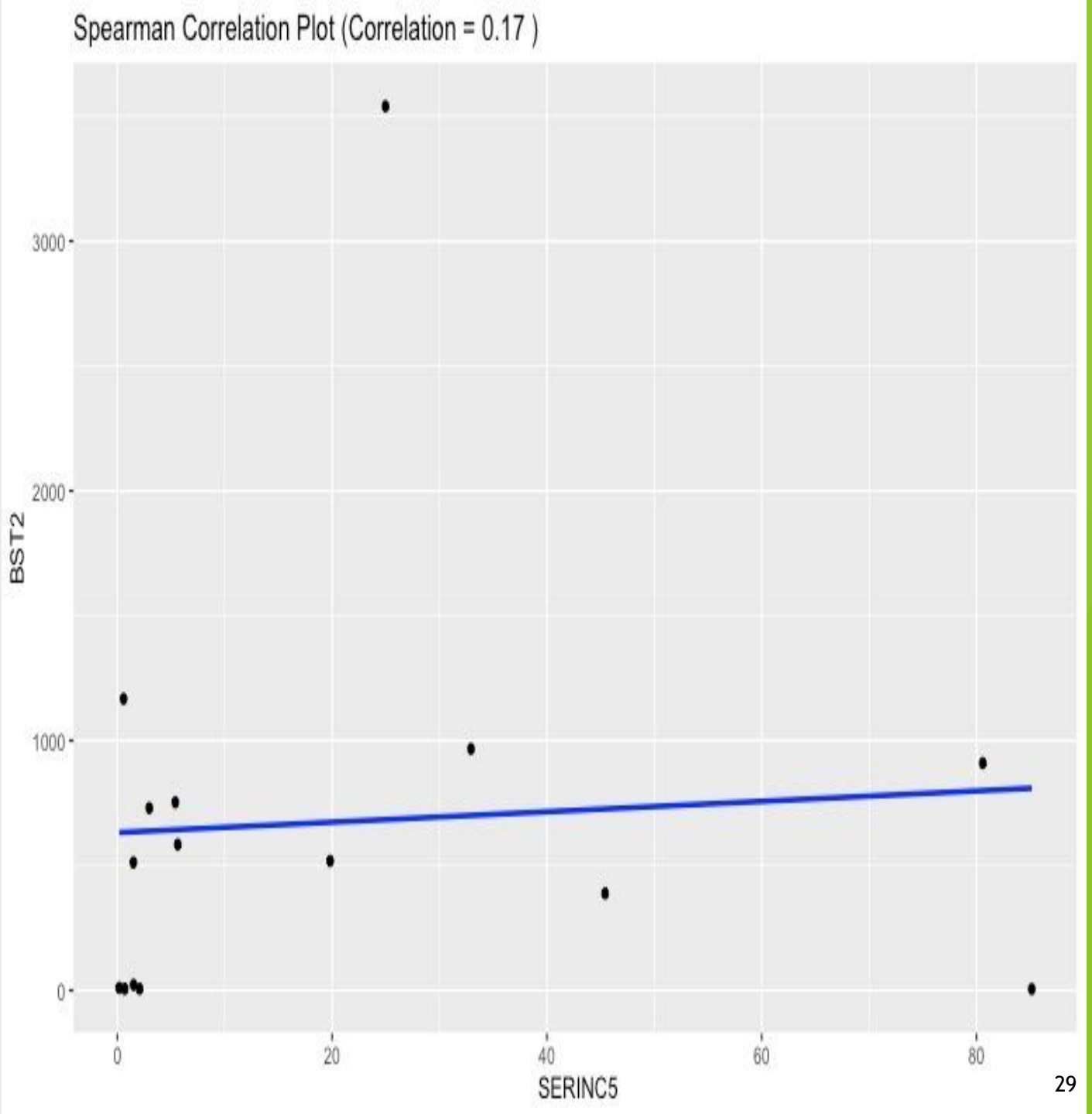

```r
#HYPOTHESIS 5
#Plotting
plot(infect_data$V2,Expression$V3, xlim=c(0,40), xlab = "Nef+/Nef- infectivity ratio", ylab = "BST2 Expression (RPM)", cex=2, col = c(rep("orangered",8),rep("purple",5),rep("darkblue",2)), pch = 16)
points(infect_data$V2,Expression$V3, col = "black", pch = 1, cex =2)


# Perform linear regression
lm_model <- lm(infect_data$V2 ~ Expression$V3)

# Extract R-squared value and p-value
r_square <- summary(lm_model)$r.squared
p_value <- summary(lm_model)$coefficients[2,4]

# Plot the data
pdf("Figure1d.pdf", width = 10, height=8)
par(mar = c(10, 10, 10, 10)) #margins

plot(infect_data$V2, Expression$V3, xlim=c(0,40), xlab = "Nef+/Nef- infectivity ratio",
     ylab = "BST2 Expression (RPM)", cex=2, col = c(rep("orangered",8),rep("purple",5),rep("navy",2)), pch = 16)
points(infect_data$V2, Expression$V3, col = "black", pch = 1, cex =2)

# Add the trendline
abline(lm_model)

# Add text with r-square and p-value
text(x = 2.5, y = 75, paste("R^2:", round(r_square, digits = 5)), pos = 4)
text(x = 2.5, y = 70, paste("p-value:", format(p_value, scientific = TRUE, digits = 2)), pos = 4)

dev.off()
```

```
        Shapiro-Wilk normality test

data:  bst2
W = 0.688, p-value = 0.0001855
```

Data is not normally distributed

Correlation plot between SERINC5 and BST2

```
        Spearman's rank correlation rho

data:  serinc5 and bst2
S = 462.91, p-value = 0.5366
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1733691
```



Results:

- BST-2 downregulates HIV-1 infection

- Rho value and p-value suggest that there may not be a significant relationship correlation between BST-2 and SERINC5 and thus changes in SERINC5 expression will not affect the expression of BST-2 in different cell lines.

30

THANK YOU