

UNIVERSIDADE FEDERAL DE MINAS GERAIS
DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

**CONSTRUÇÃO DA VIA DE INTERAÇÃO DO VÍRUS EBOLA
COM O HOSPEDEIRO E ESTIMATIVA DA ORIGEM
DOS GENES E PROCESSOS**

ÉLISSON NOGUEIRA LOPES

BELO HORIZONTE

2018

ÉLISSON NOGUEIRA LOPES

**CONSTRUÇÃO DA VIA DE INTERAÇÃO DO VÍRUS EBOLA
COM O HOSPEDEIRO E ESTIMATIVA DA ORIGEM
DOS GENES E PROCESSOS**

Dissertação de mestrado apresentada
ao Programa de Pós-Graduação em
Bioinformática da Universidade
Federal de Minas Gerais como
requisito parcial para a obtenção do
título de Mestre em Bioinformática

Orientador: Dr. José Miguel Ortega

Co-orientador: Dr. Rodrigo Siqueira Dalmolin

BELO HORIZONTE

2018

043

Lopes, Élisson Nogueira.

Construção da via de interação do vírus ebola com o hospedeiro e estimativa da origem dos genes e processos [manuscrito] / Élisson Nogueira Lopes. - 2018.

85 f. : il. ; 29,5 cm.

Orientador: Dr. José Miguel Ortega. Co-orientador: Dr. Rodrigo Siqueira Dalmolin.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas.

1. Ebola Vírus. 2. Evolução. 3. Redes e Vias Metabólicas. 4. Bioinformática - Teses. I. Ortega, José Miguel. II. Dalmolin, Rodrigo Siqueira. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004

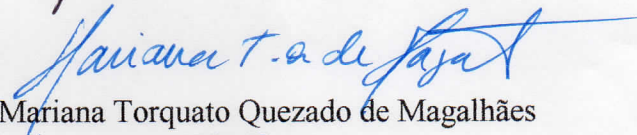


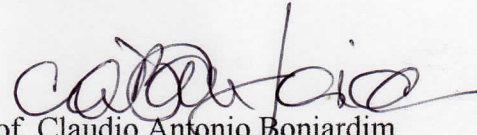
"CONSTRUÇÃO DA VIA DE INTERAÇÃO DO VÍRUS EBOLA COM O HOSPEDEIRO E ESTIMATIVA DA ORIGEM DOS GENES E PROCESSOS"

Élisson Nogueira

Dissertação aprovada pela banca examinadora constituída pelos Professores:


Prof. José Miguel Ortega - Orientador
UFMG


Profa Mariana Torquato Quezado de Magalhães
UFMG


Prof. Claudio Antonio Bonjardim
UFMG

Belo Horizonte, 27 de fevereiro de 2018.

*“Vai, vai lá, não tenha medo do pior
Eu sei que tudo vai mudar
Você vai transformar o mundo ao seu redor
Mas não vacila, moleque de vila...”
PROJOTA- Moleque de Vila*

AGRADECIMENTOS

Nem uma obra de valor foi feita sozinha, por mais talentosa que possa ser as mãos do artista, ainda foi necessário as mãos talentosas de quem primeiro fez a tela. A vida é assim, todo feito considerável é realizado através da contribuição de um grupo de pessoas. Este trabalho dedico a vocês.

Primeiramente a minha família, pai, mãe e irmã, que são a base de quem sou, que foi construído desde minha infância, que estão ao meu lado nas tempestades, e hoje, desfrutaremos juntos deste dia alegre de sol. Meus tios e padrinhos Gismar e Jaqueline, que abriram as portas da sua casa e coração e me acolheram em uma parte desta jornada. Meus amigos e irmãos, que me ajudaram neste novo desafio longe de casa, Samuel, Deivid Max e Felipe, me apoiando em todos os momentos.

Agradeço agora aos meus instrutores, onde cada um teve um peso importante nesta jornada. Os meus primeiros orientadores Ana Paula, Edina e Fabrício que me ensinaram sobre o quão fantástico o mundo pode ser, pelos olhos da ciência. Aos meus grandes amigos da Pós graduação em Bioinformática da UFMG e de Natal, Professor Dalmolin, mestre Iara e Diego, e todos do BioME sem distinção, que foram pacientes e atenciosos em me ajudar a aprender um pouco sobre o trabalho de vocês.

Em especial a todos da minha segunda família, Biodados, Professor Miguel que muitas vezes já o comparamos ao nosso pai, por tamanho suporte. Meus companheiros de café e do dia a dia, que me auxiliaram em tudo, Lissur praticamente uma segunda orientadora, Carlos que sempre era amolado nos seus momentos de folga e atendia, Rayson um grande amigo e segunda voz da nossa dupla sertaneja, Fenícia que devo algumas cervejas pela tamanha ajuda, Gilson um mestre sempre alegre assim como Ricardo (Piá), *Sensei* Tetsu uma inspiração e amigo desde o primeiro dia no laboratório.

A Deus meu guia, meu maior agradecimento, por não me desamparar, por me dar forças as batalhas diárias e alcançar esse momento de glória.

RESUMO

O vírus Ebola (EBOV) é um vírus envelopado, filamentososo, e que contém um genoma de RNA de fita negativa. O EBOV pertence à família *Filoviridae*, e é causador de uma doença devastadora, com uma taxa de mortalidade de cerca de 50-90%. Os primeiros sintomas desenvolvidos por pacientes infectados são febre, mal-estar e dor muscular e podem ser seguidos por sangramento e falha de órgãos e tecidos. Os alvos iniciais do Ebola são macrófagos e células dendríticas, mas sua capacidade de infecção abrange todos os tipos de células, com exceção de linfócitos. O Ebola conecta-se através da membrana plasmática e, depois disso, a glicoproteína viral induz a absorção por endocitose. O processo é dependente da ação das proteínas da superfície celular. Após a absorção, as partículas viajam em compartimentos onde a glicoproteína viral é clivada e fundida na membrana endossomal, o que resulta na liberação dos compartimentos virais no citoplasma do hospedeiro. Foi feito o estudo do mecanismo de infecção do EBOV e a coleta das proteínas hospedeiras conhecidas até o momento, por participar diretamente ou indiretamente na infecção. Esta abordagem de mineração recrutou 133 proteínas hospedeiras. Com elas, foi construído uma via completa do interatoma para representar o ciclo de infecção do Ebola nos hospedeiros. Foi feita a análise também dos homólogos de cada proteína humana coletada ao longo da árvore taxonômica, para inferir seu clado/época de origem. Os resultados da análise de origem evolutiva permitiram inferir que o vírus poderia infectar desde *Euleostomi*, sugerindo que animais como peixes e pets poderiam ser infectados e retransmitir o vírus para outros hospedeiros, como o homem. Além disso, analisamos duas séries de dados GEO para expressão gênica após a infecção pelo Ebola, através do enriquecimento pelo KEGG, criando um perfil de infecção viral. Inicialmente, processos envolvendo controle celular e metabolismo foram enriquecidos, seguidos de vários outros processos de imortalização e supressão do sistema imune. Em conclusão, a infecção pelo Ebola ocorre com a interação com proteínas antigas na membrana e internas, algumas porém são mais novas. A construção da via possibilita a criação de um estudo sistemático a infecção do vírus e quais seriam os possíveis alvos terapêuticos.

Palavra-chave: ebola, evolução, expressão, vias metabólicas.

ABSTRACT

Ebola virus (EBOV) is an enveloped, filamentous virus that contains a negative-stranded RNA genome. EBOV belongs to the Filoviridae family, and causes a devastating disease, with a mortality rate of about 50-90%. The first symptoms developed by infected patients are fever, malaise and muscle pain and can be followed by bleeding and failure of organs and tissues. The initial targets of Ebola are macrophages and dendritic cells, but their ability to infect all types of cells, with the exception of lymphocytes. Ebola connects through the plasma membrane and, thereafter, a viral glycoprotein induces uptake by endocytosis. The process is dependent on the action of the proteins of the cell life. After absorption, as particles travel in compartments where a viral glycoprotein is cleaved and fused to the endosomal membrane, which results in the release of viral compartments without host cytoplasm. The study of the mechanism of infection of the EBOV and a collection of the hosted proteins was carried out, by participating directly or indirectly in the infection. This mining approach recruited 133 host proteins. With them, a complete interactive route was constructed to represent the Ebola infection cycle in hosts. An analysis was also made of the homologues of each human protein collected along the taxonomic tree to infer its clade/epoch the origin. The results of the analysis of evolutionary origin have allowed inferring that the virus can infect from Euleostomi, suggesting that animals such as fish and pets might be infected and retransmit the virus to other hosts, such as man. In addition, we analyzed two datasets, through KEGG enrichment, creating a profile of viral infection. Initially, processes involving cellular control and metabolism were enriched, followed by several other processes of immortalization and suppression of the immune system. In conclusion, an Ebola infection occurs with an interaction with old membrane and internal proteins, some still are younger. The construction of the pathway allows the creation of a systematic study of virus infection and what the possible therapeutic targets might be.

Palavra-chave: ebola, evolution, expression, metabolic pathways.

LISTA DE ABREVIATURAS

ARCoBALeno	Application for Rapidly COloring Biological pathways by Ancestrality or gene ontoLogy-Enriched iNformatiOn
BBH	Bidirectional Best Hits
BLAST	Basic Local Alignment Search Tool
BOWS	Bioinformatics Open Web Services
COG	Clusters of Orthologous Groups
ELDO	External Least Divergent Ortholog
GEO	Gene Expression Omnibus
GO	Gene Ontology
GPML	Graphical Pathway Markup Language
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Ontology
LBP	Last Biological Process
LCA	Lowest Common Ancestor (menor ancestral comum)
LDO	Least Divergent Ortholog
LMF	Last Molecular Function
NCBI	National Center for Biotechnology Information
PESCADOR	Platform for Exploration of Significant Concepts Associated to co-Occurrence Relationships
PMID	PubMed Unique Identifier
PSI-BLAST	Position-Specific Iterated Blast
UEKO	UniRef50 Enriched KEGG Orthology
Uniprot	Universal Protein Resource
UniRef	UniProt Reference Clusters

LISTA DE FIGURAS

Figura 1. Estrutura do Ebola Vírus.	2
<i>Figura 2 - Mapa dos surtos de Ebola.</i>	3
Figura 3 - Exemplos de LCA.	13
Figura 4 - Resultado do MedLineRanker.	20
Figura 5 – Resultado de busca pelo PESCADOR.	21
Figura 6 - Entrada do Seed Server.	23
Figura 7 - Saída do Seed Server.	24
Figura 8 - Fluxograma da montagem da via.	26
Figura 9 – Via de Infecção pelo vírus Ebola.	32
Figura 10 – Proteínas responsáveis pela adsorção viral.	33
Figura 11 – Proteínas envolvidas na penetração e desnudamento viral.	36
Figura 12 – Proteínas responsáveis pela modulação imune do hospedeiro.	39
Figura 13 - Via de infecção do Ebola, representada em três intervalos ao longo da evolução.	45
Figura 14 - Representação da expressão dos genes da série GSE80832.	47
Figura 15 - Representação da expressão dos genes da série GSE86539.	48
Figura 16 - Endocitose para a série GSE80832 no intervalo de 8 horas.	50
Figura 17 – Endocitose para a série GSE80832 no intervalo de 30 horas.	51
Figura 18 – Endocitose para a série GSE86539 no intervalo de 8 horas.	52
Figura 19 – Endocitose para a série GSE86539 no intervalo de 24 horas.	53
Figura 20 – Endocitose para a série GSE86539 no intervalo de 48 horas.	54
Figura 21 – Endocitose para a série GSE86539 no intervalo de 72 horas.	55
Figura 22 - Resultado do ELDOgraph.	56
Figura 23 - Arvore Filogenética da ICAM2.	58
Figura 24 - Arvore Filogenética da ICAM3.	59
Figura 25 - Número de genes na via por clado de origem.	61

LISTA DE TABELAS

<i>Tabela 1 - Hospedeiros do vírus Ebola</i>	5
<i>Tabela 2 - Linhagem Evolutiva Humana</i>	7
<i>Tabela 3 - Resultado dataset 1</i>	46
<i>Tabela 4 - Resultado dataset 2</i>	46

Sumário

1. INTRODUÇÃO.....	1
1.1. O vírus Ebola	1
1.1.1. Descrição e epidemia	1
1.1.2. Hospedeiros	4
1.1.3. Patogenia.....	6
1.2. Linhagem Evolutiva Humana.....	7
1.3. Introdução às ferramentas de bioinformática utilizadas.....	8
1.3.1. MedLine Ranker	8
1.3.2. PESCADOR	9
1.3.3. PathVisio.....	9
1.3.4. SeedServer	10
1.3.4.1. Bancos de dados usados pelo SeedServer.....	11
1.3.4.2. Programas usados pelo SeedServer	12
1.3.5. Algoritmo LCA.....	13
1.3.6. TaxOnTree	13
1.3.7. ARCoBALeno.....	14
1.3.8. Bioconductor	15
1.3.9. ELDOgraph	16
2. JUSTIFICATIVA	17
3. OBJETIVOS	18
3.1. OBJETIVO GERAL	18
3.2. OBJETIVOS ESPECÍFICOS	18
4. METODOLOGIA.....	19
4.1. Criação da via	19
4.2. Desenho da representação da via de interação.....	22
4.3. Inferência do clado de origem dos componentes da via.....	22
4.4. Colorização dos componentes da via	25
4.5. Análise de expressão diferencial	27
4.6. TaxOnTree	28
4.7. ELDOGraph.....	28
5. RESULTADOS.....	30

5.1. Entrada do vírus na célula hospedeira.....	30
5.2. Modulação da resposta imune e trans-infecção.....	38
5.3. Processos e famílias proteicas envolvidas na infecção	41
5.4. Origem da via de infecção pelo vírus ebola	43
5.4. Análise de expressão diferencial de séries infectadas por ebola	46
5.5. Origem das cepas do vírus Ebola.....	55
5.6. TaxOnTree.....	57
6. DISCUSSÃO.....	60
7. REFERENCIAS.....	66
8. MATERIAL SUPLEMENTAR	77

1. INTRODUÇÃO

Neste trabalho foi elaborada uma via de interação do vírus Ebola com a célula hospedeira durante a infecção, e foram utilizadas ferramentas de bioinformática para colorizar a via com os clados da linhagem evolutiva humana onde foi estimado que os genes primeiramente ocorressem, sinapomorfias¹, portanto. Também foram utilizados dados públicos de expressão gênica, foi feito um estudo da sua variação durante a infecção mapeando os genes induzidos e reprimidos em vias *Kegg Pathways*². Assim nesta sessão será introduzido o vírus Ebola, apresentando os clados da evolução humana e as ferramentas utilizadas, para um melhor entendimento dos resultados.

1.1. O vírus Ebola

Encontra-se descrito abaixo o vírus e suas particularidades e onde ocorreram as epidemias, os hospedeiros conhecidos e os sintomas.

1.1.1. Descrição e epidemia

O vírus Ebola é classificado como categoria A de arma biológica e um organismo nível 4 de biossegurança (KATZE *et al.*, 2008; METSCHNIKOFF, 1891; NORBURY, 2006). É o agente causador de uma patologia característica de febre hemorrágica (EHF), com taxa de mortalidade de até 90% (FRIEDEL; HAAS, 2011; XUE; MILLER-JENSEN, 2012). O vírus Ebola é um vírus filamentosos, da família *Filoviridae*, envelopado e de forma variada, tendo aparência em microscopia de longos tubos. A estrutura viral pode ser dividida em três compartimentos: o nucleocapsídeo, a matriz e o envelope (FALASCA *et al.*, 2015). O Ebola contém uma fita simples de RNA *negative sense* como material genético, com cerca de 18-19 kb de tamanho, que codifica sete

¹ Caráter derivado (apomorfia) surgindo ao longo da evolução, é compartilhado por um grupo e permite assim sua identificação.

² Plataforma para depósito de vias de interação molecular, reações e relações, que permite interação com outros usuários da plataforma.

genes: NP, VP40, VP35, VP24, VP30, L e GP (GOEIJENBIER *et al.*, 2014; HOENEN *et al.*, 2006; MOHAN *et al.*, 2015; WATANABE *et al.*, 2007); o *loci* GP também codifica duas outras proteínas, não estruturais e de função extrínsecas, que são secretadas: sGP (*secreted*) e ssGP (*small secreted*). A estrutura do vírus está representada na Figura 1.

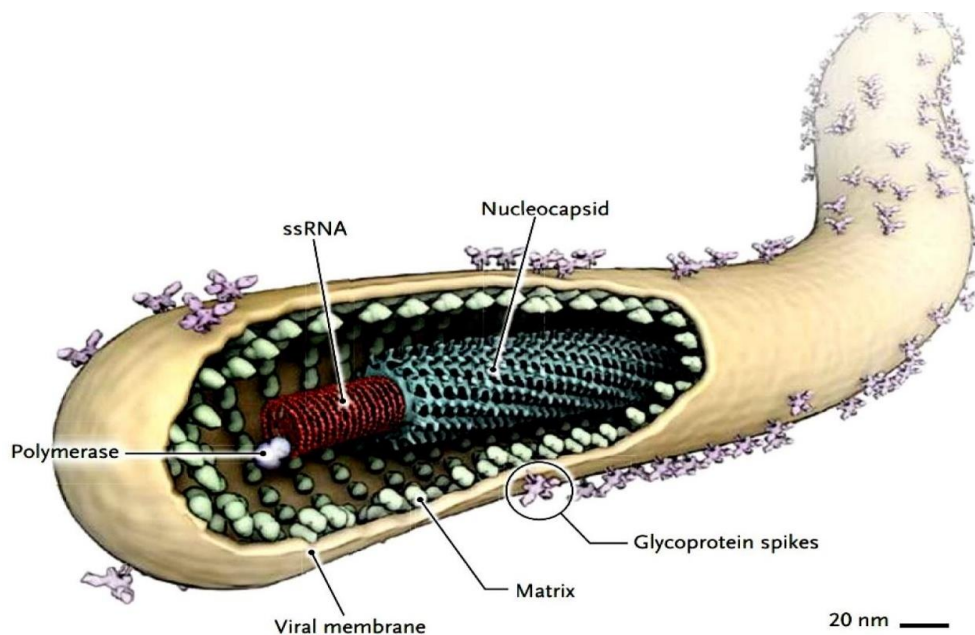


Figura 1. **Estrutura do Ebola Vírus.** Imagem representativa do vírus Ebola, em uma escala de 20 nanômetros, dividindo a estrutura em 6 partes.

Fonte: (HAMBESSE *et al.*, 2016)

Em 1976 na África, algumas pessoas começaram a apresentar sintomas de febre e hemorragia. O agente causador, até então não identificado, infectou 318 pessoas e causou a morte de 280. Seu surgimento é descrito em duas localidades diferentes do continente, primeiramente em *Sudan*, e em seguida em *Zaire* (atualmente República Democrática do Congo - DRC). Um desconhecido agente foi isolado de pacientes nestes dois locais e chamado de Ebola, devido a um rio em Zaire próximo ao local onde houve a infecção viral. As duas epidemias foram causadas por duas cepas diferentes do vírus: *Sudan Ebola vírus (SEBOV)* e *Zaire Ebola vírus (ZEBOV)*, que foram identificados apenas alguns anos depois. A terceira cepa a ser reconhecida foi *Côte d'Ivoire Ebola vírus (CIEBOV)*, descoberta em 1994, isolada de um etnologista,

que foi infectado fazendo uma necropsia de um chimpanzé, na reserva *Tai Forest* em *Côte d'Ivoire*. A quarta cepa do vírus descoberta foi o *Bundibugyo Ebola vírus (BEBOV)*, que foi encontrado na África equatorial também em 1994. Estas são as cepas do vírus Ebola que são patogênicas para humanos, todas elas descobertas no continente africano. Há, porém, uma cepa não patogênica para humanos que foi isolada de *Macaca fascicularis*, em Virginia nos Estados Unidos em 1989, animal este usado para estudos científicos. Ao se fazer o rastreamento da origem do animal, foi descoberto que ele teve a origem nas Filipinas, conseqüentemente trazendo o vírus de lá, este foi chamado de *Reston Ebola vírus (REBOV)*. As cinco cepas do vírus são distintas em alguns aspectos, as principais diferenças são a sequência, o número e a localização dos *overlaps* dos genes (GRIFONI *et al.*, 2016; LI, XINGGUANG *et al.*, 2016; LIU *et al.*, 2015). A imagem abaixo mostra as regiões da África onde ocorreram surtos do vírus Ebola que são patogênicos ao homem (Figura 2).

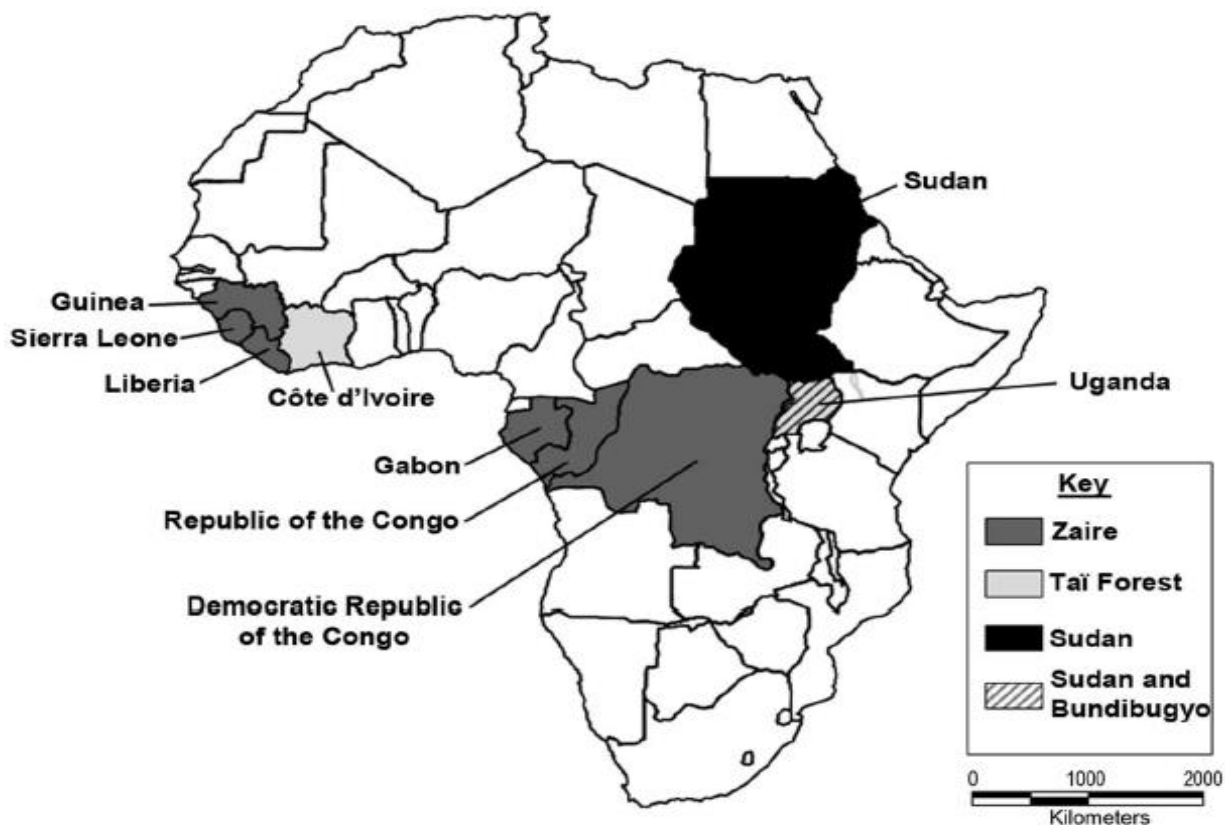


Figura 2 - Mapa dos surtos de Ebola. Representação atemporal dos locais na África onde houveram os principais surtos do vírus ebola, das cepas patogênicas ao homem.

Fonte: (BAUSCH; SCHWARZ, 2014)

Após o surgimento e contenção, em 1976, das cepas *Zaire* e *Sudan*, as duas cepas mais patogênicas para o homem, poucos casos foram relatados nos anos seguintes (AZARIAN *et al.*, 2015; GRIFONI *et al.*, 2016; ZHENG *et al.*, 2015). Entretanto, no ano de 2013, novos casos do vírus *Zaire* Ebola começaram a ser relatados em Guiné (África) e rapidamente espalhados à países ao longo do continente, sendo os países vizinhos, como Serra Leoa e Libéria, os mais afetados. Em 8 de agosto de 2014 a Organização Mundial de Saúde (WHO) declarou que a epidemia pelo vírus Ebola é um problema de saúde pública (GRIFONI *et al.*, 2016). Nesta última epidemia por Ebola, que teve início em 2013, foram registrados 28.640 casos confirmados de pacientes infectados, e um total de 11.316 óbitos, até 2016 (www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts).

As regiões mais afetadas pelos surtos virais: Guiné, Serra Leoa e Libéria, são muito carentes, ocupando as posições 178, 177, 174, entre os 187 países mais pobres do mundo, segundo o ranqueamento do *United Nations Development Programme Human Development Index*.

O vírus tem uma grande capacidade de infecção, podendo ser transmitido por contato com secreções e fluidos como urina, fezes, saliva, lágrimas, sangue e suor. O Ebola pode ser também transmitido por contato casual como aperto de mãos, contato com equipamento infectado, ou até mesmo em uma maçaneta. A infecção por vias aéreas também é uma possibilidade, porém sua capacidade de infecção por esta via não foi precisamente mensurada (LAWRENCE *et al.*, 2017). Estes seriam alguns dos motivos que possibilitaram uma disseminação rápida e de longo alcance do vírus (BAUSCH; SCHWARZ, 2014).

1.1.2 Hospedeiros

Baseado neste intervalo longo entre estas duas epidemias, 1976 e 2013, fez-se necessário realizar estudos epidemiológicos, para entender como o vírus perdurou por todos esses anos e para prevenir novas epidemias. Uma das hipóteses que explicaria a manutenção do vírus seria a existência de um organismo que sirva como seu

reservatório. Na busca por outros possíveis hospedeiros do vírus, primatas não humanos e cabritos se mostraram susceptíveis a infecção pelo vírus, mas apresentaram sintomas. Especulou-se então que o reservatório primário do vírus seria um organismo que teria pouco contato com humanos e que não apresentaria sintomas o que justificaria o intervalo das epidemias (OLIVAL; HAYMAN, 2014).

No trabalho feito por Leroy *et al.*, (2005) amostras de material genéticos de Ebola foram detectadas em 3 espécies de morcegos de fruta (*Epomops franqueti*, *Hypsignathus monstrosus*, e *Myonycteris torquata*). Estas espécies mostraram sobreviver a infecção pelo vírus, tornando-as possíveis hospedeiros primários para o vírus.

O aparecimento de casos de Ebola em humanos tem sido associado a animais selvagens, como primatas não humanos, cabritos e morcegos, que são consumidos como fonte de proteína na África. Alguns fatores têm aumentado o contato entre morcegos e outros mamíferos, como mudanças climáticas, desflorestamento, caça entre outros (BAUSCH; SCHWARZ, 2014). Na tabela abaixo estão listados alguns hospedeiros comprovados para o vírus.

Tabela 1 - Hospedeiros do vírus Ebola

Hospedeiro	Referência
Porcos	Weingartl <i>et al.</i> , 2012
Furão	Cross <i>et al.</i> , 2016
Cabritos	Takada 2012
Cachorro	Han <i>et al.</i> , 2015
Gato	Han <i>et al.</i> , 2015
Humano	Takada, 2012
Macaco	Takada, 2012
Morcego	Marí Saéz <i>et al.</i> , 2014
Orangotango	Nidom <i>et al.</i> , 2012
Mandrill	Leroy <i>et al.</i> , 2004
Drill	Leroy <i>et al.</i> , 2004
Babuíno	Leroy <i>et al.</i> , 2004
Gorila	Leroy <i>et al.</i> , 2004
Chipanzé	Leroy <i>et al.</i> , 2004

Analisando os hábitos dos animais onde os surtos surgiram em 2013, e a

distância do local do primeiro surto em 1976, especula-se que os morcegos foram os responsáveis pelo traslado do vírus Ebola *Zaire* da região central, para o oeste africano (LEENDERTZ, 2016).

1.1.3 Patogenia

A patogenicidade do vírus Ebola no corpo humano deriva-se de diversos mecanismos envolvidos na sua infecção. Estes incluem a destruição de células e tecidos, e efeitos indiretos em células próximas às infectadas que ampliam a ação destrutiva. O sistema imune inato e o adaptativo estão envolvidos na patogenicidade do vírus. Em particular, a resposta imune inata está relacionada à desregulação da resposta por interferón e a alteração dos sinais por citocinas e quimiocinas. Já a resposta adaptativa está ligada à desregulação da resposta humoral e da resposta imune específica (FALASCA *et al.*, 2015).

Os sintomas primários da infecção por Ebola são: febre alta (aproximadamente 40°C), mal-estar, fadiga e dores no corpo. A febre costuma persistir do dia três ao cinco da infecção, os sintomas gastrointestinais começam com dor, náusea, vômito e diarreia. Os pacientes apresentam após 2 ou 3 dias muito vômito e diarreia, período este de maior risco de disseminação do vírus (KORTEPETER; BAUSCH; BRAY, 2011; TATTEVIN; DURANTE-MANGONI; MASSAQUOI, 2014).

Durante os episódios recorrentes de vômito, e a inabilidade de tolerância a ingestão oral de alimentos, grandes volumes de diarreia diária persistem até o dia 7 com gradual redução. Associados a estes sintomas, aparecem uma forte dor abdominal, dor de cabeça, hiperemia conjuntival, dor no peito, dor nas junções musculares, mialgia e soluços. Os sintomas respiratórios são raros, normalmente apenas tosse. Os sintomas neurológicos mais comuns incluem delírio, hiper e hipo atividade, confusão, cognição lenta ou agitação e algumas vezes convulsões (FELDMAN; GEISBERT, 2012).

A rápida liberação de agentes vasoativos - incluindo citocinas, quimiocinas, histaminas e peroxidases - de monócitos e células endoteliais pode causar em aproximadamente 60% dos pacientes o desenvolvimento de choque, com redução da

consciência ou coma. Há perca de temperatura nas extremidades e vasoconstrição periférica. Hemorragia gastrointestinal ocorre em cerca de 5% dos pacientes antes da morte. A maioria das mortes ocorre entre os dias 7 e 12 da doença (STROM *et al.*, 2014; TAKADA, A; KAWAOKA, 2001).

A população mais vulnerável às patologias causadas pela infecção viral, são crianças com até 5 anos de idade, idosos e mulheres grávidas.

1.2. Linhagem Evolutiva Humana

Para a inferência evolutiva das proteínas encontradas na via de infecção viral, foi utilizada a linha evolutiva humana, composta por 31 níveis, ordenados do mais remoto ao mais recente na evolução humana. São eles:

Tabela 2 - Linhagem Evolutiva Humana

Nível	Nome	Descrição
1	Organismos Celulares	Todos os organismos constituídos por células
2	Eucariotos	Organismos com membrana nuclear
3	Opisthokonta	Fungos e Animais
4	Metazoa	Todos os animais
5	Eumetazoa	Animais com simetria radial e bilateral
6	Bilateria	Animais com simetria bilateral
7	Deuterostomia	Equinodermos e cordados
8	Chordata	Animais com notocorda. Cefalocordados, tunicados e craniados
9	Craniata	Animais que contém crânio
10	Vertebrata	Animais com esqueleto
11	Gnathostomata	Vertebrados com mandíbula
12	Teleostomi	Foram reclassificados em Eutelostomi
13	Euteleostomi	Vertebrados com esqueleto ósseo verdadeiro
14	Sarcopterygii	Vertebrados com membros lobados
15	Dipnotetrapodomorpha	Peixes pulmonados e tetrápodes
16	Tetrapoda	Vertebrados de 4 membros
17	Amniota	Vertebrados que apresentam âmnio
18	Mammalia	Todos os mamíferos
19	Theria	Metatérios e eutérios
20	Eutheria	Placentários verdadeiros
21	Boreoeutheria	Placentários agrupados por dados moleculares
22	Euarchontoglires	Lêmures, coelhos, roedores, musaranhos e primatas
23	Primates	Todos os primatas

24	Haplorrhini	Tarsos e símios
25	Simiiformes	Todos os símios
26	Catarrhini	Macacos e hominoídeos
27	Hominoidea	Todos os hominoídeos
28	Hominidae	Orangotangos e hominíneos
29	Homininae	Todos os hominíneos
30	Homo	Humanos extintos e atual
31	Homo Sapiens	Nível de evolução humana atual

1.3. Introdução às ferramentas de bioinformática utilizadas

Para um melhor entendimento dos resultados, será apresentada informações relevantes sobre o escopo e funcionamento das ferramentas de bioinformática utilizadas.

1.3.1. MedLine Ranker

O MedLine Ranker é uma ferramenta de mineração e classificação de texto, sua funcionalidade é definida segundo os critérios selecionados de entrada. O acesso à base de dados de literatura do software é feito através do PubMed do NCBI (<http://www.ncbi.nlm.nih.gov/pubmed>).

A classificação é feita por meio da comparação entre dois conjuntos: um grupo que representa o assunto a ser buscado, denominado *training set*. E outro chamado *test set*, que traz o assunto de forma mais geral, sem um foco, que será então classificado em função do assunto do grupo de treino. Os artigos são representados pelos seus identificadores do PubMed (PMIDs) nos dois grupos.

A primeira etapa da seleção consiste na extração de substantivos de cada resumo e título dos trabalhos, e na atribuição de valor a cada um em relação aos arquivos de busca. Na segunda etapa, o valor de cada resumo é definido segundo o resultado da soma dos valores das palavras presentes nos resumos e títulos calculados na etapa anterior. Para finalizar, a medida de *p-value* é definida através da proporção de resumos com pontuação elevada dentro de 10000 resumos recentes.

Esta ferramenta grátis está disponível através de web interface em http://cbdm-01.zdv.uni-mainz.de/~jfontain/cms/?page_id=4. É com ela que foi selecionado de forma eficaz artigos relevantes similares ao grupo usado como modelo.

1.3.2. PESCADOR

O PESCADOR (*Platform for Exploration of Significant Concepts AssociateD to co-Occurrence Relationships*) é uma ferramenta de mineração de texto, que inclui busca de redes de biointerações entre unidades biológicas (genes e produtos de genes) de trabalhos do PubMed. O software identifica co-ocorrências de entidades biológicas e suas interações, presentes em cada resumo. As co-ocorrências são classificadas nos 4 tipos abaixo:

1. Dois termos aparecem na mesma sentença e com biointeração entre eles;
2. Dois termos aparecem na mesma sentença, a biointeração também mas não está entre eles;
3. Dois termos aparecem na mesma sentença, mas sem biointeração;
4. Dois termos aparecem no mesmo resumo.

Há a possibilidade do uso de uma lista de conceitos que são utilizados como termos de procura. A função peculiar exercida pelo PESCADOR oferece a possibilidade de se complementar o resultado encontrado pelo MedLine Ranker.

O PESCADOR é uma ferramenta grátis que está disponível em interface web em <https://pescador.uni.lu/>.

1.3.3. PathVisio

O PathVisio é um software de livre acesso, que permite a criação, visualização e análise de vias biológicas (KUTMON *et al.*, 2015). O programa apresenta vários objetos para a representação das vias. Para os moldes necessários a esse trabalho, foram usados o formato equivalente às vias *Wikipathways* e ao ARCoBALeno (descrito

adiante neste trabalho).

Os principais objetos para a representação disponíveis são: *DataNode* usado para entidades biológicas (genes, produtos gênicos e moléculas), *Lines* representam as interações diretas entre as entidades biológicas (linhas, setas, *Tbars*), e *Shapes* usados em anotações gráficas (KUTMON *et al.*, 2015).

O PathVisio ainda traz funções para análise estatística e funções estendidas através da instalação de novos plug-ins (KUTMON *et al.*, 2015). Ele está disponível para download em <http://www.pathvisio.org/>.

1.3.4. SeedServer

O SeedServer é um programa que integra diferentes bancos de dados e programas com a finalidade de agrupamento de proteínas homólogas.

A primeira etapa realizada pelo programa é entrada da *query*³ através dos identificadores UniProtKB (*seeds*), logo é requerida uma curadoria para se chegar ao acesso correto quando se parte de um símbolo ou descrição do gene, na literatura. Em seguida é feita a associação com seus respectivos identificadores taxonômicos e assim obtendo-se a taxonomia completa das proteínas (reino, filo, classe e etc).

A próxima etapa realizada pelo programa é a execução do *Seed Linkage* (descrito à frente) que busca parálogos⁴ e ortólogos⁵ das *seeds* na base UniProt utilizando apenas proteínas completas. Em seguida ele consulta grupos de ortólogos das bases KO/UEKO (descritas a seguir). Caso a *seed* não seja encontrada no banco de dados KO (e conseqüentemente na sua extensão UEKO), o procedimento de enriquecimento com grupos UniRef50 utilizado para construção do UEKO é realizado com a *Seed*.

Para finalizar o processo, é feita a validação do grupo produzido através do PSI-BLAST. O programa faz ao final a adição de informações que descrevem as proteínas. O SeedServer foi desenvolvido por Rafael Lucas Muniz Guedes durante seu doutorado em Bioinformática na Universidade Federal de Minas Gerais (não publicado).

³ Proteína usada como modelo de estudo.

⁴ Termo referente a duas espécies que partilham um ancestral comum.

⁵ Entidades homólogas separados por um evento de especiação.

1.3.4.1. Bancos de dados usados pelo SeedServer

O UniProt (*Universal Protein Resource*) é um banco de dados de proteínas. O UniprotKB (*UniProt Knowledgebase*) é um componente do UniProt onde está armazenado as informações das proteínas e suas sequências. O SeedServer usa UniProtKB para a obtenção das sequências e as informações das *queries* de sua busca. O UniRef (*UniProt Reference Clusters*) é onde está depositado os agrupamentos de proteínas do UniProt com níveis de identidade (50%, 90% e 100%). O UniRef50 é um agrupamento utilizado pelo SeedServer durante o procedimento UE-Seed (*UniRef50 Enriched-Seed*) e na construção da base UEKO. O acesso ao Uniprot é gratuito, pelo seguinte endereço: <http://www.uniprot.org/>.

O KO (*Kegg Orthology*) é baseado em um componente da base de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*) e nele é depositado os grupos de ortólogos.

O UEKO (*UniRef50 Enriched Kegg Orthology*) é uma base de dados que abriga os grupos de ortólogos do KO enriquecidos com proteínas do cluster UniRef50. O UEKO foi criado através do alinhamento das sequências originais dos grupos de ortólogos do KO e as sequências dos respectivos agrupamentos UniRef50 que contêm membros do KO, tendo como exigência o mínimo de 50% de recobrimento (tamanho do alinhamento dividido pelo tamanho da sequência original). O KO é de livre acesso no endereço: <http://www.genome.jp/kegg/ko>. O UEKO foi desenvolvido por Gabriel da Rocha Fernandes durante seu doutorado em Bioinformática (não publicado), utilizando um algoritmo desenvolvido para enriquecer a base COG (Fernandes *et al.*, 2008).

O NCBI-Taxonomy é um banco de dados de acesso livre, que armazena a classificação e nomenclatura para todos organismos que tenham sequencias depositadas em repositórios públicos. Os dados usados pelo SeedServer referentes a taxonomia foram obtidos por esse banco. O acesso é livre, e é feito pelo seguinte link: <https://www.ncbi.nlm.nih.gov/taxonomy>.

1.3.4.2. Programas usados pelo SeedServer

O Seed Linkage é um programa que cria agrupamentos de proteínas homólogas (KUTMON *et al.*, 2015), utilizado pelo SeedServer. O Seed Linkage utiliza relações de melhor alinhamento recíproco (BBH, *Bidirectional Best Hit*) para a determinação destes homólogos. O funcionamento deste *software* se dá através do alinhamento de uma sequência de interesse com as sequências de outros organismos, e usando os valores de BBH para a determinação de *out*-parálogos e o melhor valor BBH para a determinação de *in*-parálogos. O Seed Linkage apresenta como vantagem a criação de grupos de homólogos a partir de uma única proteína de interesse e não requer que a busca seja feita com genomas completamente sequenciados.

O PSI-BLAST (*Position-Specific Iterated Blast*) é uma modalidade de BLAST que faz alinhamentos sucessivos, e após cada alinhamento é calculada uma matriz PSSM⁶, a matriz de pontuação é, portanto, refeita a cada busca e utilizada no alinhamento seguinte. O SeedServer primeiramente realiza o alinhamento entre a *seed* e ela mesma e obtém o *auto-score*, este valor serve então de referência para limitar as iterações do PSI-BLAST. No próximo passo, a *seed* é usada contra as proteínas recrutadas pelo SeedServer de forma sucessiva, até que o valor de *score* seja superior a 70% do *auto-score*. O PSI-BLAST ao alcançar o valor almejado, passará então a considerar as sequências presentes como válidas. Este procedimento foi projetado para eliminar possíveis contaminações com famílias relacionadas distintas da família da Seed. O módulo do PSI-BLAST usado no SeedServer foi desenvolvido por Henrique de Assis Lopes Ribeiro durante seu doutorado em Bioinformática pela Universidade Federal de Minas Gerais (não publicado).

⁶ Position-Specific Scoring Matrix.

1.3.5. Algoritmo LCA

No estudo da filogenia, o LCA (*Lowest Common Ancestor*) é o último ancestral comum entre dois pontos da árvore filogenética. A figura abaixo exemplifica um LCA:

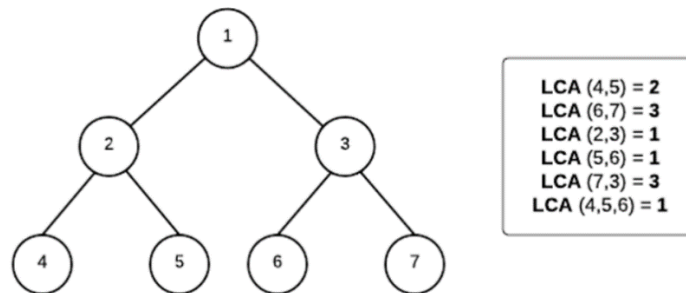


Figura 3 - Exemplos de LCA. Modelo de análise para o *Lowest common Ancestor*.

Fonte: Tese de doutorado de Henrique Velloso Ferreira Melo.

Ao analisar os pontos 4 e 5 da figura pode-se concluir que o LCA entre eles é o 2. O último ancestral comum que tem os dois nós como descendentes.

O algoritmo inicia a busca com os identificadores taxonômicos fornecidos na entrada, a árvore filogenética completa é carregada na memória e varrida até encontrar o nó parental dos dados de entrada, e calcula-se assim o LCA.

Este algoritmo foi implementado em um *webservice* por Henrique Velloso Ferreira Melo durante seu doutorado em Bioinformática na Universidade Federal de Minas Gerais (não publicado). É por meio desse serviço que determinamos LCA com os resultados de grupos de ortólogos fabricados com SeedServer.

1.3.6. TaxOnTree

A TaxOnTree é um *software* que permite a montagem de árvores filogenéticas para a visualização dos clados taxonômicos, de acordo com as sequências dos organismos pesquisados. Os diferentes clados são diferenciados através de cores atribuídas aos ramos da árvore.

A TaxOnTree incorpora dados taxonômicos em uma representação de árvore

filogenética. A partir do resultado gerado, o usuário pode ter acesso às informações taxonômicas das amostras presentes nas árvores, tendo como base de informações do banco de dados do NCBI Taxonomy, assim como a relação evolutiva entre elas.

Este software foi desenvolvido por Tetsu Sakamoto durante seu doutorado em Bioinformática na Universidade Federal de Minas Gerais (não publicado) e está disponível para uso em interface web em: <http://bioinfo.icb.ufmg.br/taxontree/>. É com ele e com análise manual que determinamos com a melhor aproximação possível o LCA dos genes aqui pesquisados.

1.3.7. ARCoBALeno

O ARCoBALeno (*Application for Rapidly COloring Biological pathways by Ancestrality or gene ontoLogy-Enriched iNformatiOn*) é uma ferramenta que atribui cor à arquivos de formato “.gpml”. Formato este de saída do programa PathVisio e apropriado para o uso no repositório Wikipathways. As cores atribuídas no programa são referentes a clados de origem dos genes presentes no arquivo gpml. Esta função permite a identificação de uma cronologia evolutiva dos componentes da via.

O software ainda permite a atribuição de cores de acordo com a origem da função molecular mais recente do gene (LMF, *Last Molecular Function*), do processo biológico mais recente em que o gene participa (LBP, *Last Biological Process*) e segundo a quantidade de processos biológicos em que os genes estão envolvidos, para isso essa plataforma faz o uso de informações obtidas segundo o banco de dados do GO (*Gene Ontology*).

O ARCoBALeno é uma ferramenta de livre acesso, e foi desenvolvida por Carlos Alberto Xavier Gonçalves durante seu mestrado em Bioinformática na Universidade Federal de Minas Gerais (não publicado) e está disponível para uso em interface web em: <http://biodados.icb.ufmg.br/arcobaleno/run.html>.

1.3.8. Bioconductor

O Bioconductor é um projeto *open source*, desenvolvido para programação em R que disponibiliza pacotes para análises principalmente de *microarrays*, *pathways* e *organisms annotations* (GENTLEMAN *et al.*, 2004; HUBER *et al.*, 2015). Dentre os pacotes disponibilizados pelo bioconductor de interesse para a bioinformática, pode-se listar:

Limma - pacote com métodos estatísticos para inferir expressão diferencial (RITCHIE *et al.*, 2015);

KeggRest - pacote em R que faz a conexão com o API (*application programming interface*) do KEGG, usado para integração ao banco de dados do Kegg (Dan Tenenbaum, 2018);

BioMart - pacote para busca e conversão dos identificadores das sonda de microarray (DURINCK *et al.*, 2005, 2009);

Pbapply – pacote para estimar média e variância de análises de expressão diferencial (LOVE; HUBER; ANDERS, 2014);

Genomicper- pacote com funções para fazer testes hiper geométricos (LOVE; HUBER; ANDERS, 2014).

HsAgilentDesign026652.db – pacote específico da tecnologia Agilent, das séries pesquisadas (CARLSON M, 2016).

Os pacotes específicos para cada tecnologia utilizada nos *dataset* estudados, permitem o tratamento dos dados, como normalização, dados *outlier*, etc.

Os pacotes disponibilizados pelo Bioconductor foram utilizados para a análise de expressão diferencial das séries de células infectadas pelo vírus Ebola, usadas neste trabalho.

1.3.9. ELDOgraph

O ELDOgraph é um programa que realiza análises comparativas em uma árvore filogenética, com base nas distâncias evolutivas entre as espécies, permitindo calcular quais espécies possuem mais genes similares com a em análise.

O nome ELDO deriva do termo LDO (*Least Diverged Orthology*) que significa o ortólogo que menos divergiu entre duas espécies. O ELDO (*External Least Diverged Ortholog*) de uma proteína é o ortólogo mais próximo que pertence a um táxon distinto (externo) ao da proteína analisada. A diferença entre os termos é que o ELDO é uma propriedade individual de cada proteína.

O programa ELDOgraph permite ao usuário escolher o nível taxonômico (gênero, família, ordem, etc.) de cada proteína analisada, para que seja feita a busca pelo ortólogo menos divergente pertencente a um táxon distinto. Na interface do *software* é permitido a visualização entre as relações através de grafos.

Este software foi desenvolvido por Tetsu Sakamoto durante seu doutorado em Bioinformática na Universidade Federal de Minas Gerais.

2. JUSTIFICATIVA

Os dados existentes na literatura sobre a interação do vírus Ebola com as proteínas das células hospedeiras durante a infecção nunca foram organizados de uma forma sistêmica no formato de uma via. Uma via coloca a informação organizada em um cenário, onde cada gene de interesse de um investigador está integrado em um sistema. A coevolução do vírus com o hospedeiro tem sido estudada em alguns aspectos, mas nunca se perguntou em que estágio da evolução o vírus encontra proteínas com as quais hoje interage, pois seria plausível que alguns receptores tenham vindo a existir somente na forma de sinapomorfias de clados mais recentes na evolução, bem como algumas proteínas de combate à presença do vírus. Justifica-se também desenvolver-se uma abordagem que permita a visualização de mudanças de expressão gênica pleiotrópicas durante a infecção, investigando-se um grande conjunto de vias como as do Kegg Pathways. Estudos sistêmicos com o vírus Ebola não só cobrem os aspectos evolutivos, mas também destacam possíveis alvos para interferência terapêutica na enfermidade. A relevância do Ebola como patógeno é indiscutível, então estudos abrangentes como o proposto para o vírus, são justificados.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Criar uma via de infecção do vírus Ebola e verificar se proteínas mais recentes na evolução são requeridas para a infecção ou se originam para defesa, sugerindo co-evolução.

3.2. OBJETIVOS ESPECÍFICOS

- Obter os genes (e as biointerações entre eles) envolvidos em cada fase do ciclo de multiplicação viral, a partir da literatura usando métodos de mineração de texto.
- Elaborar uma figura ilustrativa da via de infecção viral.
- Estudar as famílias de proteínas presentes na via e suas funcionalidades.
- Inferir a origem de cada gene do hospedeiro envolvido na infecção viral, usando metodologias baseadas no agrupamento de homólogos e na determinação do ancestral comum.
- Analisar os genes diferencialmente expressos em células infectadas com o vírus Ebola e os processos envolvidos e planejar uma plataforma de análise de infecções virais e seus efeitos pleiotrópicos⁷ nas vias do *Kegg Pathways*.

⁷ Aquele que tem múltiplos efeitos.

4. METODOLOGIA

4.1. Criação da via

Para a criação da via de representação da infecção pelo vírus Ebola, primeiramente foram realizadas várias buscas por trabalhos científicos que caracterizavam o envolvimento de proteínas do hospedeiro durante o processo infeccioso do vírus. Várias rotinas foram criadas a fim de abranger determinado ponto sobre a via, estas só foram dadas como satisfeitas quando não houvessem novos resultados sobre o determinado assunto pesquisado.

As rotinas consistiam basicamente de uma busca inicial, que foi feita utilizando a plataforma PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>), usando a palavra Ebola como *query*. Os artigos recuperados através desta busca foram selecionados e utilizados na plataforma de mineração de texto a seguir como *test set*. Mais de 7 mil artigos foram coletados. A segunda busca foi feita também no PubMed, mas desta vez com termos mais específicos como “*Ebola infection*”, “*Ebola proteins*”, “*Ebola hosts*” e entre outros termos que referenciassem o vírus ou as etapas da infecção viral. Do resultado encontrado, foram curados manualmente aproximadamente 10 artigos sobre determinado aspecto a ser coberto, que formaram o *training set*.

O MedLine Ranker (<http://cbdm-01.zdv.uni-mainz.de/~jfontain/cms/>), foi o primeiro software a ser usado, nele o *test set* e o *training set* foram dados de entrada e cruzados. O resultado desta análise é uma lista de classificação de artigos referentes ao *test set* com um valor atribuído ao nível de correspondência com o *training set*. Deste resultado, os 1000 melhores trabalhos foram separados ($p\text{-value} < 10^{-2}$). Um exemplo de resultado encontrado com o MedLine Ranker é mostrado na figura 4.

Parameters


The training set: 6 / 6 abstracts
The background set: the profile of the whole Medline database.
The test set: 4800 abstracts
Other: Scoring scheme: Bayes, Min word weight: 0.5, Target database: medline
Initialisation: less than 1 second.

Top - Results - Discriminative words - Download

Results

Processing 3948 abstracts: 0% - 9% - 19% - 29% - 39% - 49% - 59% - 69% - 79% - 89% - 99% - done.

Ranked in less than 1 second.

colors: yellow for training set pmids, green for background set pmids, and brown for discriminative words Color code: high  low weight.

Click on the pmid to read the abstract with highlighted discriminative words

Rank	PMID	Abstract Title	P-value
1	28768865	Ubiquitin Ligase WWP1 Interacts with Ebola Virus VP40 To Regulate Egress .	2.40e-05
2	27489272	ITCH E3 Ubiquitin Ligase Interacts with Ebola Virus VP40 To Regulate Budding.	3.15e-05
3	28076420	Chaperone-Mediated Autophagy Protein BAG3 Negatively Regulates Ebola and Marburg VP40 -Mediated Egress .	4.50e-05
4	24741084	Small-molecule probes targeting the viral PPxY -host Nedd4 interface block egress of a broad range of RNA viruses.	4.89e-05
5	25786915	ALIX Rescues Budding of a Double PTAP/PPEY L-Domain Deletion Mutant of Ebola VP40 : A Role for ALIX in Ebola Virus Egress .	4.92e-05
6	20032189	Conserved motifs within Ebola and Marburg virus VP40 proteins are important for stability, localization, and subsequent budding of virus -like particles .	5.53e-05
7	15811650	Virus -like particles exhibit potential as a pan-filovirus vaccine for both Ebola and Marburg viral infections .	5.76e-05
8	16026254	Filovirus -like particles as vaccines and discovery tools .	5.79e-05
9	28679761	The Host E3-Ubiquitin Ligase TRIM6 Ubiquitinates the Ebola Virus VP35 Protein and Promotes Virus Replication .	5.91e-05
10	20702783	Identification of essential filovirion-associated host factors by serial proteomic analysis and RNAi screen.	6.00e-05
11	21994610	Evasion of the interferon-mediated antiviral response by filoviruses .	6.00e-05
12	19114059	No exit: targeting the budding process to inhibit filovirus replication .	6.02e-05
13	28622346	Flavivirus and Filovirus EvoPrinters: New alignment tools for the comparative analysis of viral evolution .	6.08e-05
14	17940980	Ebola virus -like particle -based vaccine protects nonhuman primates against lethal Ebola virus challenge.	6.09e-05
15	26513362	Calcium Regulation of Hemorrhagic Fever Virus Budding: Mechanistic Implications for Host -Oriented Therapeutic Intervention.	6.12e-05
16	27122575	Characterization of Human and Murine T-Cell Immunoglobulin Mucin Domain 4 (TIM-4) IgV Domain Residues Critical for Ebola Virus Entry.	6.17e-05
17	26468532	Macaque Monoclonal Antibodies Targeting Novel Conserved Epitopes within Filovirus Glycoprotein.	6.25e-05
18	21987757	Characterization of filovirus protein -protein interactions in mammalian cells using bimolecular complementation.	6.36e-05
19	19000317	The YPLGVG sequence of the Nipah virus matrix protein is required for budding.	6.38e-05

Figura 4 - Resultado do MedLineRanker. Os resultados são ranqueados segundo o training set e recebem assim um P-value de correspondência.

Os mil trabalhos selecionados pelo MedLine Ranker, a cada rotina, foram então processados pelo software *PESCADOR* (<https://pescador.uni.lu/>) de forma a realçar eletronicamente os termos presentes nos resumos que denotem entidades biológicas, como nomes de genes, proteínas, e biointerações (“interact”, “activate”, “bind”, etc). O processamento destes resumos pelo *PESCADOR* foi importante para facilitar a extração manual das interações entre proteínas do vírus e do hospedeiro descritas na literatura. Um modelo de resultado do *PESCADOR* é apresentado abaixo na figura 5.



PESCADOR

Platform for Exploration of Significant Concepts AssociateD to co-Occurrences Relationships.

Results for ID 1507919632: [Summary](#) | [Terms](#) | [Concepts](#) | [Abstracts](#) | [Network](#) | [Validations](#) | [Help](#) | [Home](#) |

Selected PubMed ID: [15613320](#) [Highlight all NLPROT terms](#) [?]

Sentence 1: Studies of ebola virus glycoprotein-mediated entry and fusion by using pseudotyped human immunodeficiency virus type 1 virions: involvement of cytoskeletal proteins and enhancement by tumor necrosis factor alpha. The Ebola filoviruses are aggressive pathogens that cause severe and often lethal hemorrhagic fever syndromes in humans and nonhuman primates.

Sentence 2: To date, no effective therapies have been identified.

Sentence 3: To analyze the entry and fusion properties of ebola virus, we adapted a human immunodeficiency virus type 1 (hiv-1) virion-based fusion assay by substituting ebola virus glycoprotein (GP) for the hiv-1 envelope.

Sentence 4: Fusion was detected by cleavage of the fluorogenic substrate CCF2 by beta-lactamase-Vpr incorporated into virions and released as a result of virion fusion.

Sentence 5: Entry and fusion induced by the ebola virus GP occurred with much slower kinetics than with vesicular stomatitis virus G protein (VSV-G) and were blocked by depletion of membrane cholesterol and by inhibition of vesicular acidification with bafilomycin A1.

Sentence 6: These properties confirmed earlier studies and validated the assay for exploring other properties of ebola virus GP-mediated entry and fusion.

Sentence 7: Entry and fusion of ebola virus GP pseudotypes, but not VSV-G or hiv-1 Env pseudotypes, were impaired in the presence of the microtubule-disrupting agent nocodazole but were enhanced in the presence of the microtubule-stabilizing agent paclitaxel (Taxol).

Sentence 8: Agents that impaired microfilament function, including cytochalasin B, cytochalasin D, latrunculin A, and jasplakinolide, also inhibited ebola virus GP-mediated entry and fusion.

All loaded PMIDs

[Download](#) list

- | | |
|-----|--------------------------|
| 1: | 15613320 |
| 2: | 15140952 |
| 3: | 21994610 |
| 4: | 16026254 |
| 5: | 21062490 |
| 6: | 22935026 |
| 7: | 21866103 |
| 8: | 19694547 |
| 9: | 23650881 |
| 10: | 19114059 |
| 11: | 14990685 |
| 12: | 23441171 |
| 13: | 12941909 |
| 14: | 22072961 |

Figura 5 – Resultado de busca pelo PESCADOR. São realçados termos que tem relação ao assunto de interesse, como GP e Env.

Ao final, foi feita uma conferência manual de cada artigo selecionado e validado segundo concordância com outros trabalhos, posteriormente foi então, gerada uma

tabela de biointerações contendo a relação das proteínas do hospedeiro com as proteínas virais e/ou com processos da infecção viral (Tabela suplementar T1).

4.2. Desenho da representação da via de interação

Após reunir em uma tabela de biointerações informações sobre as proteínas do hospedeiro envolvidas no processo infeccioso do Ebola e sua relação com as proteínas virais, o próximo passo foi criar uma representação visual dos dados desta tabela na forma de uma via biológica. A via de interação foi confeccionada aos moldes do repositório *Wikipathways* (wikipathways.org), seguindo o padrão de outras vias já depositadas. Esta plataforma permite que usuários depositem vias de diversos cenários de interesse científico. A via confeccionada no presente trabalho terá uma significativa contribuição já que esta representa a primeira tentativa de conciliar as informações presentes na literatura científica sobre as etapas de infecção do vírus Ebola na forma de uma via.

A figura foi confeccionada usando o *software Pathvisio* (<https://www.pathvisio.org/>) que traz opções de simbologia para a representação da via. Na confecção da via feita, foi padronizado que: genes, produtos gênicos e substâncias são representados por retângulos. As setas significam relação positiva ou direção, TBar significa relação negativa e linha significa relação neutra. Outras estruturas presentes foram identificadas com texto, na via ou na legenda dela.

4.3. Inferência do clado de origem dos componentes da via

A estimativa da origem dos genes envolvidos na via de infecção do vírus do Ebola foi feita utilizando o programa *SeedServer* para agrupamento de homólogos e o algoritmo LCA (*Lowest Common Ancestor*), que determina o último ancestral comum dos organismos aos quais as proteínas que compõem um grupo de ortólogos pertencem.

A entrada para o programa *SeedServer* é o identificador UniProtKB de cada proteína presente na via. Estes foram buscados manualmente, através da etapa final

da mineração de texto e associados ao *Gene Symbol* de seu componente correspondente. O programa está disponível para acesso local ou via *Bioinformatics Open Web Services* (BOWS) (<http://maxixe.icb.ufmg.br:8080/BOWSWeb/>). No presente trabalho foi utilizada a versão local que recebe os parâmetros através de um arquivo de texto (**Erro! Fonte de referência não encontrada.6**).

```
<processID>TPCN2_out</processID>
<SLtable>allorg</SLtable>
<input>Q8NHX9</input>
<parameter_r>50</parameter_r>
<parameter_c>50</parameter_c>
<parameter_s>50</parameter_s>
<parameter_l>0.3</parameter_l>
<e_value>1e-10</e_value>
<self_score>0.7</self_score>
```

Figura 6 - Entrada do Seed Server. Script dado de entrada para o programa Seed Server, com os parâmetros nativos.

O resultado do *SeedServer* é uma lista de identificadores das proteínas tidas como homólogas àquela fornecida de entrada ao programa. A **Erro! Fonte de referência não encontrada.7** ilustra parte do arquivo de saída do *SeedServer* listando os homólogos do gene TPCN2, que codifica um canal poroso de cálcio (acesso Uniprot: Q8NHX9).

1	uniprot_ID	cluster_number	SL	KO	size	txi	clade	validated	PSI	description
2	Q8NHX9	1	2	1	752	9606	1	1	1	TPC2_HUMAN Two pore calcium channel protein 2 OS=Homo sapiens GN=TPCN2 PE=1 SV=2
3	K7CUS7	1	1	2	752	9598	1	1	1	K7CUS7_PANTR Two pore segment channel 2 OS=Pan troglodytes GN=TPCN2 PE=2 SV=1
4	G7PNZ5	1	1	2	752	9541	1	1	1	G7PNZ5_MACFA Voltage-dependent calcium channel protein TPC2 OS=Macaca fascicularis GN=EGM_05332 PE=4 SV=1
5	G7NBU4	1	1	2	752	9544	1	1	1	G7NBU4_MACMU Voltage-dependent calcium channel protein TPC2 OS=Macaca mulatta GN=EKG_05900 PE=4 SV=1
6	F7GUY6	1	1	2	1525	9544	1	1	1	F7GUY6_MACMU Uncharacterized protein OS=Macaca mulatta GN=TPCN2 PE=4 SV=1
7	H2NCK2	1	1	0	717	9601	1	1	1	H2NCK2_PONAB Uncharacterized protein OS=Pongo abelii GN=TPCN2 PE=4 SV=1
8	U3CZM9	1	1	2	752	9483	1	1	1	U3CZM9_CALJA Two pore calcium channel protein 2 OS=Callithrix jacchus GN=TPCN2 PE=2 SV=1
9	G1R4C4	1	1	2	752	61853	1	1	1	G1R4C4_NOMLE Uncharacterized protein OS=Nomascus leucogenys GN=TPCN2 PE=4 SV=1
10	E2QYG4	1	1	2	751	9615	1	1	1	E2QYG4_CANFA Uncharacterized protein OS=Canis familiaris GN=TPCN2 PE=4 SV=2
11	GSBSR8	1	1	1	753	10181	1	1	1	GSBSR8_HETGA Two pore calcium channel protein 2 OS=Heterocephalus glaber GN=GW7_00301 PE=4 SV=1
12	G1L487	1	1	2	759	9646	1	1	1	G1L487_AILME Uncharacterized protein OS=Ailuropoda melanoleuca GN=TPCN2 PE=4 SV=1
13	E1B1B9	1	1	1	751	9913	1	1	1	E1B1B9_BOVIN Uncharacterized protein OS=Bos taurus GN=TPCN2 PE=4 SV=1
14	I3NDL7	1	1	2	748	43179	1	1	1	I3NDL7_SPETR Uncharacterized protein OS=Spermophilus tridecemlineatus GN=TPCN2 PE=4 SV=1
15	M3XZJ7	1	1	2	746	9669	1	1	1	M3XZJ7_MUSPF Uncharacterized protein OS=Mustela putorius furo GN=TPCN2 PE=4 SV=1
16	B2RUK7	1	1	2	731	10090	1	1	1	B2RUK7_MOUSE Two pore segment channel 2 OS=Mus musculus GN=tpcn2 PE=2 SV=1
17	Q8BWC0	1	1	1	731	10090	1	1	1	TPC2_MOUSE Two pore calcium channel protein 2 OS=Mus musculus GN=tpcn2 PE=1 SV=1
18	L8HTI9	1	1	2	754	72004	1	1	1	L8HTI9_9CETA Two pore calcium channel protein 2 OS=Bos mutus GN=M91_19939 PE=4 SV=1
19	D3ZTJ6	1	1	1	711	10116	1	1	1	D3ZTJ6_RAT Protein Tpcn2 OS=Rattus norvegicus GN=tpcn2 PE=4 SV=1
20	L5LF18	1	1	0	668	225400	1	1	1	L5LF18_MYODS Two pore calcium channel protein 2 OS=Myotis davidii GN=MDA_GLEAN10017788 PE=4 SV=1
21	U3KZC1	1	1	2	766	59894	1	1	1	U3KZC1_FICAL Uncharacterized protein OS=Ficedula albicollis GN=TPCN2 PE=4 SV=1
22	F1NSV9	1	1	2	766	9031	1	1	1	F1NSV9_CHICK Uncharacterized protein OS=Gallus gallus GN=TPCN2 PE=4 SV=2
23	K7G0L2	1	1	2	684	13735	1	1	1	K7G0L2_PELSI Uncharacterized protein OS=Pelodiscus sinensis GN=TPCN2 PE=4 SV=1
24	ALL1H8	1	1	2	772	8355	1	1	1	ALL1H8_XENLA MGCL15225 protein OS=Xenopus laevis GN=MGCL15225 PE=2 SV=1
25	F6ZZ82	1	1	2	771	8364	1	1	1	F6ZZ82_XENTR Uncharacterized protein OS=Xenopus tropicalis GN=nadsynl PE=4 SV=1
26	U6DUX5	1	1	0	744	452646	1	1	1	U6DUX5_NEOVI Two pore calcium channel protein 2 OS=Neovison vison GN=TPC2 PE=2 SV=1
27	A0JMD4	1	1	1	774	7955	1	1	1	TPC2_DANRE Two pore calcium channel protein 2 OS=Danio rerio GN=tpcn2 PE=2 SV=1
28	F1QKC7	1	1	2	774	7955	1	1	1	F1QKC7_DANRE Two pore calcium channel protein 2 OS=Danio rerio GN=tpcn2 PE=4 SV=1
29	WSMCH1	1	1	2	772	7918	1	1	1	WSMCH1_LEPOC Uncharacterized protein OS=Lepisosteus oculatus PE=4 SV=1
30	WSKQX3	1	1	2	764	7994	1	1	1	WSKQX3_ASTMX Uncharacterized protein OS=Astyanax mexicanus PE=4 SV=1
31	I3J656	1	1	2	765	8128	1	1	1	I3J656_ORENI Uncharacterized protein OS=Oreochromis niloticus GN=LOC100695350 PE=4 SV=1
32	I3J657	1	1	2	1360	8128	1	1	1	I3J657_ORENI Uncharacterized protein OS=Oreochromis niloticus GN=LOC100695350 PE=4 SV=1
33	M4AHE5	1	1	2	776	8083	1	1	1	M4AHE5_XIPMA Uncharacterized protein OS=Xiphophorus maculatus PE=4 SV=1
34	B8PKJ9	1	1	1	798	7668	1	1	1	B8PKJ9_STRPU Two-pore calcium channel 2 OS=Strongylocentrotus purpuratus PE=2 SV=1
35	D1J6X8	1	1	2	796	7668	1	1	1	D1J6X8_STRPU Two-pore channel 2 OS=Strongylocentrotus purpuratus GN=tpcn2 PE=2 SV=1
36	W4Y2S6	1	1	2	706	7668	1	1	1	W4Y2S6_STRPU Uncharacterized protein OS=Strongylocentrotus purpuratus GN=Sp-tpcn2 PE=4 SV=1
37	V3Z0U0	1	1	0	854	225164	1	1	1	V3Z0U0_LOTGI Uncharacterized protein OS=Lottia gigantea GN=LOTGIDRAFT_229618 PE=4 SV=1

Figura 7 - Saída do Seed Server. Resultado da análise com lista de homólogos do organismo dado de entrada ao programa

O resultado recebido do *SeedServer* foi posteriormente usado para determinar o clado de origem da proteína estudada utilizando o algoritmo do LCA (*Lowest Common Ancestor*). Este algoritmo busca obter as linhagens taxonômicas de todos os organismos listados no grupo de homólogos e determinar o último ancestral que seja comum em todas as linhagens taxonômicas fornecidas pelo NCBI Taxonomy. O LCA de cada grupo foi encontrado através de linha de comando, com um *script* que tem como entrada o resultado do *SeedServer* e que requisita o serviço LCA presente em <http://biodados.icb.ufmg.br/services>, utilizando o protocolo SOAP.

Adicionalmente, utilizamos também uma curadoria manual para inferir o LCA. Cada identificador UniProt foi utilizado como entrada na ferramenta TaxOnTree (mais detalhes descritos a seguir neste capítulo), a qual executa uma busca com BLAST em bases de dados de proteínas completas (sem fragmentos). Neste trabalho utilizamos a base Uniprot Reference Proteomes. Com as sequencias recrutadas a ferramenta cria

uma árvore filogenética anotada taxonomicamente que pode ser aberta com o programa FigTree (RAMBAUT, 2009). Primeiro é escolhido o táxon mais distante para enraizar a árvore (LCA com menor número que aparece na legenda) e em seguida é feita o destaque de todas as sequencias de *Homo sapiens* com a função *Find*. Assim é identificado os ramos que apresentam proteínas distintas e o LCA de cada ramo. Em alguns casos, quando ocorre uma duplicação recente, anota-se o ramo que agrupa as duas e os sub-ramos que contêm cada uma, ou seja, coletando dois LCA. Em caso de as duas proteínas executarem a mesma função, o LCA do ramo que as agrupa seria o mais apropriado.

4.4. Colorização dos componentes da via

Para a melhor identificação visual da origem das proteínas da via, o programa ARCoBALeno foi utilizado. Através deste programa uma via fornecida pelo usuário no formato GPML (*Graphical Pathway Markup Language*) pode ter seus componentes coloridos de forma programática de acordo com as especificações do usuário listadas em um arquivo. Neste trabalho, as proteínas que compõem a via de infecção do Ebola foram coloridas de acordo com o nível de seu clado de origem determinado anteriormente (LCA). O diagrama da via no formato GPML (*Graphical Pathway Markup Language*) é um dos formatos exportados pelo software PathVisio e é utilizado como entrada no ARCoBALeno.

O programa ARCoBALeno também traz uma opção de representação de cores com base na última função molecular LMF (*Last Molecular Function*) das proteínas. Para realizar essa tarefa, o *software* usa uma base de dados local, onde estão armazenados os LCAs dos termos do *Gene Ontology* (GO). Cada gene está associado a diversas funções moleculares no GO, cada qual com uma distribuição taxonômica diferente e é possível a partir dela inferir o LCA. O resultado desse processo é a representação que permite indicar na via o momento em que cada gene adquiriu sua mais recente função molecular. Como as funções mais genéricas como enzima são mais distribuídas e pouco informativas, o uso da LMF visa capturar a função mais específica da proteína.

Todo o procedimento realizado até a representação em cores do último ancestral comum está representado na **Erro! Fonte de referência não encontrada.8:**

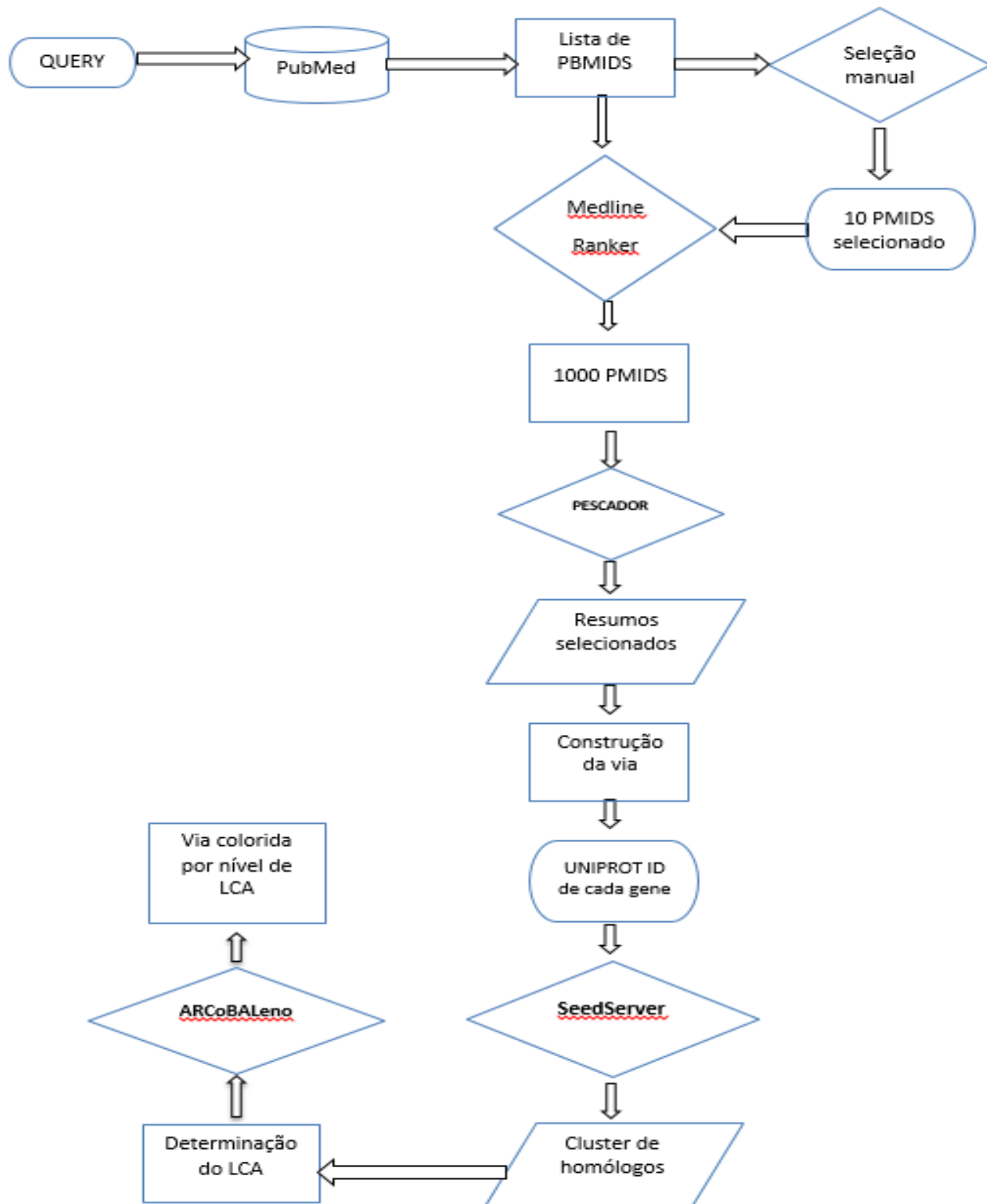


Figura 8 - Fluxograma da montagem da via. Sequência de ações tomadas até alcançar o resultado final da via colorida.

4.5. Análise de expressão diferencial

Para a análise da expressão diferencial, foram selecionadas séries de células infectadas com o vírus Ebola. As séries estão disponíveis no banco de dados GEO (*Gene Expression Omnibus - NCBI*), para a seleção foram avaliadas quanto à quantidade de amostras, presença de amostras controle e a tecnologia utilizada.

No fim da busca, foram selecionadas 2 séries que analisaram células infectadas com a cepa Zaire, que é considerada a maior responsável pelo último surto do vírus e a mais patogênica. As séries selecionadas foram:

- GSE86539 – Hepatócitos – Zaire;
- GSE80832 – Monócito – Zaire.

O primeiro tratamento das amostras foi através do pacote Agilent, específico da tecnologia de preparo das séries, onde foi feita a normalização e tratamento dos dados *outliers*.

A avaliação da qualidade das amostras foi feita através da PCA (Análise de componente principal), foi gerado também um dendograma⁸ onde apenas as que seguiam perfil de distribuição de amostras de acordo com o experimento, foram mantidas.

Foi montada a matriz *design* do experimento, onde as 2 séries foram analisadas separadamente. Todas as amostras foram comparadas contra o tempo zero ou controle, para a verificação do perfil de expressão. As séries foram analisadas baseadas em estatística bayesiana do pacote *Limma*, usando um *p-value* de 0.001 para limiar de significância da expressão diferencial.

Com o pacote *KEGGREST* foi realizada uma análise de enriquecimento dos genes encontrados diferencialmente expressos, juntamente com os encontrados por *text meaning*, contra a base de dados do KEGG. A partir deste pacote é possível

⁸ Diagrama que apresenta grupos formados em suas similaridades.

verificar quais vias do KEGG, tem correspondência com os genes dados de *input*. No entanto foi dado mais ênfase às vias com mais genes diferencialmente expressos.

4.6. TaxOnTree

Para complementar a análise evolutiva das proteínas relacionadas na via de infecção do vírus Ebola, o programa TaxOnTree foi utilizado para gerar árvores filogenéticas destas proteínas. Este *software* possui um pipeline de montagem de árvores filogenéticas e facilita a inspeção visual dos organismos e dos clados taxonômicos presentes nas árvores geradas.

Cada proteína da via foi carregada no programa através de seu identificador UniProt. A recuperação dos ortólogos putativos foi realizada através de uma análise de busca por proteínas similares utilizando o algoritmo blastp (ALTSCHUL *et al.*, 1990) contra o banco de dados de sequências provenientes do UniProt, mas contendo apenas aquelas sequências que constituem o proteoma completo dos organismos (Uniprot Reference Proteomes). Além disso, utilizamos o *threshold* de 50% de identidade, de modo que apenas proteínas com mais de 50% de identidade pudessem ser incorporadas na análise. Para as etapas posteriores, foram utilizados o MUSCLE (EDGAR, 2004), para o alinhamento de sequências, trimAl (CAPELLA-GUTIÉRREZ; SILLA-MARTÍNEZ; GABALDÓN, 2009), para a análise de qualidade do alinhamento, e FastTree (PRICE; DEHAL; ARKIN, 2009) para a inferência da árvore filogenética. Em todas as análises foram incluídas um máximo de 500 proteínas.

4.7. ELDOGraph

O ELDOGraph foi usado para gerar grafos que ilustrassem os ortólogos menos divergentes externos (ELDO - *External Least Divergent Ortholog*) das proteínas virais de cepas de Ebola. Para isso, 7 proteínas do Ebola (NP, L, VP30, VP24, VP35, VP40 e GP) de cinco cepas virais (Zaire, Sudan, Reston, Bundibugyo e Cotê d'Ivoire) foram recuperadas do banco de dados do Uniprot e submetidos a análise filogenética. Para a análise filogenética, cada grupo de ortólogos foi submetido ao software MUSCLE, para

o alinhamento de sequências, ao trimAl para a retirada de sítios do alinhamento de baixa qualidade, e FastTree para realizar a inferência filogenética. As árvores não-enraizadas foram utilizadas como dados de entrada para o programa ELDOgraph. Como resultado é gerado um grafo onde cada cepa encontra-se representada por um nó do grafo. Uma aresta que aponta para o nó “X” tendo como origem o nó “Y” indica que “X” possui um dada fração ou porcentagem de suas proteínas que tem como ELDO o nó “Y”. A frequência das proteínas de “X” que possui esta relação com o “Y” é indicada por uma escala de cores das arestas.

5. RESULTADOS

Neste trabalho foi almejada a construção da via de infecção do vírus Ebola e sua origem. Para tanto, partiu-se da coleta das informações presentes no estado da arte, que relacionaram o vírus e proteínas de contato direto ou indiretamente. Para que a via pudesse seguir uma linha de apresentação eficaz, foi feita a organização baseada nas funções e localização das proteínas nas células (**Erro! Fonte de referência não encontrada.9**).

Uma vez produzida a via, foi feito o estudo da evolução por meio da busca pelo LCA, objetivando uma inferência evolutiva que remontasse o surgimento da capacidade de penetração do vírus e o aparecimento de proteínas de defesa.

Para a construção da via de infecção de modo que objetivasse a interação proteína e vírus, as diferentes possibilidades de contato vírus e célula foram representadas em uma única figura, sendo posicionada, da esquerda para direita, a representação da sequência de processos.

5.1. O vírus na célula hospedeira

O vírus Ebola infecta os hospedeiros principalmente através da mucosa, acometendo uma grande variedade de células e tecidos; sua glicoproteína GP é a principal responsável pela adsorção do vírus nas células (ZHENG *et al.*, 2015).

O vírus consegue infectar quase todos os tipos de células, exceto linfócitos, no qual é induzida apoptose quando é feito o contato. Uma vez dentro do hospedeiro, as primeiras células a terem contato com o vírus são as fagocíticas: macrófagos, células dendríticas e monócitos, por representarem a primeira linha de defesa contra patógenos. O vírus tira proveito do processo de obtenção de nutrientes e renovação da membrana das células, a macropinocitose, para realizar a penetração (GROVE; MARSH, 2011).

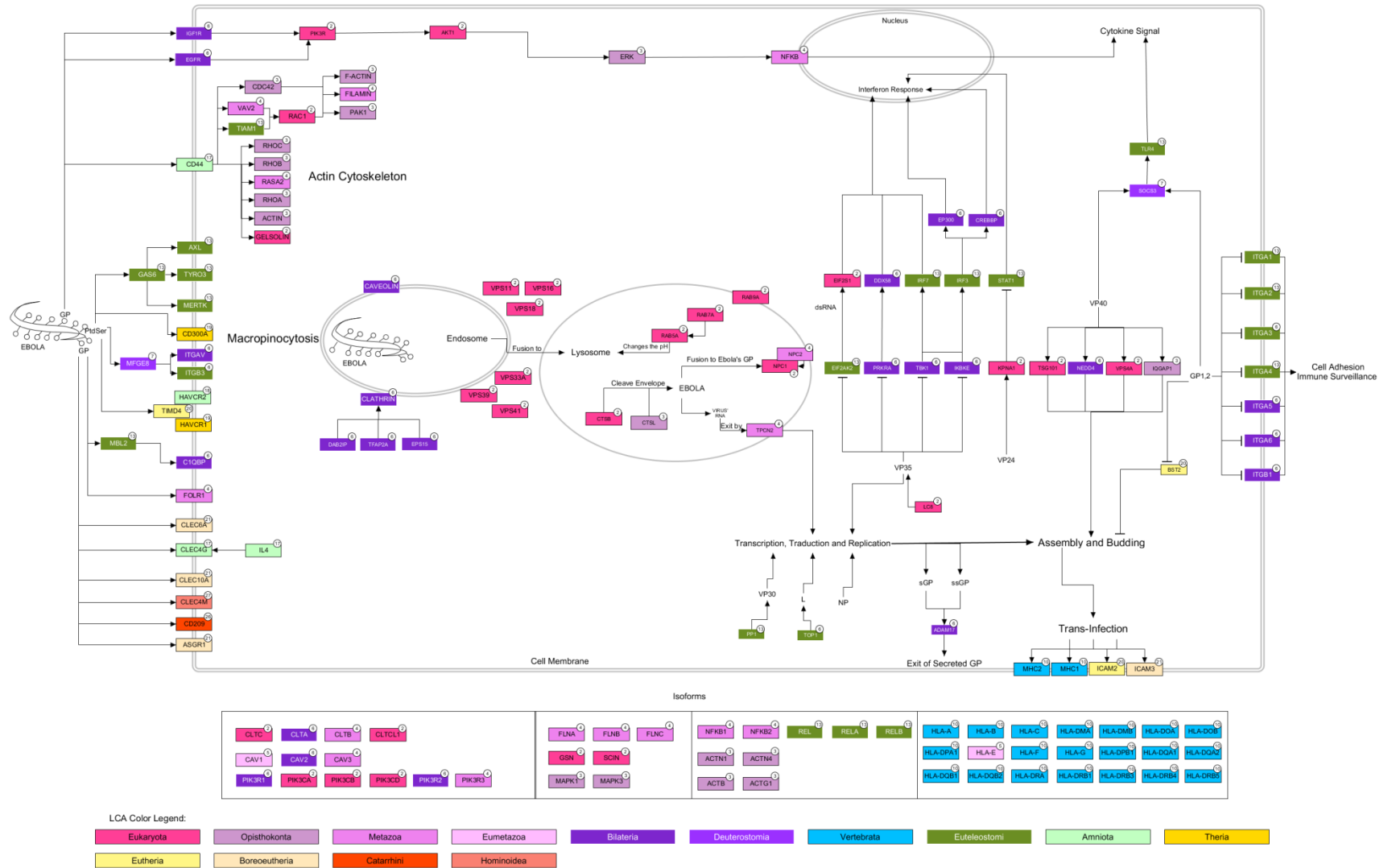


Figura 9 – Via de Infecção pelo vírus Ebola. Na figura está representado como retângulos menores as proteínas humanas, setas e tbar interações, proteínas virais estão representadas apenas por nome para melhor distinção. Algumas proteínas presentes na via, têm suas isoformas descritas embaixo da figura principal.

A glicoproteína é a parte do vírus que se conecta às células, utilizando dos mais diferentes receptores: lectinas, ligantes de manose, fator de crescimento epidermal entre outros, presentes na célula-alvo. Os meios pelos quais essas ligações são feitas, variam de acordo com a natureza do receptor. As proteínas presentes na membrana do hospedeiro a que o vírus se liga e podem funcionar como receptores para a adsorção nas células, segundo a literatura, são as: ASGR1, CD209, CLEC4M, CLEC10A, CLEC6A, FOLR1 e CD300A (ALVAREZ *et al.*, 2002; BHATTACHARYYA; HOPE; YOUNG, 2011; BRECHER *et al.*, 2012; BRUDNER *et al.*, 2013; CARETTE *et al.*, 2011; CHAN *et al.*, 2001; GRAMBERG *et al.*, 2008; MORIZONO, K; CHEN, 2014; SIMMONS, GRAHAM *et al.*, 2003; SINN *et al.*, 2003; TAKADA, AYATO *et al.*, 2004). A IL4 tem uma relação não muito bem esclarecida até o momento com o receptor CLEC4G, mas é descrito que esta interleucina estimula a expressão deste receptor na presença do Ebola, sendo um receptor para o vírus (DOMINGUEZ-SOTO *et al.*, 2008; DOMÍNGUEZ-SOTO *et al.*, 2009; GRAMBERG *et al.*, 2005, 2008; PIPIROU *et al.*, 2011; POWLESLAND *et al.*, 2008; ZHAO *et al.*, 2016). Foi notado que, já no momento de adsorção do vírus aos receptores encontrados, alguns só aparecem em Boroetheria, ASGR1, CLEC10A e CLEC6A. CD209 é mais recente, restrita a Catarrhini e CLEC4M, a Hominoidea. Somente CLEC4G daria suporte à ligação do vírus em Amniota, que é um clado recente na evolução e combina aves e mamíferos. De acordo com essa possibilidade, IL4 também é presente em Amniota. E CD300A aparece no clado seguinte a Amniota, Theria.

O vírus tem maneiras atípicas de adsorver e penetrar na célula, por meio da sua glicoproteína, na figura 10 é apresentado a etapa da infecção viral, mais detalhes são descritos a seguir.

A proteína MBL2, um ligante de manose, devido à sua característica de ligação a carboidratos, se liga na glicoproteína viral e em seguida quando é feita a ligação ao seu receptor C1QBP, o vírus consegue realizar a adsorção indiretamente. Apesar de o

receptor ser antigo, de Bilatéria, a MBL2 aparece em Euteleostomi, seria mais um meio responsável pela penetração já presente no ancestral de peixes e mamíferos (BRUDNER *et al.*, 2013; JI *et al.*, 2005).

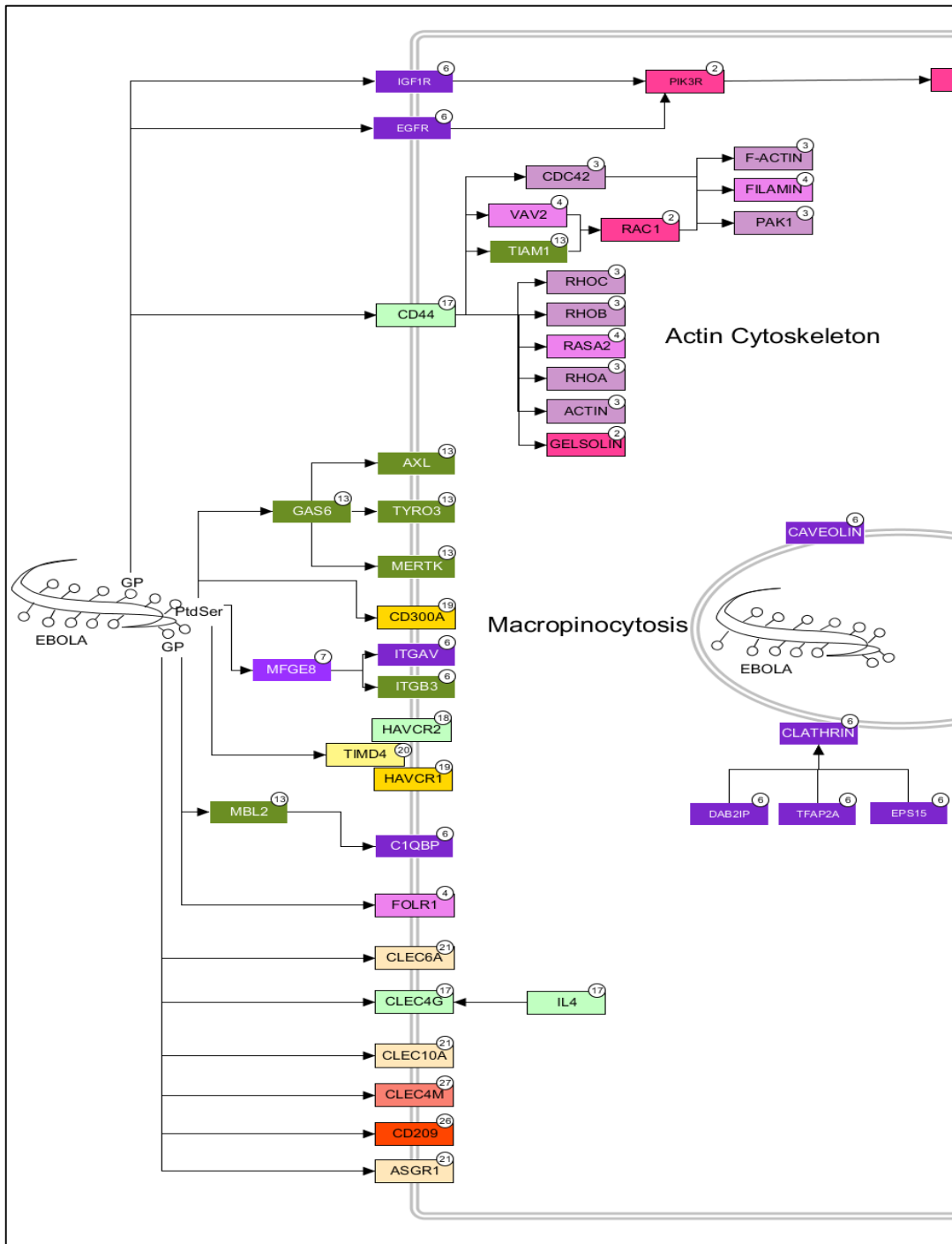


Figura 10 – Proteínas responsáveis pela adsorção viral.

A proteína humana TIMD4 é um ligante para fosfatidilserina, composto presente no envelope viral, e após esta interação, TIMD4 liga-se a HAVCR1 ou HAVCR2, estas que são proteínas de membrana, que culminam por outro meio de adsorção e posterior penetração viral (JEMIELITY *et al.*, 2013; KONDRATOWICZ *et al.*, 2011; KURODA *et al.*, 2015; LI, M. *et al.*, 2014; MIAO *et al.*, 2016; MOLLER-TANK *et al.*, 2013; MOLLER-TANK; MAURY, 2014; YUAN *et al.*, 2015). Como o mecanismo se inicia com uma proteína de Eutheria (TIMD4), já é mais recente e seria restrito a placentários. As proteínas de membrana HAVCR1 e HAVCR2 também são relativamente recentes, respectivamente de Theria e Aminiota.

MFGE8 é uma proteína humana ligante para fosfatidilserina, como dito uma molécula presente no vírus, que ao se ligar a ele se conecta ao dímero ITGAV/ITGB3, na membrana celular, possibilitando a adsorção (MORIZONO, K; CHEN, 2014). A ligante de fosfatidilserina remonta a Bilatéria, assim como ITGAV. A entrada pode acontecer em Bilatéria mas a depender de ITGB3 se limitaria e Euteleostomi.

A proteína CD300A (Theria), receptora humana direta do vírus, também se utiliza da ligação a fosfatidilserina como as descritas acima.

A proteína solúvel GAS6 envolvida em muitos processos de sinalização, se liga também a fosfatidilserina presente na glicoproteína viral, e após essa ligação se liga a um dos receptores MERTK, TYRO3 ou AXL, sendo feita de ponte para a adsorção do vírus (HUNT *et al.*, 2011; MOLLER-TANK; MAURY, 2014; MORIZONO, KOUKI *et al.*, 2012; SHIMOJIMA, M. *et al.*, 2006; SHIMOJIMA, MASAYUKI; IKEDA; KAWAOKA, 2007). Essas quatro proteínas existem desde os Euteleostomi, portanto um meio de infecção que estaria disponível no ancestral de peixes e animais de terra firme.

A proteína CD44 serve como receptor viral, que desencadeia a macropinocitose por citoesqueleto; este processo ocorre através da passagem por uma série de processos que modificam a estrutura de actinas da membrana para que se gere um compartimento e a penetração seja feita; envolvidas neste mecanismo estão as proteínas: TIAM1, VAV2, RAC1, F-ACTIN (isoformas ACTN1 e ACTN4), FILAMIN (isoformas FLNA, FLNB, FLNC), PAK1, CDC42, RHOC, RHOB, RASA2, RHOA, ACTIN (isoformas ACTB e ACTB1) e GELSOLIN (isoformas GSN e SCIN) (BHATTACHARYYA; MULHERKAR; CHANDRAN, 2012; GARRETT *et al.*, 2000;

MERCER; GREBER, 2013; QUINN *et al.*, 2009; SAEED *et al.*, 2008). Como a adsorção é mediada por CD44, o processo ficaria restrito a Amniota. Em seguida a mais recente seria TIAM1, de Euteleostomi. Quanto às demais, quando se trata de transdução de sinais ou processos parecidos, é comum que proteínas já existentes e participantes de muitos outros processos sejam recrutadas a participar. Assim, desta resposta participam ainda sete proteínas de Opisthokonta (ancestral de fungos e animais), três de Metazoa e duas de Eukaryota. Assim, à exceção do início da resposta em Amniota e participação de TIAM1 de Euteleostomi, o sinal é passado a proteínas bem antigas.

EGFR e IGF1R são receptores do tipo tirosina quinase envolvidos na resposta primária do vírus quando este entra em contato com a célula. Por meio da passagem de sinal, através de PIK3, AKT1, ERK e NFkB, como resultado citocinas e quimiocinas são liberadas (MARTINEZ, OSVALDO; VALMAS; BASLER, 2007; SAEED *et al.*, 2008; VANDE BURGT; KALETSKY; BATES, 2015; ZAMPIERI *et al.*, 2007). Os receptores estão presentes desde Bilatéria, a transdução de sinal passa por proteínas mais antigas, chegando ao fator de transcrição NF-KB, cuja maioria de suas isoformas é de Euteleostomi (NFkB1, NFkB2, REL, RELA e RELB), mas NFkB2 é de Metazoa. Assim esta seria uma sinalização já existente em Bilatéria.

Após ter feito a adsorção a uma célula hospedeira, o vírus pode penetrar por duas maneiras: macropinocitose ou endocitose, a parte da via que representa estes processos é mostrado na figura 11.

Na macropinocitose, após se ligar a uma das proteínas citadas anteriormente será feita a penetração do vírus que em seguida seguirá para o endossomo; este processo pode ser dependente de Caveolina (CAV1, CAV2 ou CAV3) ou Clatrina (CLTC, CLTA, CLTB ou CLTCL1), proteínas associadas à formação da estrutura do compartimento de internalização (ALEKSANDROWICZ *et al.*, 2011; BHATTACHARYYA; HOPE; YOUNG, 2011; EMPIG C.J. GOLDSMITH M.A., 2002; HUNT *et al.*, 2011; SANCHEZ, 2007). Auxiliam na formação da vesícula criada pela Clatrina, as proteínas estruturais: DAB21P, TFAP2A e EPS15, todas datadas em Bilatéria, assim como a Clatrina. A CAV3 é a mais antiga, já presente em Metazoa, e CAV1 e CAV2 atuam junto, o que pode ocorrer já em Bilatéria. A ancestralidade das clatrininas variam de Eukaryota a Metazoa, é uma ocorrência antiga qualquer que seja isoforma participante.

Já o processo da endocitose é feito através da modificação do citoesqueleto de actina, onde através da passagem dos sinais, que tem início na CD44, ocorre o processo de invaginação da membrana, que se tornará o compartimento para a degradação do vírus. As isoformas de actina alfa-cito actina (ACTB) e beta-cito actina (ACTG1) são de Metazoa, um processo antigo.

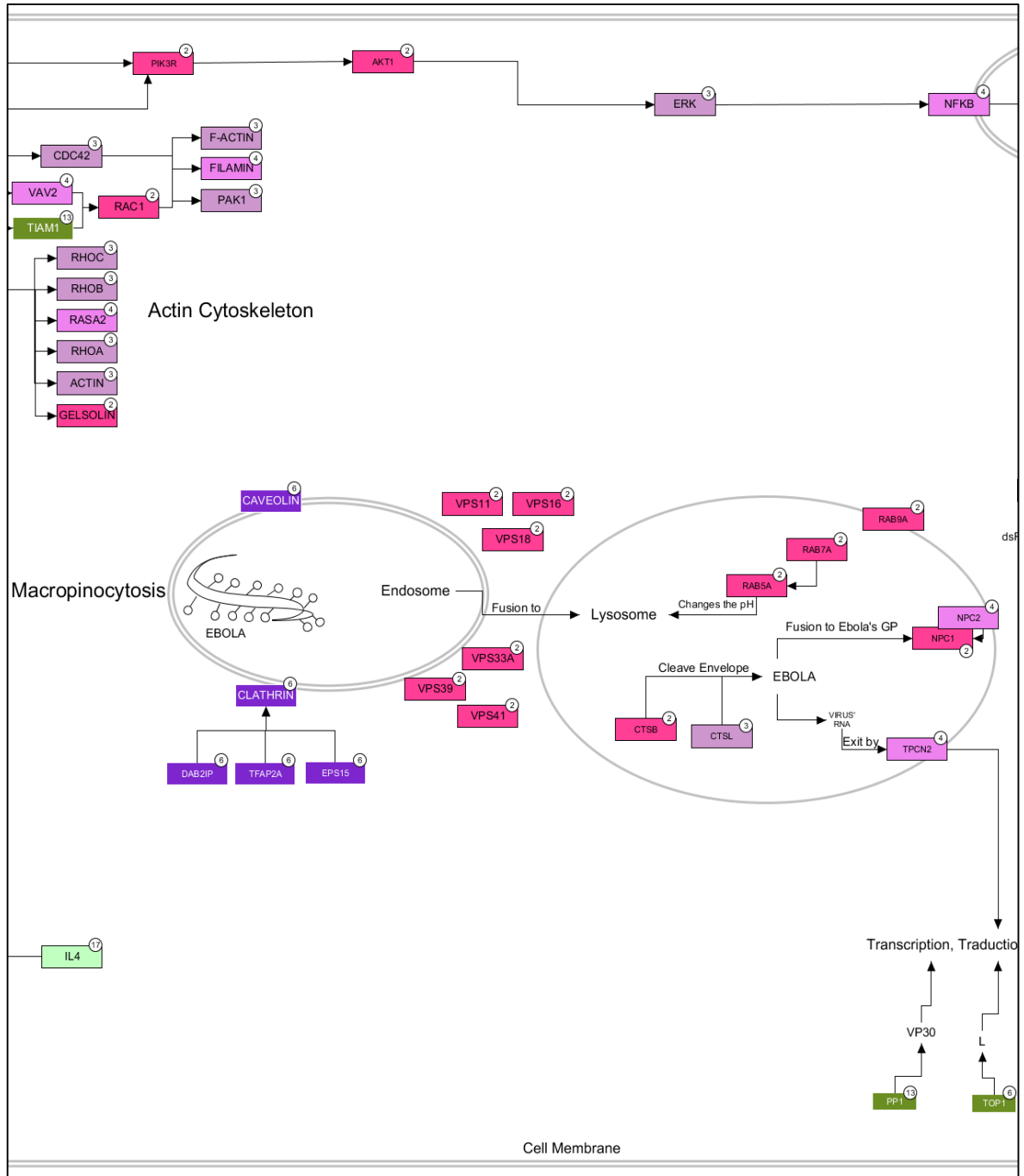


Figura 11 – Proteínas envolvidas na penetração e desnudamento viral.

Para a manutenção da estrutura do endossomo estão envolvidas as proteínas no complexo HOPS (VPS11, VPS16, VPS18, VPS33A, VPS39 e VPS4), associadas também à maturação deste (CARETTE *et al.*, 2011; MARTINEZ, O. *et al.*, 2013; MILLER *et al.*, 2012). São todas antigas, existindo desde Eukaryota.

Para a lise do vírus internalizado, o endossomo é fundido em lisossomo, onde as características do meio começaram a mudar. No lisossomo as proteínas RAB5A, RAB7A e RAB9A, são responsáveis pela sua estrutura, enquanto as catepsinas, CTSL e CTSB, tentam clivar o envelope do vírus. Porém devido à natureza do envoltório proteico, ele tende por sobrepujar a defesa celular liberando partes funcionais do vírus no lisossomo. A clivagem da glicoproteína viral nesta etapa, tem grande importância na fusão proporcionada entre a GP e o transportador de colesterol NPC1 e NPC2, que é um dos responsáveis pela possibilidade de desnudamento viral no citoplasma da célula e o material genético, este último sendo transposto ao citoplasma pelo canal de cálcio do lisossomo, TPCN2 (BHATTACHARYYA; MULHERKAR; CHANDRAN, 2012; BRECHER *et al.*, 2012; CARETTE *et al.*, 2011; CHEN *et al.*, 2004; GONG *et al.*, 2016; MISASI *et al.*, 2012; MURRAY *et al.*, 2005; NDUNGO *et al.*, 2016; NG *et al.*, 2015; SAKURAI *et al.*, 2016; SIMMONS, JAMES A *et al.*, 2016; SPENCE JS, KRAUSE TB, MITTLER E, JANGRA RK, 2016; WANG *et al.*, 2016). Todo este sistema envolve proteínas de origem em Eukaryota a Metazoa.

Uma vez no citoplasma da célula hospedeira o vírus inicia seu processo de expressão gênica e montagem/maturação de novas partículas virais. A proteína VP30 é um importante regulador deste complexo; ela precisa ser desfosforilada pela proteína PP1, para que não interfira na interação entre o nucleocapsídeo (NP) e a VP35. A topoisomerase, TOP1, interage diretamente com a RNA polimerase (L), sendo necessária para a atividade polimerase (BIEDENKOPF; LIER; BECKER, 2016; ILINYKH *et al.*, 2014; MARTIN; CANARD; DECROLY, 2017). Esse processo só poderá ocorrer em Euteleostomi, já que PP1 e TOP1 têm origem neste clado.

A atividade de replicação do genoma viral é aumentada com a interação de LC8 e VP35. Inicialmente são feitas transcrições do RNA viral em mRNA e, posteriormente traduzido, no entanto, com o acúmulo das proteínas virais no citoplasma celular, o vírus passa então a replicar o RNA viral *negative sense*, para a completa maturação de

novos vírus (BIEDENKOPF; LIER; BECKER, 2016; ILINYKH *et al.*, 2014; MARTIN; CANARD; DECROLY, 2017). LC8 já está disponível, no entanto, desde Eukaryota.

Todos os genes do Ebola são monocistrônicos, exceto o da GP, que codifica as formas solúveis da GP, sGP e ssGP, pequenos peptídeos que têm uma função importante na resposta imune, que são liberados no espaço extracelular com o auxílio da proteína ADAM17 (DOLNIK *et al.*, 2004; MEHEDI *et al.*, 2011). Esta é de Bilatéria.

As proteínas virais modulam a atividade celular do hospedeiro para que ocorra o completo ciclo de replicação e não haja interferência por mecanismos de defesa, tendo assim liberdade para desempenhar seu ciclo completo, como será apresentado adiante.

5.2. Modulação da resposta imune e trans-infecção

A resposta imune desencadeada pelo vírus Ebola é alternada. No início da infecção uma resposta inflamatória é produzida pelo hospedeiro, através da liberação de citocinas, após a interação com o receptor IGF1R e EGFR. Uma vez iniciada a infecção a resposta inflamatória tende a diminuir, sendo restabelecida apenas quando as proteínas virais chegam ao citoplasma da célula hospedeira. Cada proteína viral tem uma participação notável neste momento, conforme é descrito adiante deste capítulo.

A VP35, proteína viral que tem um essencial papel na replicação e transcrição, age na resposta anti-interferón produzida pela célula. Um dos mecanismos mais modernos produzidos pela célula para neutralizar um vírus invasor, é a tentativa de liberação de um micro RNA complementar ao material genético do vírus, o que ativa a degradação de dsRNA (RNA de fita dupla). No entanto, este mecanismo é reprimido, através da modulação das proteínas EIF2AK2 e EIF2S1, PRKRA e DDX58, pela VP35, que seriam as proteínas responsáveis pela reação de sinalização e ativação de defesa por interferón antiviral (CARDENAS *et al.*, 2006; FENG *et al.*, 2007; LUTHRA *et al.*, 2013). O primeiro par de proteínas tem a sua segunda proteína já existente no nível Eukaryota, mas a regulação pela EIF2AK2 só aparece em Euteleostomi. Já o segundo é datado em bilatéria.

Outra ação anti-interferón desempenhada pela VP35 é através da modulação de

duas cinases, TBK1 e IKKε, que são precursoras de resposta por interferón alfa e beta. Estas agiriam através da ativação dos fatores transcrição IRF3, que ainda ativa EP300 e CREBBP, e IRF7, onde os sinais de ambos culminam agindo no núcleo celular ativando a resposta por interferón (AUDET; KOBINGER, 2015; BASLER; AMARASINGHE, 2009; HARTMAN *et al.*, 2006, 2008; KIMBERLIN *et al.*, 2010; LEUNG *et al.*, 2011; PRINS; CARDENAS; BASLER, 2009; SCHUMANN; GANTKE; MUHLBERGER, 2009). Percebe-se que esta ação utiliza proteínas não tão antigas, com origem em Bilateria, mas só se exerce com as que se originariam em Euteleostomi, IRF3 e IRF7.

Na figura 12, têm-se a representação dos processos de modulação imune, durante a infecção por Ebola.

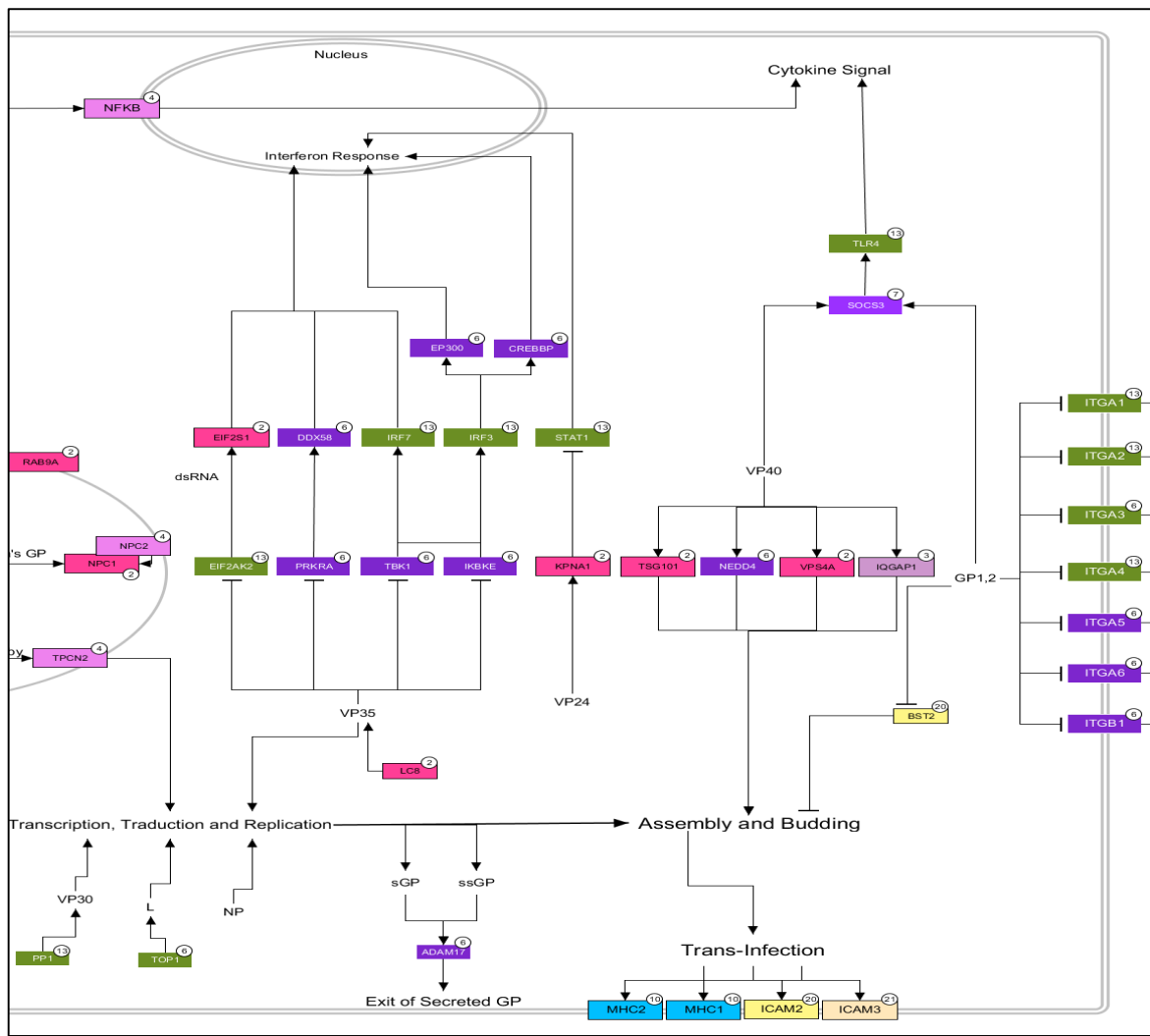


Figura 12 – Proteínas responsáveis pela modulação imune do hospedeiro.

A proteína VP24 antagoniza a resposta por interferón através da ligação a KPNA1, o que evita o transporte do fator de transcrição de interferón, STAT1, para o núcleo da célula (LEE; LEBEDA; OLSON, 2009; OKUMURA, A. *et al.*, 2010; REID *et al.*, 2006; ZHANG *et al.*, 2012). KPNA1 é de Eukaryota, mas uma vez mais, a resposta só se estabelecerá com STAT1 em Euteleostomi.

A liberação de citocinas é uma maneira que a célula infectada tem de sinalizar o seu estado, podendo assim ativar uma resposta de defesa, a qual o vírus combate. A proteína SOCS3 é uma proteína de supressão de citocinas, sua expressão é induzida pela GP e VP40, e uma vez expressa, evita a ação do TLR4, que ativa liberação de citocinas e mediadores inflamatórios de forma geral (DENNER, 2015; OKUMURA, A. *et al.*, 2010; OKUMURA, ATSUSHI *et al.*, 2015). SOCS3 é de Bilatéria mas já TLR4 aparece em Euteleostomi apenas.

Uma forma de comunicação entre as células e de manutenção de estruturas, são as proteínas de adesão celular, como as integrinas, que passam sinais de sobrevivência às células vizinhas. Essa comunicação é modulada pela proteína GP, quando já presente no citoplasma, para que a célula que está infectada não seja detectada. Relacionadas à via de infecção do vírus Ebola estão as seguintes integrinas: ITGA1, ITGA2, ITGA3, ITGA4, ITGA5, ITGA6 e ITB1 (VANDE BURGT; KALETSKY; BATES, 2015; YAN *et al.*, 2015). As integrinas citadas têm origem em Bilatéria (três delas) e Euteleostomi (quatro delas). Assim, é um processo que se inicia em Bilatéria, mas se efetiva em Euteleostomi, como muitos mencionados acima.

Os mamíferos aperfeiçoaram uma forma de resposta específica de tentar evitar a replicação de partículas virais, através da proteína BST2 (ou CD317), que tem como ação evitar com que o vírus saia da célula e o direciona para o lisossomo para ser lisado. Tal proteína já foi relatada tendo ação contra outros vírus, como HIV. Porém, esta resposta não consegue ser eficaz contra o Ebola, devido à modulação da ação da BST2 pela GP (AUDET; KOBINGER, 2015; BRINKMANN *et al.*, 2016; DOUGLAS *et al.*, 2010; GOFFINET *et al.*, 2010; KALETSKY *et al.*, 2009; TETHERIN *et al.*, 2009; VANDE BURGT; KALETSKY; BATES, 2015). A BST2 é restrita a mamíferos placentários, Eutheria, um mecanismo de defesa extremamente recente e que já é combatido pelo vírus.

Uma forma de modulação de resposta imune extracelular é desempenhada pelo produto secretado da Glicoproteína viral, sGP e/ou ssGP (variação transcricional), que é encontrada ativando células não infectadas, fazendo com que haja a produção de citocinas e ativação imune contra células que têm esses peptídeos internalizados, evitando com que as células verdadeiramente hospedeiras do vírus não sejam reconhecidas com uma resposta específica, além da modulação do mecanismo de opsonização desempenhado pelos anticorpos (MEHEDI *et al.*, 2011).

As células apresentadoras de antígenos mesmo infectadas, tendem a fazer a tentativa de apresentação ao sistema imune para que ocorra a resposta específica contra o patógeno. Porém, ao tentar fazer o contato com a célula infectada com o vírus, os linfócitos entram em apoptose, pela ação viral. O mecanismo ainda não está bem definido, mas a linfopenia é uma característica já bem definida durante a infecção pelo vírus Ebola, o que representa a deficiência de resposta do hospedeiro contra esse patógeno. Estão envolvidos nesse processo de apresentação as proteínas ICAM2, ICAM3, e os complexos MHC1 e MHC2 (ALEKSANDROWICZ *et al.*, 2011; MASON; TARR, 2015; MELANSON; KALINA; WILLIAMS, 2014; SULLIVAN *et al.*, 2005; VLIET *et al.*, 2000). Os complexos MHC são formados por muitas isoformas, datados na sua maioria como de Vertebrada, mas por sua vez ICAM2 é de Eutheria e ICAM3 de Boroetheria (excluídos os Afrotheria, como os elefantes). Assim aprecia-se o aparecimento de mecanismos de interação com o sistema imune bem recentes.

5.3. Processos e famílias proteicas envolvidas na infecção

Algumas famílias, complexos e processos especializados de proteínas foram encontrados na construção da via de interação do vírus Ebola, sendo apresentados abaixo.

A família TIM, conhecida por ser um regulador de apoptose, é também um ligante de fosfatidilserina, composto este presente também no envelope do vírus, responsável pela adsorção às proteínas presentes na membrana. Os representantes dessa família presentes no interatoma do Ebola são: TIMD4, HAVCR2 e HAVCR1, que facilitam a internalização do vírus.

A família TAM, que tem sua função associada ao crescimento e sobrevivência da célula, é também ligante de fosfatidilserina. Através da ligação da proteína GAS6 à fosfatidilserina, é formada uma ponte entre o vírus e os receptores, AXL, TYRO3, MERTK, para a adsorção indireta do vírus.

As lectinas agem em diversos processos, sua função está associada à capacidade de se ligar a carboidratos. Devido a esta característica o vírus Ebola, através de sua glicoproteína, consegue aproveitar-se destas proteínas presentes na membrana: ASGR1, CD209, CLEC4M, CLEC10A, CLEC6A e CLEC4G.

As integrinas são proteínas transmembrana, que participam de sistemas complexos de sinalização, relacionados à proliferação, migração, longevidade celular, entre outros, e principalmente tendo uma função relevante no sistema imune. Em um modelo de infecção, a célula hospedeira utilizaria deste mecanismo para passar o sinal de seu estado ao sistema imune, através das ITGA1, ITGA2, ITGA3, ITGA4, ITGA5, ITGA6 e ITGB1, por exemplo, presentes. Porém, dentre as modulações imunológicas desempenhadas pelo vírus, para sobressair às defesas, a sua glicoproteína já na sua forma livre, age reduzindo a expressão das integrinas interrompendo a comunicação celular.

A família das caveolinas participa de vários processos celulares, como transporte vesicular, transdução de sinal, supressão de tumor entre outros. Relacionado à via de infecção é importante a função de transporte vesicular exercida na endocitose, onde as proteínas CAV1, CAV2, CAV3, possibilitam a penetração do vírus a partir da criação da vesícula, direcionando o Ebola ao endossomo.

A família das clatrininas está envolvida no revestimento da membrana, e tem como representantes na via as proteínas CLTC, CLTA, CLTB e CLTCL1. A endocitose mediada por clatrina é um dos processos utilizados pelo vírus Ebola para sua penetração.

O complexo HOPS, representado na via pelas proteínas VPS11, VPS16, VPS18, VPS33A, VPS39 e VPS41, medeia a fusão do endossomo e lisossomo, e tem efeito na maturação do endossomo, com associação indireta com o vírus.

As famílias e complexos aqui listados têm a função de exemplificar as interações e modulações funcionais desencadeadas pelo vírus durante a infecção, de forma direta

ou indireta. Estes aspectos têm extrema relevância para o entendimento do desenvolvimento da infecção em função dos locais afetados, sintomas e como evitá-los.

5.4. Origem da via de infecção pelo vírus ebola

A origem da via de infecção do vírus foi feita inicialmente através de rotinas automatizadas realizando agrupamentos de homólogos com SeedServer e executando o algoritmo LCA. Foram feitas comparações com a ferramenta ARCoBALeno com o LCA determinado por duas bases de dados, *GeneAges* e *Kegg Orthology*. Entretanto, para maior confiabilidade dos resultados, optou-se por apresentar a ancestralidade dos genes definida por conferência manual utilizando a ferramenta TaxOnTree. Para referência, as demais determinações estão na Tabela suplementar T2.

A via de infecção pelo vírus Ebola colorida de acordo com a origem evolutiva de suas proteínas é mostrada na imagem abaixo (**Erro! Fonte de referência não encontrada.**¹³). Para esta forma de apresentação com muitas cores, recomenda-se a análise em uma apresentação PDF ou em PowerPoint, estes arquivos estão disponíveis em biodados.icb.ufmg.br/ebola.

Ao analisar o resultado do surgimento das proteínas, foram buscados pontos em que os processos apresentados estariam mais desenvolvidos e próximos à completude funcional. Partindo deste critério, foram separados três intervalos ao longo do estudo evolutivo para se dar enfoque, apresentados na **Erro! Fonte de referência não encontrada.**¹³. Os genes aparecem coloridos em três opções de cores, para orientação quanto a seu clado de origem, um círculo com um número referente ao nível evolutivo foi adicionado. Os cladogramas são especificados por números.

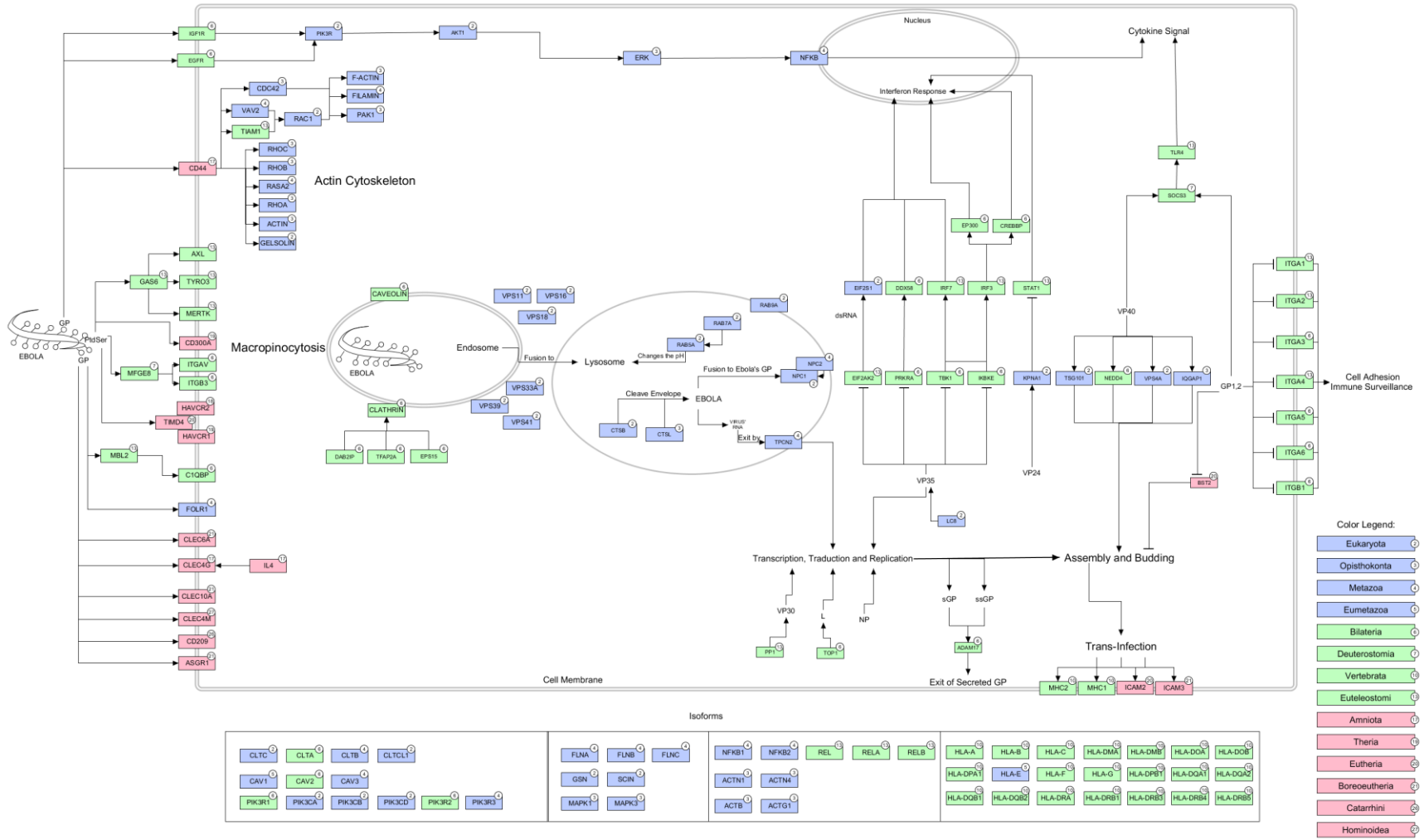


Figura 13 - Via de infecção do Ebola, representada em três intervalos ao longo da evolução. De Eukaryota a Eumetazoa, representado de azul. De Bilatéria a Euteleostomi, representado de verde. De Aminiota a Hominoidea, representado de rosa.

Conforme mostrado na **Erro! Fonte de referência não encontrada.**, as famílias e processos de proteínas foram surgindo em uma ordem cronológica, mantendo certa proximidade entre as que se assemelham em sua função.

O primeiro intervalo definido foi de Eukaryota a Eumetazoa, nível evolutivo que inclui os ancestrais de eucariotos até o ancestral de todos os animais, com legenda cor azul. Nesse nível é destacado principalmente os processos internos realizados para o desnudamento do vírus. Estão presentes nesse ponto de evolução, por exemplo, as proteínas que realizam a lise do envoltório proteico, as catepsinas. As que auxiliam a formação do endo-lisossomo, além da formação de suas condições de degradação específicas, como RAB's que alteram o pH. Membros do processo de endocitose por citoesqueleto começam a surgir, faltando poucos genes para este processo ser concluído.

O segundo intervalo destacado é entre Bilatéria e Euteleostomi, representado na cor verde, ponto este que reúne nosso ancestral com a *Drosófila* até nosso ancestral com peixes. São destacados neste marco evolutivo, os processos de adsorção quase completos, somados as proteínas envolvidas na penetração, o que traz consigo grande possibilidade de atividade de infecção completa, baseada em representação dos processos. A família de receptores de membrana TIM está completa, os módulos que começaram a surgir no ponto passado citado, aparecem completos relativos à endocitose por citoesqueleto. Os complexos de manutenção do endossomo, Clathrin, Caveolin, Hops estão presentes. Genes relativos à defesa do hospedeiro aparecem completos, como as integrinas, precursores de resposta por interferón como IRF3 e IRF7, entre outros.

No último e mais recente intervalo, que abrange Sarcopterygii a Hominoidea, divisão esta que compreende desde os peixes pulmonados até os “grandes símios”, representou-se na imagem pela cor rosa. Neste ponto evolutivo surgem as lectinas, a última parte de receptores que ainda não estava presente, completando todas possibilidades de adsorção citadas no presente trabalho. As ICAM's surgem, proteínas

ligadas a trans-infecção celular, um meio de infecção do vírus de uma célula para outra.

5.4. Análise de expressão diferencial de séries infectadas por ebola

Foi desenvolvida a análise de expressão diferencial de duas séries de células infectadas, hepatócito e monócito, pelo vírus Ebola da cepa Zaire disponíveis no banco de dados NCBI-GEO: GSE86539 e GSE80832, respectivamente. Foi feita a comparação do perfil de expressão do controle e as células passando por infecção viral em vários tempos.

Para cada série foi obtido o resultado de comparação do tempo 0 e a amostra ao longo do *time course* do experimento. Abaixo é apresentado o delineamento de cada experimento:

Tabela 3 - Resultado dataset 1.

Série	Tipo celular	Delta	Genes Dif Exp
GSE80832	Monócito	8h-0h	1632
GSE80832	Monócito	18h-0h	0
GSE80832	Monócito	30h-0h	1645

Tabela 4 - Resultado dataset 2.

Série	Tipo celular	Delta	Genes Dif Exp
GSE86539	Hepatócito	8h-0h	1157
GSE86539	Hepatócito	24h-0h	1209
GSE86539	Hepatócito	48h-0h	2070
GSE86539	Hepatócito	72h-0h	3065

A partir deste resultado foi feito o enriquecimento dos genes diferencialmente expressos juntamente com os genes encontrados na via aqui proposta por *text meaning*, através do pacote KeggRest do Rstudio, que buscou pelas vias no Kegg que mais tinham recobrimento com os genes diferencialmente expressos. Os resultados deste enriquecimento são apresentados a seguir:

Na **Erro! Fonte de referência não encontrada.** 14, são apresentados os genes que mais tiveram sua expressão alterada ao longo da análise, para a série GSE80832. No eixo Y é representado o número de genes, com o limiar de corte de significância de 10^{-3} . No eixo X, as vias que corresponderam aos genes em cada tempo.

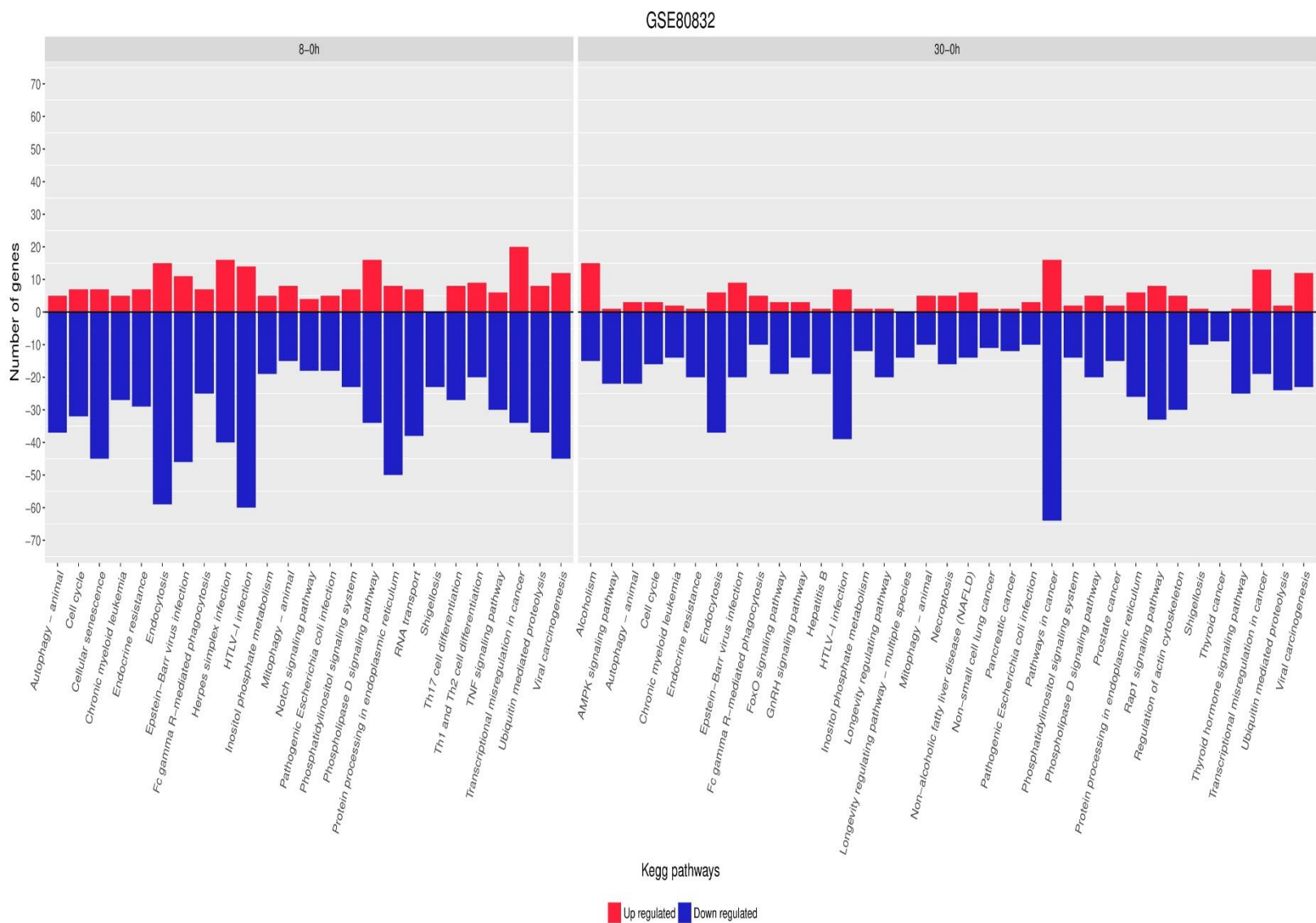


Figura 14 - Representação da expressão dos genes da série GSE80832. No eixo Y há os números de genes e no eixo X as vias correspondentes do KEGG. Cor azul representa *down-regulated*, cor vermelha *up-regulated*.

Na **Erro! Fonte de referência não encontrada.5**, são apresentados os resultados para a série GSE86539, considerando os mesmos critérios para a série anterior. Uma das diferenças no perfil apresentado por cada *dataset* é relativo ao *time-course* usado no experimento.

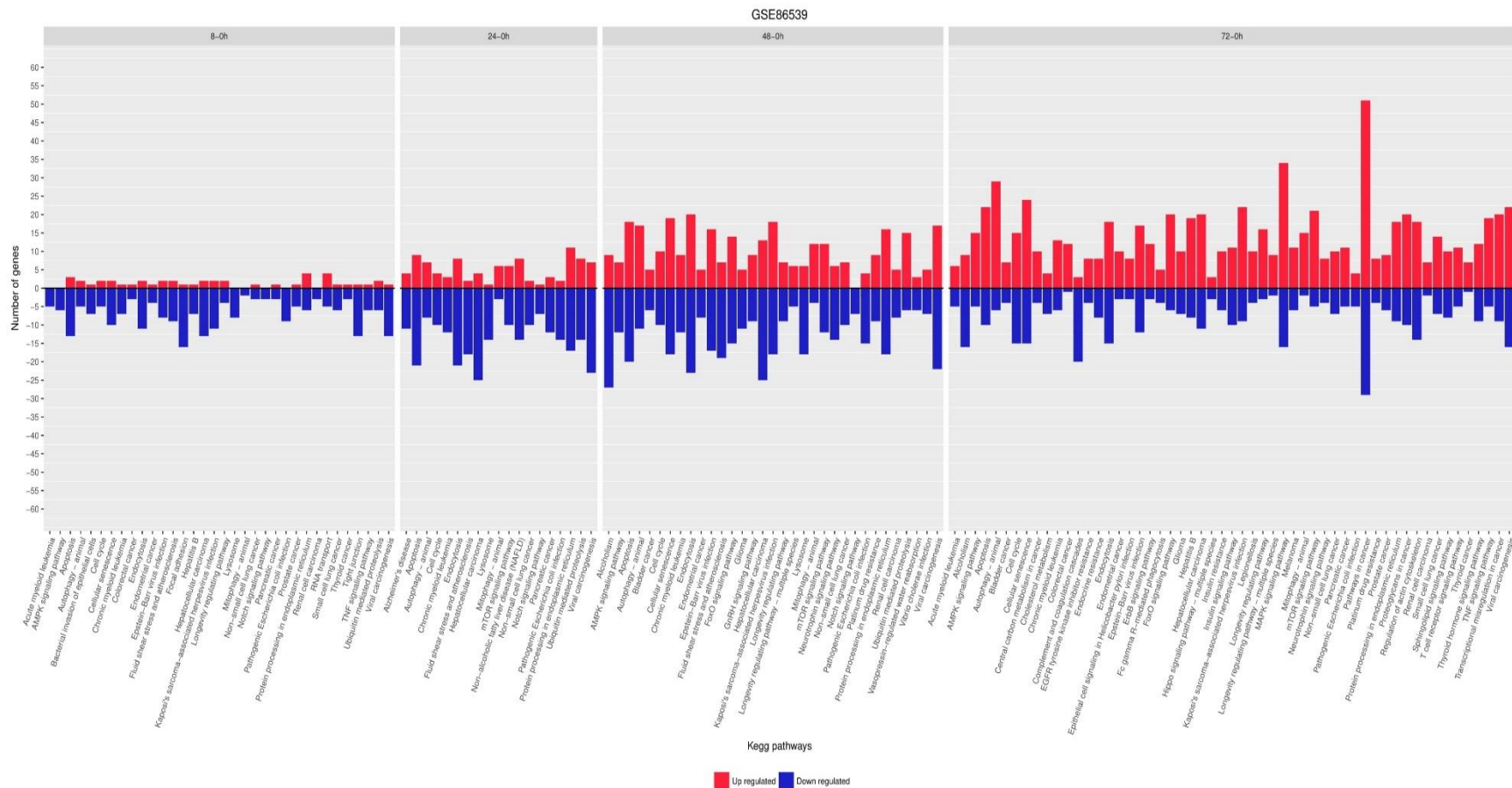


Figura 15 - Representação da expressão dos genes da série GSE86539. No eixo Y há os números de genes e no eixo X as vias correspondentes do KEGG. Cor azul representa *down-regulated*, cor vermelha *up-regulated*.

Dentre as vias que tiveram expressão alterada presentes, foi dado destaque especial à da endocitose, por ter representação em todos os tempos com genes diferencialmente expressos e pela relação direta ao processo de infecção viral na célula-alvo. Abaixo são apresentadas figuras, relativas às alterações passadas pelos genes da via, ao longo dos experimentos. Os genes da via da endocitose receberam cores, genes com aumento da expressão (vermelho) e genes com redução da expressão (azul), relação que representa a alteração da expressão, tendo um *cut-off*⁹ mínimo de -0,1 e +0,1.

Na **Erro! Fonte de referência não encontrada.6**, é apresentado a endocitose para o experimento GSE80832, realizado com monócitos infectados com o Ebola. Nos genes diferencialmente expressos no intervalo entre 0 e 8 horas, destacam-se em especial os seguintes: MHC1, FOLR1 e CAVEOLIN, apontados como envolvidos na adsorção e penetração do vírus, e que na via aparecem com a expressão elevada. Possivelmente ligada ao contato do vírus com a membrana celular e os processos desencadeados como consequência.

⁹ Valor referente ao limiar de aceite, ponto de corte.

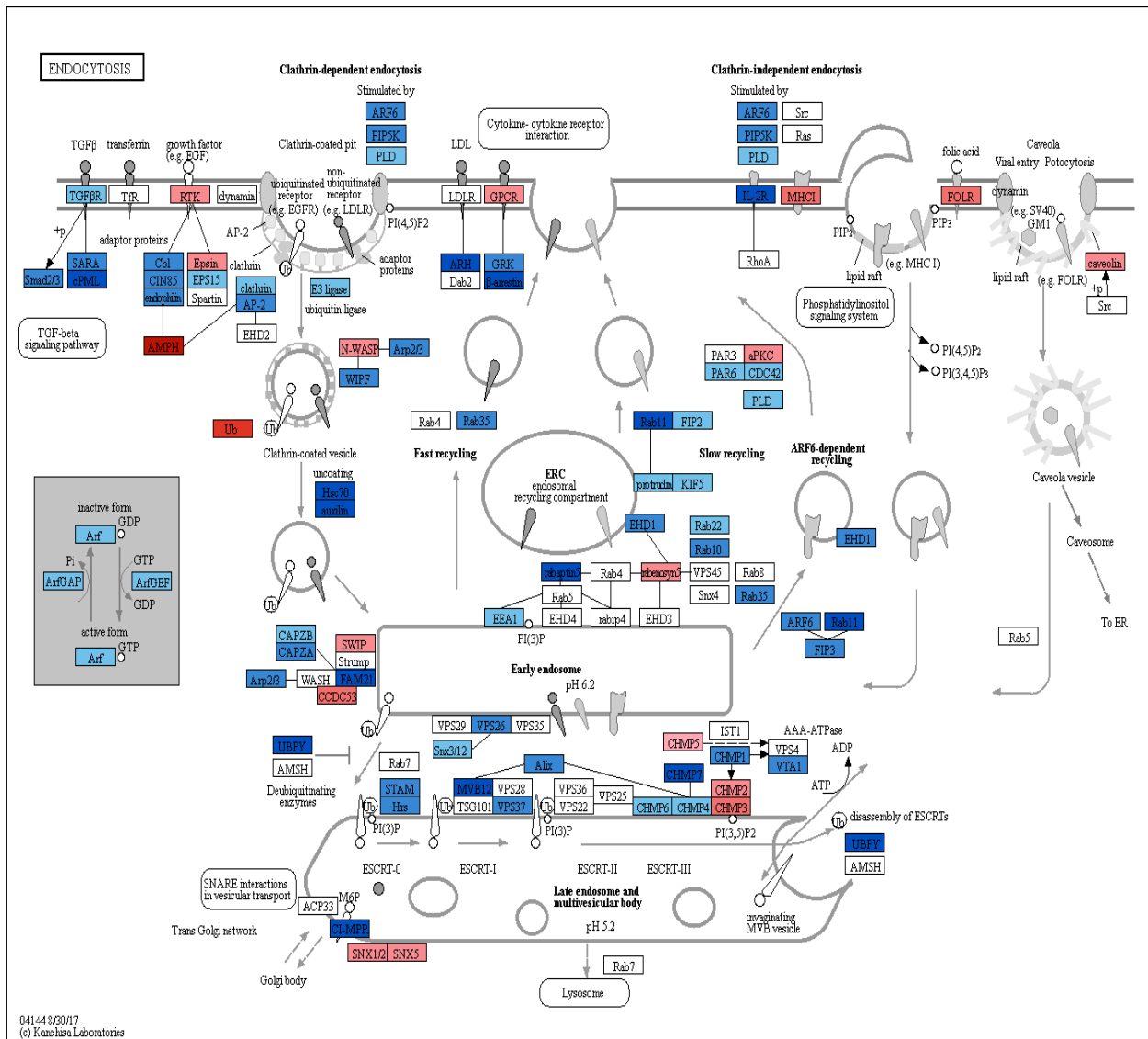


Figura 16 - Endocitose para a série GSE80832 no intervalo de 8 horas. Via da endocitose extraída do repositório KEGG Pathways, com os genes que tiveram alteração dos níveis de expressão. Aumento em escala de vermelho e redução em escala de azul.

No próximo intervalo do experimento GSE80832, não houve genes expressos que satisfizessem o *cut-off* escolhido, logo não haverá via apresentada.

No próximo intervalo, **Erro! Fonte de referência não encontrada.**7, relativa ao experimento GSE80832, no tempo de 30h, os genes MHC1 e CHM estão com sua expressão aumentada, ambos estão relacionados a formação de vesículas. O receptor FOLR1 tem uma redução na sua expressão apresentada inicialmente.

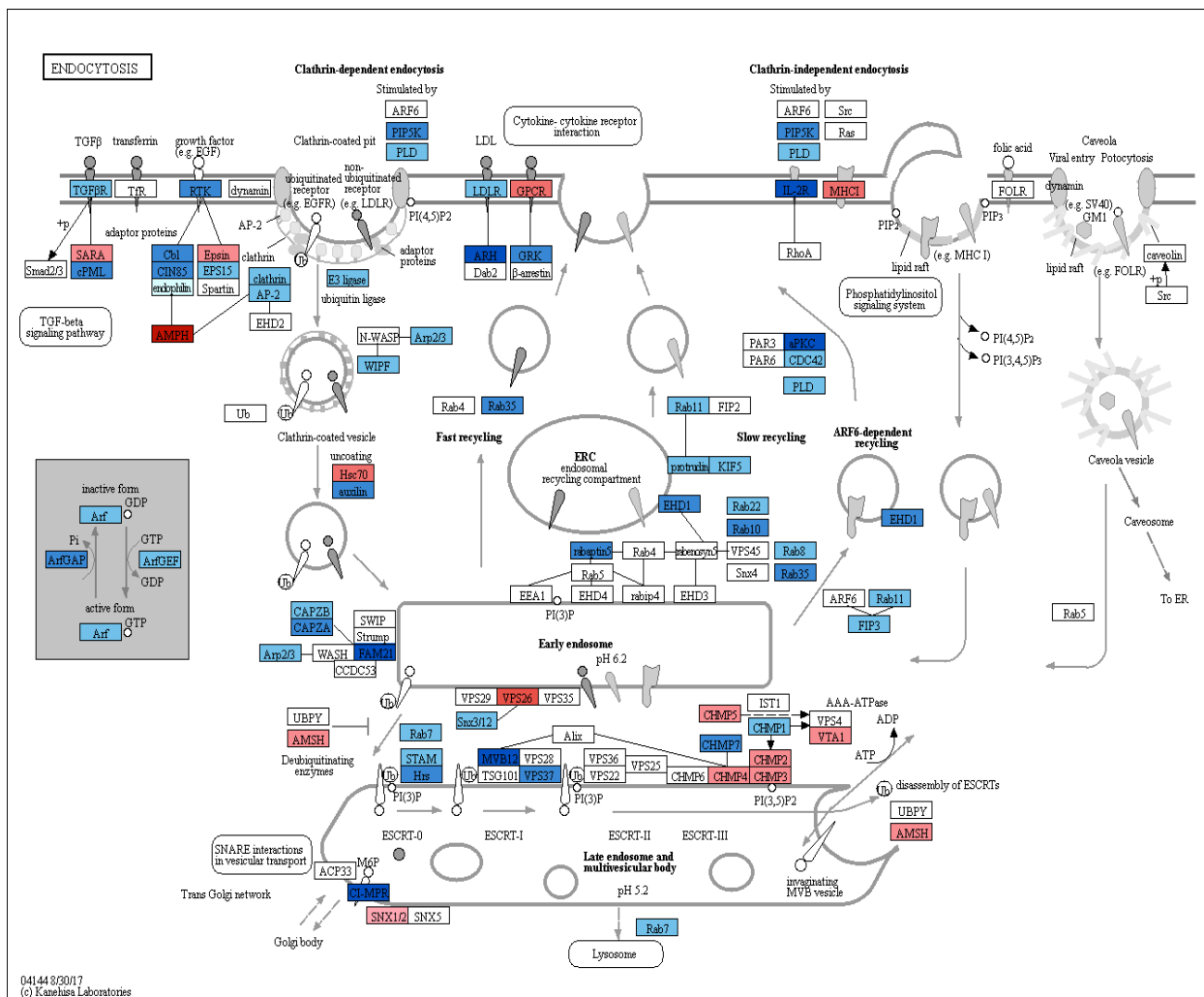


Figura 17 – Endocitose para a série GSE80832 no intervalo de 30 horas. Via da endocitose extraída do KEGG Pathways, com os genes que tiveram alteração dos níveis de expressão. Aumento em escala de vermelho e redução em escala de azul.

No próximo experimento GSE86539, com células de hepatócito infectadas, a via de endocitose apresentou variação da expressão em todos os intervalos do experimento.

Na Figura 188, está representado o intervalo de 0 a 8 horas, neste modelo de infecção são apresentados com expressão elevada genes do complexo Caveolin e Hops, ligados à penetração viral.

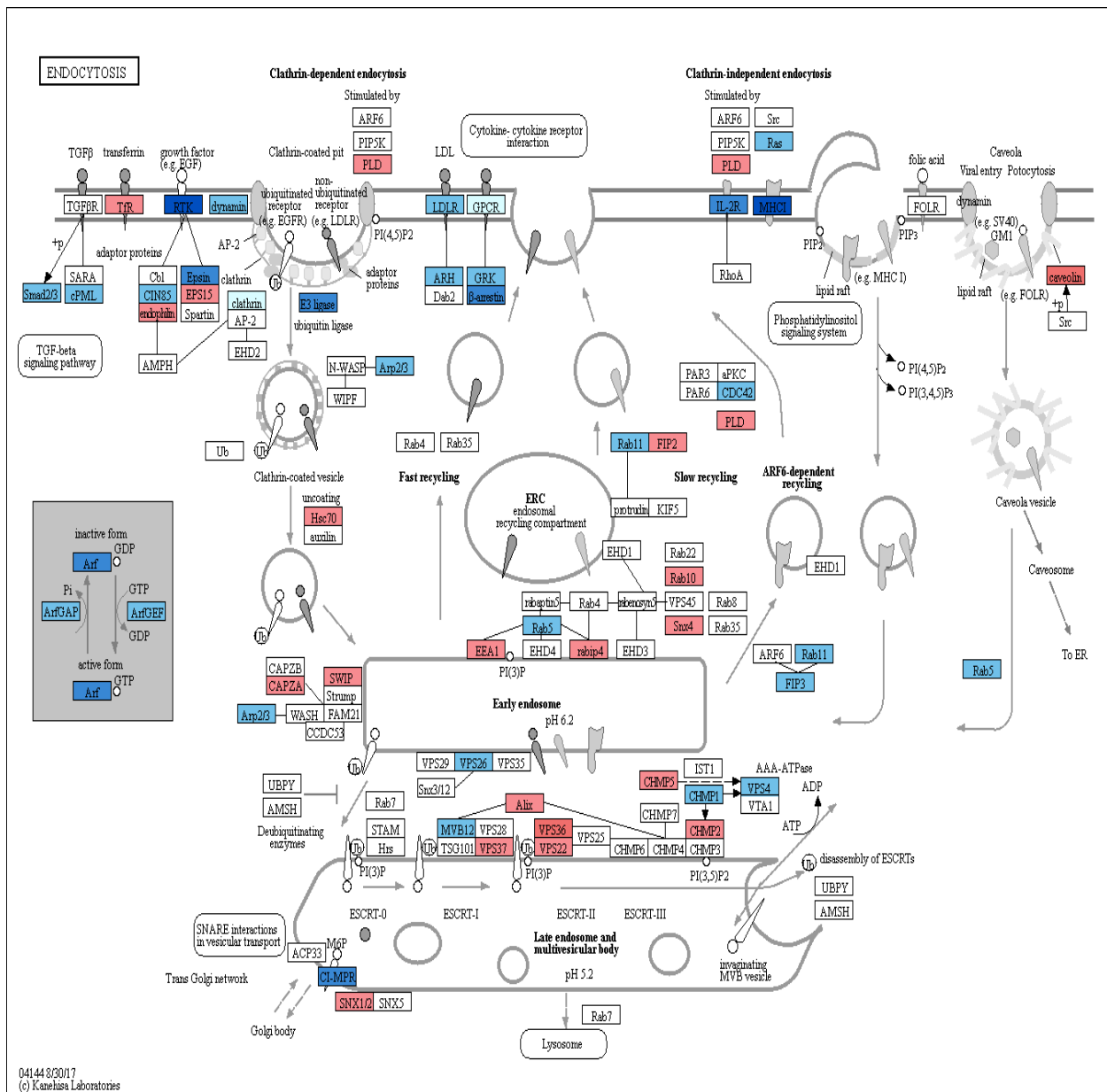


Figura 18 – Endocitose para a série GSE86539 no intervalo de 8 horas. Via da endocitose extraída do KEGG Pathways, com os genes que tiveram alteração dos níveis de expressão. Aumento em escala de vermelho e redução em escala de azul.

Na Figura 199, é apresentado o intervalo entre 0h e 24 horas, onde manteve-se expresso os genes Hops, Caveolin e genes de maturação e formação de endossomo começam a aparecer com expressão elevada, RAB5 e RAB10.

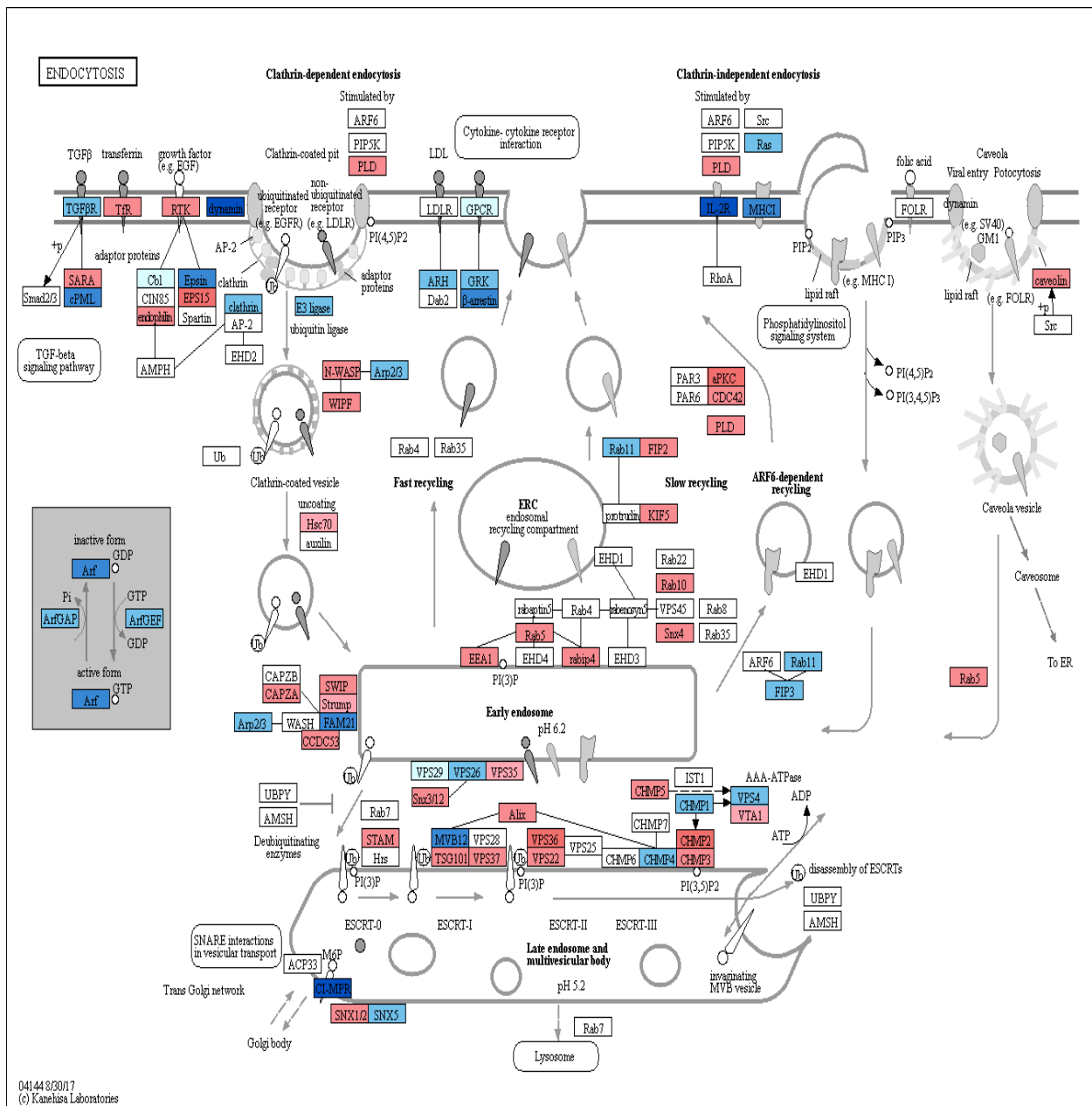


Figura 19 – Endocitose para a série GSE86539 no intervalo de 24 horas. Via da endocitose extraída do KEGG Pathways, com os genes que tiveram alteração dos níveis de expressão. Aumento em escala de vermelho e redução em escala de azul.

Na **Erro! Fonte de referência não encontrada.20**, é apresentado o intervalo de 0h a 48 horas, neste intervalo obteve-se o maior pico de genes com expressão elevada, destaque especial aos genes de membrana, os principais receptores com expressão aumentada, e os genes diretamente ligados a formação e maturação do endossomo, como RAB5, RAB10, RAB7 e complexo HOPS.

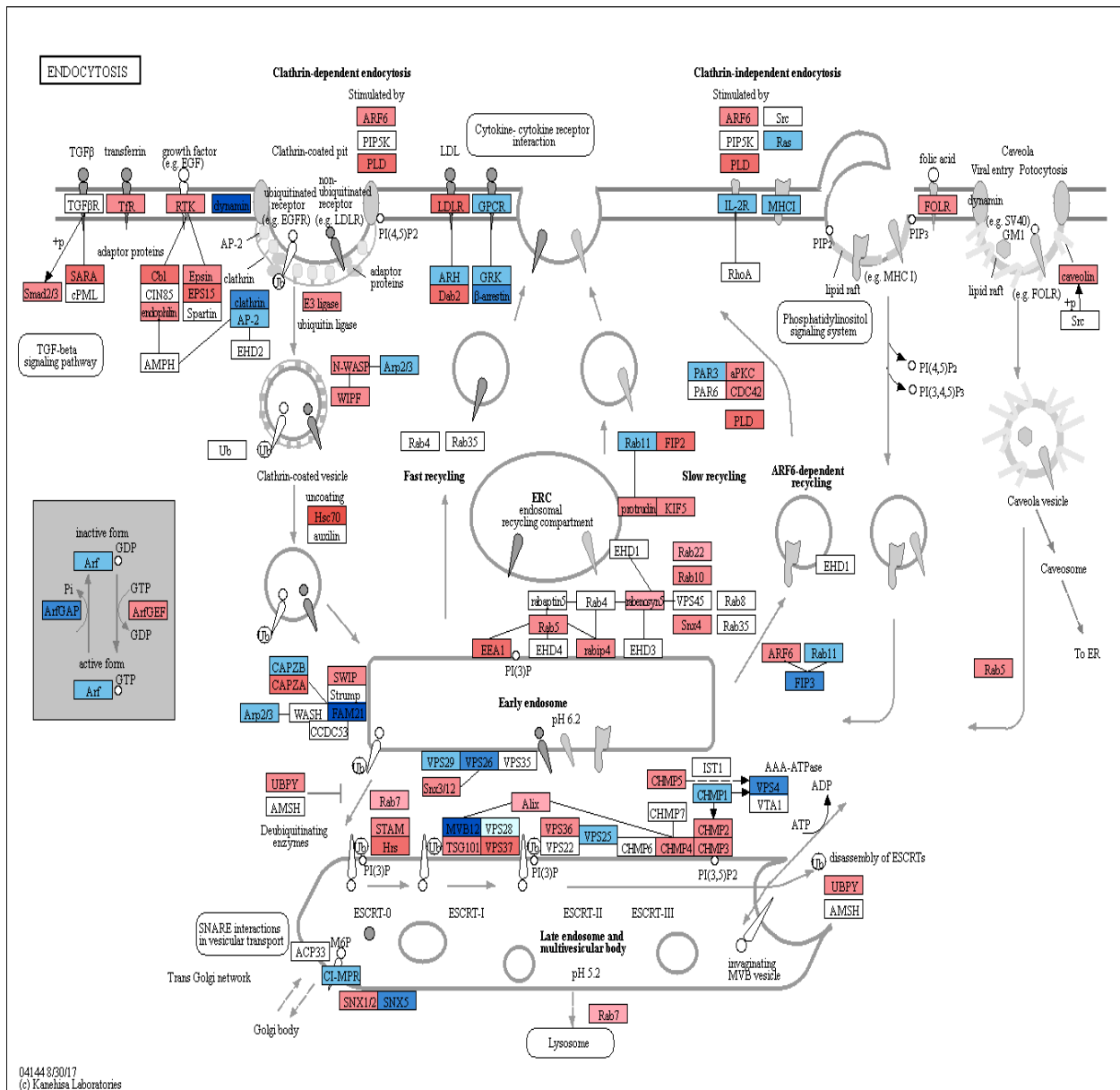


Figura 20 – Endocitose para a série GSE86539 no intervalo de 48 horas. Via da endocitose extraída do KEGG Pathways, com os genes que tiveram alteração dos níveis de expressão. Aumento em escala de vermelho e redução em escala de azul.

Na **Erro! Fonte de referência não encontrada.21**, é apresentado o último intervalo deste experimento 0h a 72 horas, onde os genes relacionados ao endossomo (*early* e *late*) mantiveram sua expressão mais alta, e os genes dos receptores de membrana tiveram uma redução comparados ao intervalo anterior.

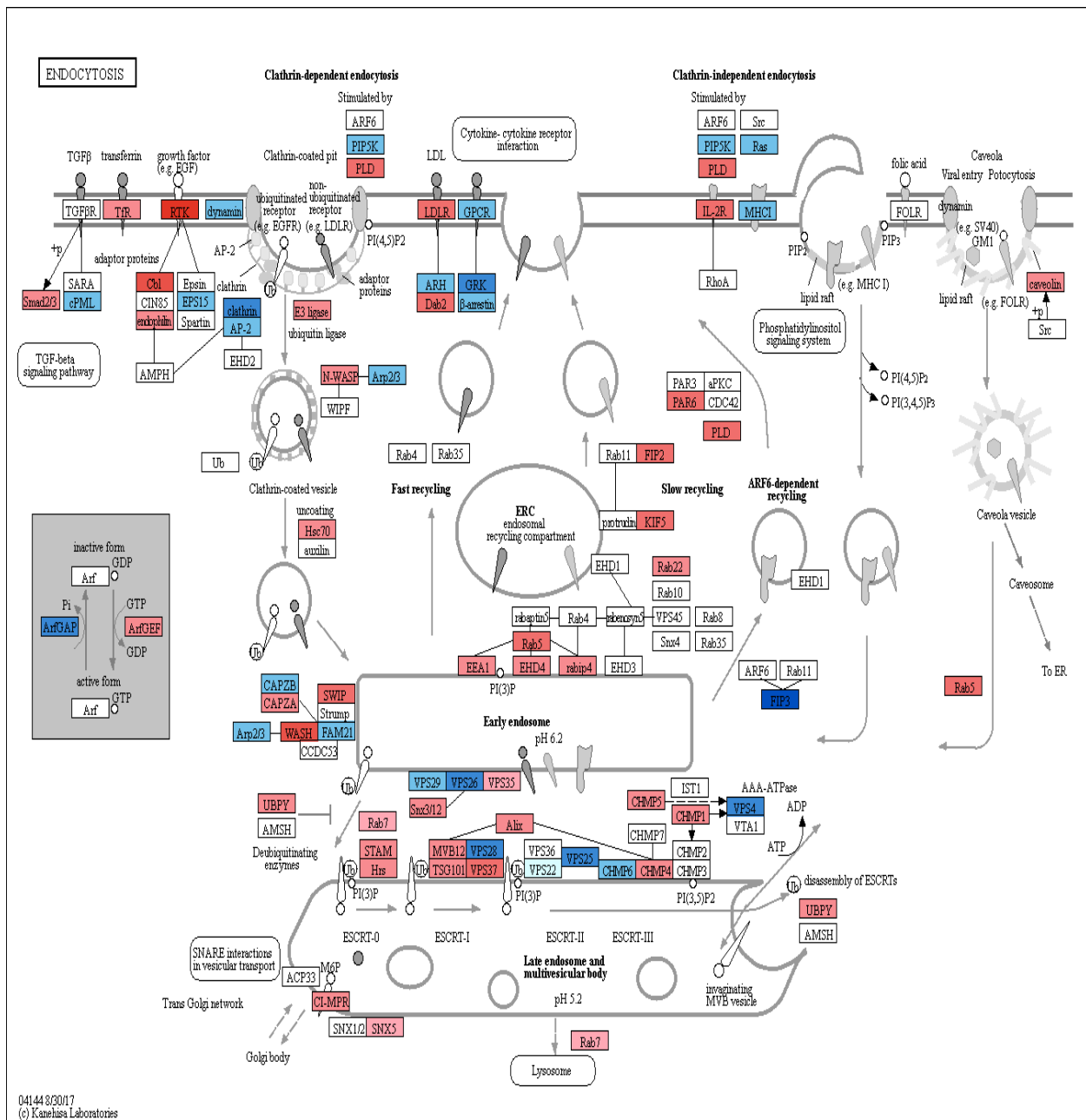


Figura 21 – Endocitose para a série GSE86539 no intervalo de 72 horas. Via da endocitose extraída do KEGG Pathways, com os genes que tiveram alteração dos níveis de expressão. Aumento em escala de vermelho e redução em escala de azul.

5.5. Origem das cepas do vírus Ebola

Com o objetivo de entender a diferença do perfil de infecção de cada uma das cinco cepas do vírus Ebola (Zaire, Sudan, Reston, Bundibugyo e Tai Forest), foi feito o estudo da distância evolutiva entre os produtos gênicos das cepas. Para tanto, foram

comparadas as sete proteínas de cada vírus com o programa ELDOgraph, e com isso gerou-se uma representação gráfica da distância entre as cepas mais similares, possibilitando inferência de proximidade entre as cepas.

Na **Erro! Fonte de referência não encontrada.**²², é apresentado o resultado do ELDOgraph com a legenda relativa à fração das proteínas da cepa a qual está apontada pela flecha que tem como mais próximas as proteínas de uma outra cepa, na origem da flecha. O programa é desenhado para comparar proteomas completos, então neste caso uma cepa pode ter como mais próximas 1, 2, 3, 4, 5, 6 ou 7 de outra cepa, o que em fração se traduz por 0,14, 0,29, 0,43, 0,57, 0,71, 0,86 ou 1. Quando uma proteína tem a mesma distância ou até um pequeno limiar de distância para outras duas, ele considera um empate e conta uma ocorrência de a mais próxima para as outras duas cepas. Assim, a linha vermelha significa acima de 6 a 7 proteínas, verde de 4 a 5, azul 3 e cinza de 1 a 2. Assim, se buscamos as proteínas mais similares a Reston ebolavirus devemos utilizar as proteínas de Sudan ebolavirus (origem da seta vermelha que se origina em Sudan e aponta para Reston). O mesmo para Tai Forest versus Bundibugyo. A cepa do Zaire tem 3 proteínas mais similares em Tai Forest (azul) e 4 a 5 em Bundibugyo já que a seta é verde (no caso, X). As demais comparações têm menos proteínas mais similares, setas em cinza. O conceito do ELDOgraph não é filogenético, mas de biologia comparativa. Mas revela a proximidade funcional entre as cepas das extremidades da Figura e uma proximidade maior da do Zaire com as da direita.

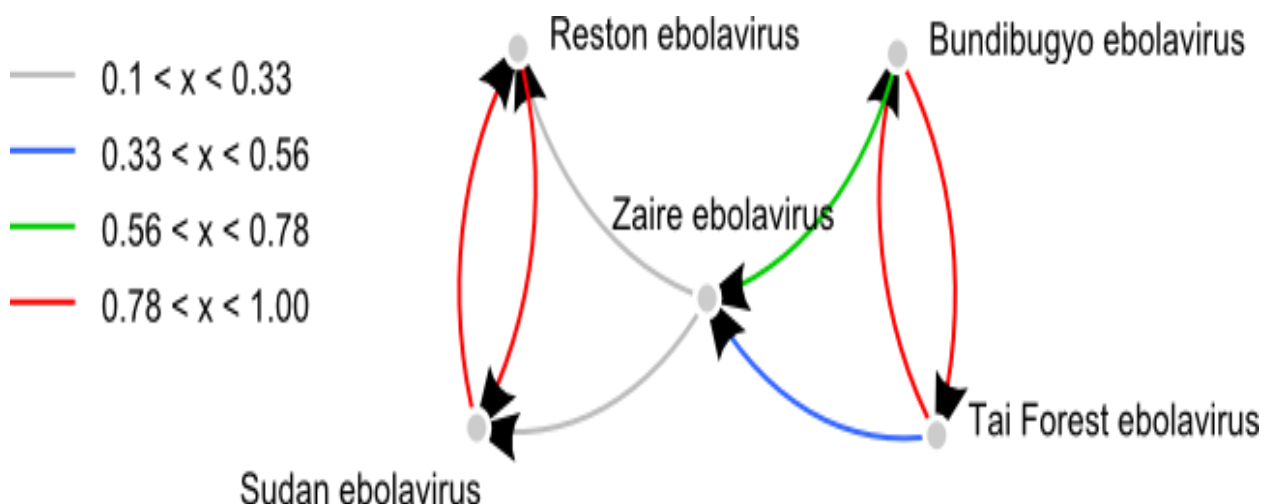


Figura 22 - Resultado do ELDOgraph. Representação em cores relativa as comparações de similaridade entre as proteínas das 5 cepas do vírus Ebola.

5.6. TaxOnTree

Todo o estudo evolutivo realizado neste trabalho foi feito através de softwares que calculam automaticamente a posição mais antiga e por curadoria manual para confirmar nenhuma inconcludente informação.

Alguns genes ao serem estudados apresentaram resultados curiosos em relação a outros genes relativamente próximos evolutivamente, estruturalmente e em função. Dentre estes casos foram selecionados alguns para que fossem apresentados os resultados das árvores filogenéticas e os dados extraídos a partir da informação disponibilizada por ela.

Na **Erro! Fonte de referência não encontrada.**²³, é apresentada a árvore da proteína ICAM 2, que ao ser analisada pelo programa SeedServer, teve seu LCA em Euteleostomi (13). Mas ao se fazer a curadoria manual, foi verificado que a proteína humana estava em um ramo separado, e que a melhor atribuição de nível evolutivo para esta proteína seria Eutheria (20).

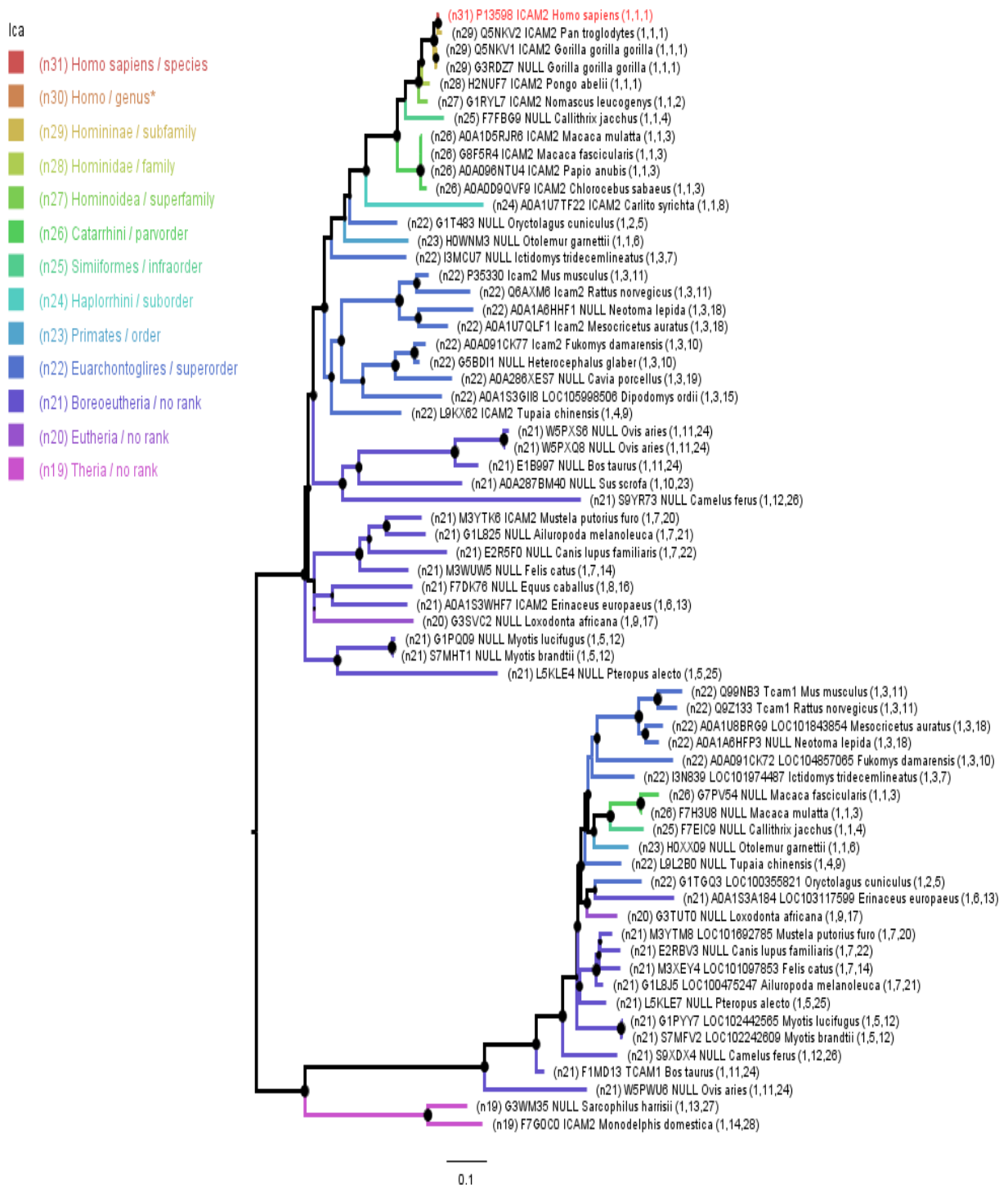


Figura 23 - Arvore Filogenética da ICAM2. Figura elaborada pela TaxOnTree, a proteína ICAM2, objeto de estudo, está destacada em vermelho.

Na **Erro! Fonte de referência não encontrada.24**, se dá a representação da proteína ICAM3. Onde o resultado para o *SeedServer* era dado em Mammalia (18), e

ao fazer a análise pela *TaxOnTree* constatou-se que o nível mais apropriado seria o Boreoeutheria (21). Neste caso um grupo de ortólogos irmão com relações BBH continha a ICAM1 humana.

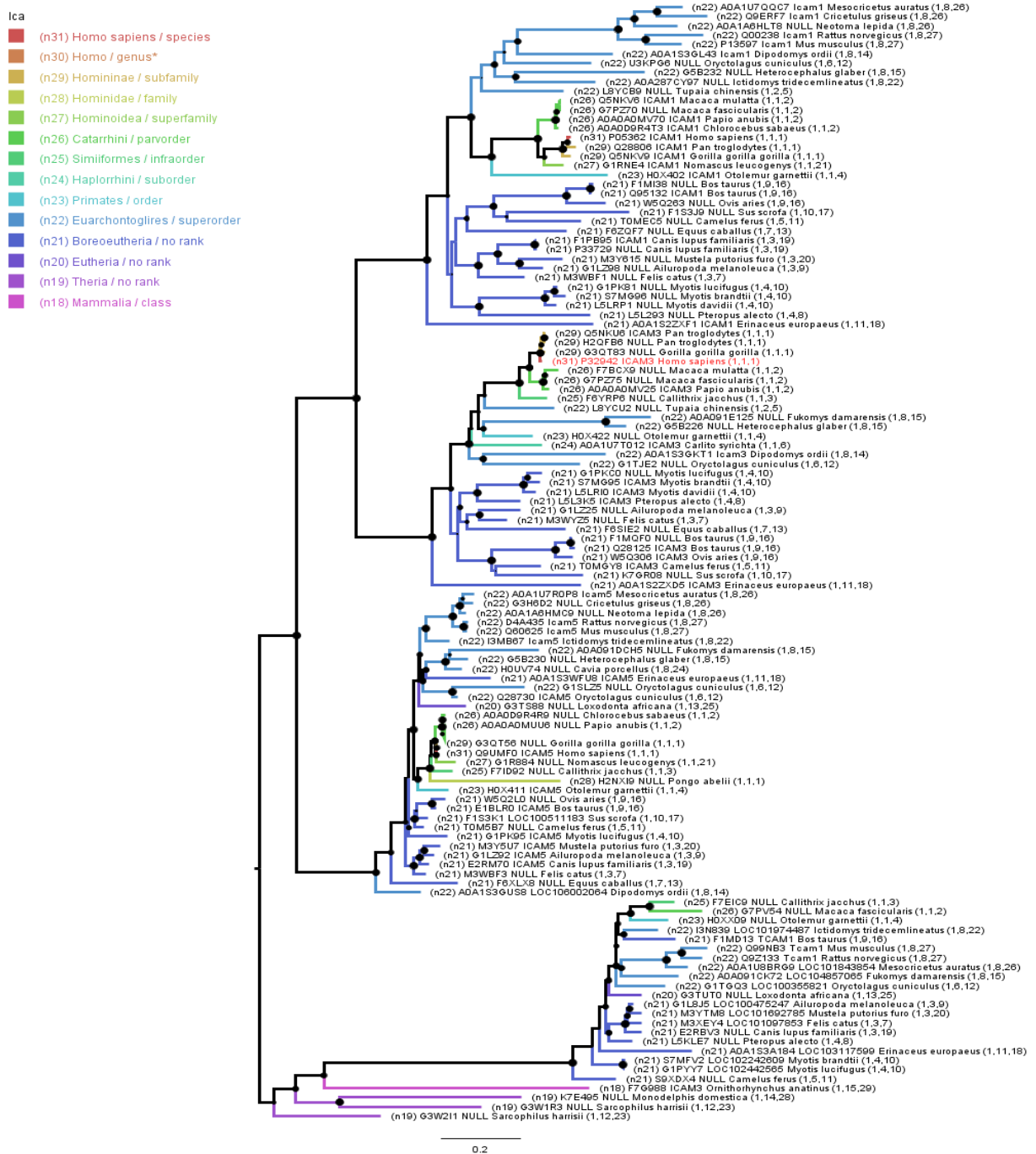


Figura 24 - Arvore Filogenética da ICAM3. Figura elaborada pela *TaxOnTree*, a proteína ICAM3, objeto de estudo, está destacada em vermelho.

6. DISCUSSÃO

Esse trabalho tem como principal contribuição para a comunidade científica, uma via completa de interação do Ebola com proteínas celulares durante a infecção: são 133 genes recrutados por análise da literatura, muitos encontrados de forma não relacionada, que foram todos contextualizados com suas biointerações ao longo do percurso da infecção. Isso representa um cenário que traz possibilidades para novos alvos terapêuticos a serem explorados, novos processos a serem abordados e compreendidos ainda não descritos, relacionados a este vírus. Exploramos também por meio da distribuição taxonômica de ortólogos, a origem cladística das proteínas, pois era plausível que o vírus não estivesse apto a infecções de organismos invertebrados, caso proteínas restritas a vertebrados fossem relevantes. A par disso, o presente trabalho possibilita a visão ao pesquisador interessado no assunto, de quais genes da base Kegg Pathways estariam afetados, os que têm sua expressão aumentada e reduzida, devido a alguma relação com os processos necessários à sobrevivência do vírus em condições internas ao hospedeiro, ou efeitos pleiotrópicos causados pela infecção.

A mineração de texto se mostrou como uma ferramenta de grande importância, cada resumo selecionado contribuiu com uma ou duas biointerações, ajudando a criar o contexto do cenário neste trabalho descrito, por isso ferramentas como MEDLINERANKER e PESCADOR auxiliaram muito, tornando acessível as informações presentes no estado da arte sobre o assunto explorado até o momento. O PESCADOR, em especial, permitiu que traçássemos distintos focos, de modo que foi possível explorar as possibilidades que a infecção pudesse alcançar, como: modulação imune, mimetismo celular, relação direta e indireta, etc.

A literatura geralmente cita símbolos gênicos ou descrições de genes, sem uma curadoria profunda não se chega às sequências de referência que podem ser usadas para pesquisas genômicas. Temos como exemplo os complexos HOPS, MHC, CAVEOLIN, CLATHRIN que são citados de forma generalizada, mas são compostos

por várias proteínas, que como este estudo mostrou não são totalmente dependentes umas das outras, muitas vezes surgiram em épocas diferentes e foram passando por aperfeiçoamento até alcançarem o ponto de especialização descrito pela literatura. Na **Erro! Fonte de referência não encontrada.**5 são apresentados o número de genes da via pelo nível de evolução. É interessante notar que boa parte dos genes surgiu já em Eucariotos, mas as funções e processos mostraram-se completos em níveis mais recentes, como Euteleostomi. O estudo individual de cada um desses componentes permitiu a determinação mais aprofundada das interações, no caso de uma infecção tão versátil, onde cada gene tem um peso importante a ser considerado.

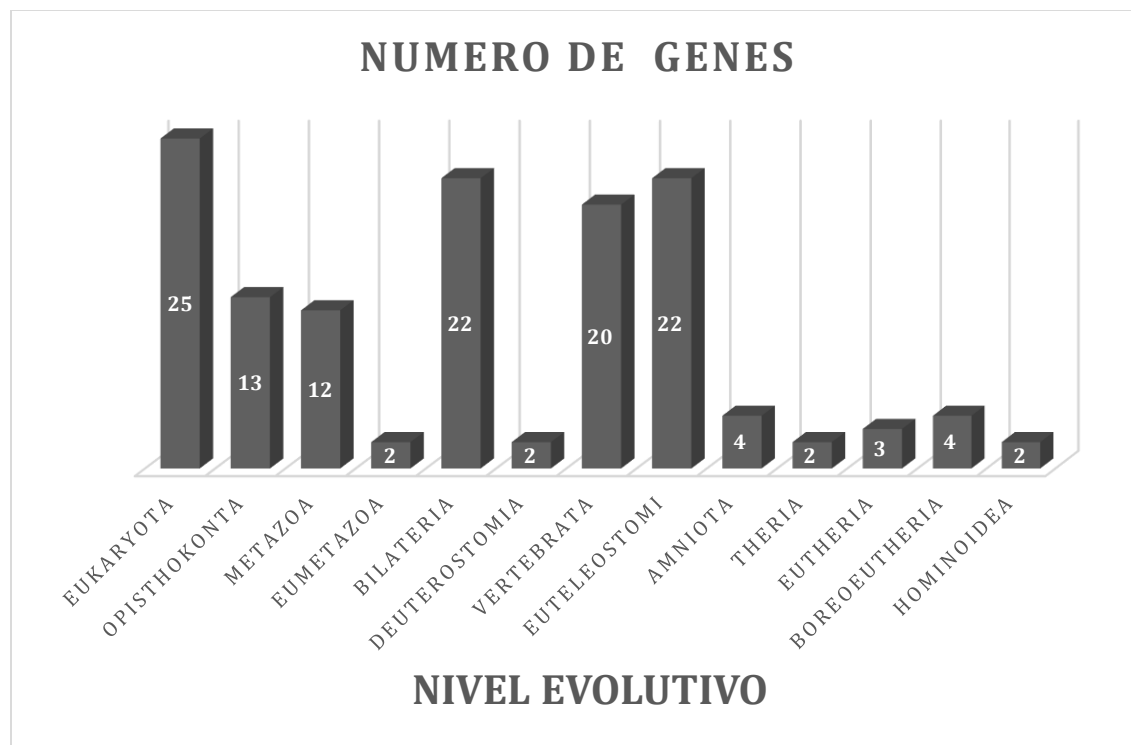


Figura 25 - Número de genes na via por clado de origem. No eixo X estão apresentados os níveis evolutivos e nas barras os números de genes presentes.

O estudo evolutivo de vias tem muitas funcionalidades. Aspectos como, informar a um pesquisador que tenha estudado a expressão diferencial, qual a função de genes que foram encontrados na rota de infecção. Relacionando à via de infecção do vírus, é possível perceber como foram se originando as partes do que se tornou um processo

chave, à infecção viral. As proteínas relacionadas aos processos internos realizados pelo vírus se mostraram mais antigas que as externas, sendo provavelmente recrutadas de processos primordiais à sobrevivência celular, como as catepsinas que estão diretamente ligadas ao aproveitamento energético celular partindo de macro compostos absorvidos pela célula. As regiões que se mostraram mais recentes, como no clado Euteleostomi, estavam ligadas a processos de sinalização e comunicação, provavelmente advindas de comunidades celulares já mais complexas com mais estruturas.

A inferência evolutiva sugere que, hospedeiros ainda não descritos através de experimentos *in vivo* para o vírus Ebola, como seria o caso de peixes, onde o nível evolutivo Euteleostomi tem proteínas suficientes para a multiplicação viral na célula, seriam susceptíveis a serem hospedeiros para o vírus. Com a análise de associação de ortólogos há também a concordância com trabalhos recentes realizados após a epidemia de 2014 do vírus Ebola, onde *pets* foram descritos como hospedeiros, e seriam do grupo Boreoeutheria (mamíferos, com exceção de Afrotheria: elefantes por exemplo) (HAMBESSE *et al.*, 2016; HAN *et al.*, 2016).

O nível Euteleostomi foi selecionado a partir das análises desempenhadas neste trabalho, como ponto chave para que a infecção pudesse ocorrer. Foram utilizados como critério os processos que a partir deste ponto de evolução já estão presentes na via, como a via de entrada pela família TAM, a conexão através do ligante de manose, receptor de ácido folato e os complexos estruturais Caveolina e Clathrina. Apesar do processo de formação da via de endocitose por citoesqueleto não estar completo estes receptores presentes, segundo a literatura, seriam possivelmente suficientes para que o vírus conseguisse realizar a adsorção a célula e possivelmente a penetração. Ainda neste nível encontram-se presentes, as catepsinas e os transportadores de colesterol (CTSL, CTSB, NPC1 e NPC2) essenciais para a clivagem e fusão da GP, para o desnudamento do vírus no citoplasma. NPC1 ganha destaque especial por ser encarregada pela maioria dos trabalhos relacionados ao desenvolvimento de tolerância à infecção pelo vírus Ebola. No trabalho feito por Carette *et al.*, 2011, foi constatado que, uma mutação nesta proteína é suficiente para impedir o desenvolvimento da infecção.

Após o nível Eutelostomi até o Hominoidea, o mais recente, surgem novas proteínas na via, receptores que são quase todos lectinas, CD209, CLEC4M, CLEC10A, CLEC4G e CLEC6A. Apesar de novos, tem como característica mediar a adsorção do vírus sem precisar de um intermediário. Segundo o estudo por este trabalho realizado, estes surgem na via em organismos mais recentes, a partir de Amniota.

A determinação do clado de origem não é algo trivial, porque formar o grupo de ortólogos é uma atividade difícil, as várias bases de dados podem concordar, ou por vezes divergirem. No laboratório de Biodados foram utilizadas informações de LCA de diferentes bancos, usados através do *software* SeedServer e consulta automática à base GeneAges por meio do ARCoBALeno. Para a maior confiabilidade dos resultados, optou-se pela determinação manual do LCA. A conferência manual permite a visualização das relações evolutivas. Todavia nos dá a situação atual no homem, onde proteínas de grupos de ortólogos separados, por exemplo, não tendo relação de BBH (*Bidirectional Best Hit*) com o ramo filogenético vizinho, poderiam ser consideradas mais ancestrais se considerarmos que o ancestral comum dos dois ramos já executaria a função, embora atualmente a função tenha sido caracterizada em uma delas somente.

A análise de expressão diferencial trouxe dados sobre vias e processos desencadeados pelo vírus, não descritos na literatura. A similaridade de sua infecção à alguns pontos da infecção por HTLV (dados no site: biodados.icb.ufmg.br/ebola), é um interessante aspecto a ser considerado, uma vez que são vírus de características e famílias diferentes. Esta abordagem abre portas para uma comparação sistemática de infecções e processos não só virais como de outros patógenos.

A via *Pathways in cancer* (dados no site: biodados.icb.ufmg.br/ebola), trouxe processos que abordam o efeito da imortalidade celular, uma vez infectada a célula tende a sinalizar o agente estranho e/ou entrar em apoptose, mas o vírus precisa se replicar, então vias de sinalização relacionadas a morte celular são inibidas. Como vimos, esta fase do trabalho acabou por apresentar muito mais material que seja plausível um grupo analisar, para tanto será feito deste um projeto piloto, para a criação de um banco de dados com interface interativa para que possa atender e receber

contribuições da comunidade científica.

A via da Endocitose merece destaque neste trabalho por ser o mecanismo utilizado para a penetração viral, a análise de duas células distintas como o hepatócito e o monócito mostraram como os genes são modulados para funcionarem a favor do vírus de maneiras diferentes. O monócito representa um alvo primário, é uma das primeiras células a terem contato com agentes infecciosos, diferente do hepatócito alvo acometido pelo ebola dias após estar presente no hospedeiro. A maior atividade de expressão celular dos genes da endocitose no monócito foi no início do contato com o vírus, no intervalo de 0 a 8 horas. Especula-se que por ser uma célula de defesa, é feita a tentativa de fagocitose, mas devido ao envoltório proteico o vírus sobrevive. No hepatócito, alvo já mais tardio à infecção, a variação de expressão teve seu ápice no intervalo entre 48h e 72h, devido à diferença no modelo celular e como o processo precisou ativar muito mais genes para que ocorresse para que ocorresse a internalização do vírus.

Apesar da simplicidade do organismo estudado, com apenas sete proteínas e um RNA, o vírus Ebola mostrou-se muito dinâmico, pela capacidade de infecção, sobrevivência e virulência em todos os seus hospedeiros. A análise feita a partir do ELDOgraph permite visualizar a relação entre as cepas, as patogênicas ao homem ou não. É interessante notar que as cepas que mais se aproximaram geograficamente, são as mais próximas em similaridades. É possível que a relação da cepa Sudan, patogênica ao homem, e Reston, não patogênica, que apresentaram similaridade entre 6-7 das proteínas analisadas, seja justificada com alguma mutação na infecção a outra espécie e necessidade de adaptação para a sobrevivência viral. Tai Forest e Bundibugyo também se mostraram com similaridade equivalente, ambos apresentaram proximidade geográfica em seu primeiro relato, possivelmente tendo algum ancestral em comum. Já a cepa Zaire foi a que se mostrou mais distante, tendo como mais próxima a cepa Tai Forest, mas com apenas 3 de 7 proteínas como as mais semelhantes (ELDO significa ortólogo menos divergente no grupo externo, no caso as diferentes cepas). Esta cepa tem características únicas de patogenicidade, possivelmente o que justifica esta diversidade. Alguns trabalhos descrevem que há mutações no vírus ao infectar outra espécie (DIETZEL *et al.*, 2017; HOFFMANN *et al.*,

2017). O que levanta hipóteses sobre um ancestral comum para as cinco cepas, que ao infectar hospedeiros diferentes foi sofrendo mutações e dando origem a outras.

Com a via de interação do vírus Ebola e o hospedeiro, e a origem dos processos envolvidos espera-se trazer mais respostas sobre como o vírus infecta tantos hospedeiros, como seus sintomas são desencadeados a nível molecular, quais os mecanismos são usurpados e quais seriam possíveis meios de evitar uma nova epidemia. A análise apresentada é binária, todavia pode inspirar estudos sistemáticos, com enfoque alterações de expressão, para entendimento do universo de infecção viral.

7. REFERENCIAS

ALEKSANDROWICZ, Paulina *et al.* Ebola virus enters host cells by macropinocytosis and clathrin-mediated endocytosis. *Journal of Infectious Diseases*, v. 204, n. SUPPL. 3, p. 957–967, 2011.

ALTSCHUL, S F *et al.* Basic local alignment search tool. *Journal of molecular biology*, v. 215, n. 3, p. 403–10, 1990. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022283605803602>>.

ALVAREZ, Carmen P *et al.* C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans. *Journal of Virology*, v. 76, n. 13, p. 6841–6844, 2002.

AUDET, Jonathan; KOBINGER, Gary P. Immune Evasion in Ebolavirus Infections. *Viral Immunology*, v. 28, n. 1, p. 10–18, 2015. Disponível em: <<http://online.liebertpub.com/doi/abs/10.1089/vim.2014.0066>>.

AZARIAN, Taj *et al.* Impact of spatial dispersion, evolution, and selection on Ebola Zaire Virus epidemic waves. *Scientific reports*, v. 5, p. 10170, 2015. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4431419&tool=pmcentrez&rendertype=abstract>>.

BASLER, Christopher F.; AMARASINGHE, Gaya K. Evasion of Interferon Responses by Ebola and Marburg Viruses. *Journal of Interferon & Cytokine Research*, v. 29, n. 9, p. 511–520, 2009. Disponível em: <<http://www.liebertonline.com/doi/abs/10.1089/jir.2009.0076>>.

BAUSCH, Daniel G.; SCHWARZ, Lara. Outbreak of Ebola Virus Disease in Guinea: Where Ecology Meets Economy. *PLoS Neglected Tropical Diseases*, v. 8, n. 7, p. 8–12, 2014.

BHATTACHARYYA, Suchita; HOPE, Thomas J.; YOUNG, John a T. Differential requirements for clathrin endocytic pathway components in cellular entry by Ebola and Marburg glycoprotein pseudovirions. *Virology*, v. 419, n. 1, p. 1–9, 2011. Disponível em: <<http://dx.doi.org/10.1016/j.virol.2011.07.018>>.

BHATTACHARYYA, Suchita; MULHERKAR, Nirupama; CHANDRAN, Kartik. Endocytic pathways involved in filovirus entry: Advances, implications and future directions. *Viruses*, v. 4, n. 12, p. 3647–3664, 2012.

BIEDENKOPF, Nadine; LIER, Clemens; BECKER, Stephan. Dynamic Phosphorylation of VP30 Is Essential for Ebola Virus Life Cycle. *Journal of Virology*, v. 90, n. 10, p. 4914–4925, 2016. Disponível em: <<http://jvi.asm.org/lookup/doi/10.1128/JVI.03257-15>>.

BRECHER, M *et al.* Cathepsin cleavage potentiates the Ebola virus glycoprotein to undergo a subsequent fusion-relevant conformational change. *J Virol*, v. 86, n. 1, p. 364–372, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22031933>%5Cn<http://jvi.asm.org/content/86/1/364.full.pdf>>.

BRINKMANN, Constantin *et al.* Tetherin antagonism by the Ebola virus glycoprotein requires an intact receptor-binding domain and can be blocked by GP1-specific antibodies. *Journal of Virology*, v. 90, n. October, p. JVI.01563-16, 2016. Disponível em: <<http://jvi.asm.org/lookup/doi/10.1128/JVI.01563-16>>.

BRUDNER, Matthew *et al.* Lectin-Dependent Enhancement of Ebola Virus Infection via Soluble and Transmembrane C-type Lectin Receptors. *PLoS ONE*, v. 8, n. 4, 2013.

C??T??, Marceline *et al.* Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. *Nature*, v. 477, n. 7364, p. 344–348, 2011.

CAPELLA-GUTIÉRREZ, Salvador; SILLA-MARTÍNEZ, José M.; GABALDÓN, Toni. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, v. 25, n. 15, p. 1972–1973, 2009.

CARDENAS, W. B. *et al.* Ebola Virus VP35 Protein Binds Double-Stranded RNA and Inhibits Alpha/Beta Interferon Production Induced by RIG-I Signaling. *Journal of Virology*, v. 80, n. 11, p. 5168–5178, 2006. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.02199-05>>.

CARETTE, Jan E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature*, v. 477, n. 7364, p. 340–343, 2011. Disponível em: <<http://dx.doi.org/10.1038/nature10348>>.

CHAN, Stephen Y. *et al.* Folate receptor-?? is a cofactor for cellular entry by Marburg and Ebola viruses. *Cell*, v. 106, n. 1, p. 117–126, 2001.

CHEN, Liqing *et al.* High resolution crystal structure of human Rab9 GTPase: a novel antiviral drug target. *The Journal of biological chemistry*, v. 279, n. 38, p. 40204–8, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15263003>>.

DENNER, Joachim. Treatment of Ebola virus infections with inhibitors of TLR4. *Medical Hypotheses*, v. 85, n. 3, p. 253–257, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.mehy.2015.05.009>>.

DIETZEL, Erik *et al.* Functional Characterization of Adaptive Mutations during the West African Ebola Virus Outbreak. *Journal of virology*, v. 91, n. 2, p. e01913-16, 2017. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/27847361>%5Cn<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5215343>>.

DOLNIK, Olga *et al.* Ectodomain shedding of the glycoprotein GP of Ebola virus. *EMBO*

Journal, v. 23, n. 10, p. 2175–2184, 2004.

DOMINGUEZ-SOTO, Angeles *et al.* The DC-SIGN – related lectin LSECTin mediates antigen capture and pathogen binding by human myeloid cells The DC-SIGN – related lectin LSECTin mediates antigen capture and pathogen binding by human myeloid cells. v. 109, n. 12, p. 5337–5345, 2008.

DOMÍNGUEZ-SOTO, Ángeles *et al.* The pathogen receptor liver and lymph node sinusoidal endothelial cell C-type lectin is expressed in human Kupffer cells and regulated by PU.1. *Hepatology*, v. 49, n. 1, p. 287–296, 2009.

DOUGLAS, Janet L. *et al.* The great escape: Viral strategies to counter BST-2/ tetherin. *PLoS Pathogens*, v. 6, n. 5, p. 1–12, 2010.

DURINCK, Steffen *et al.* BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*, v. 21, n. 16, p. 3439–3440, 2005.

DURINCK, Steffen *et al.* Mapping identifiers for integration of genomic datasets with the R/Bioconductor package biomaRt. v. 4, n. 8, p. 1184–1191, 2009.

EDGAR, Robert C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, v. 32, n. 5, p. 1792–1797, 2004.

EMPIG C.J. GOLDSMITH M.A.: Empig C.J. Goldsmith M.A.: Association of the caveola vesicular system with cellular entry by filoviruses. *Journal of Virology*, 2002; 76(10):5266- 70. *Journal of Virology*, v. 76, n. 10, p. 5266–70, 2002.

FALASCA, L. *et al.* Molecular mechanisms of Ebola virus pathogenesis: Focus on cell death. *Cell Death and Differentiation*, v. 22, n. 8, p. 1250–1259, 2015.

FELDMAN, Heinz; GEISBERT, Thomas. Ebola. *Lancet*, v. 377, n. 9768, p. 849–862, 2012.

FENG, Z. *et al.* The VP35 Protein of Ebola Virus Inhibits the Antiviral Effect Mediated by Double-Stranded RNA-Dependent Protein Kinase PKR. *Journal of Virology*, v. 81, n. 1, p. 182–192, 2007. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.01006-06>>.

FRIEDEL, Caroline C.; HAAS, Jürgen. Virus-host interactomes and global models of virus-infected cells. *Trends in Microbiology*, v. 19, n. 10, p. 501–508, 2011.

GARRETT, Wendy S *et al.* Developmental Control of Endocytosis in Dendritic Cells by Cdc42. *Cell*, v. 102, n. 3, p. 325–334, 2000. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0092867400000386>>.

GENTLEMAN, RC *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, v. 5, n. 10, p. R80, 2004.

GOEIJENBIER, M *et al.* Ebola virus disease: a review on epidemiology, symptoms, treatment and pathogenesis. *The Netherlands journal of medicine*, v. 72, n. 9, p. 442–8, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25387613>>.

GOFFINET, C. *et al.* Endogenous CD317/Tetherin Limits Replication of HIV-1 and Murine Leukemia Virus in Rodent Cells and Is Resistant to Antagonists from Primate Viruses. *Journal of Virology*, v. 84, n. 21, p. 11374–11384, 2010. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.01067-10>>.

GONG, Xin *et al.* Structural insights into the Niemann-Pick C1 (NPC1)-mediated cholesterol transfer and ebola infection. *Cell*, v. 165, n. 6, p. 1467–1478, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.cell.2016.05.022>>.

GRAMBERG, Thomas *et al.* Interactions of LSEctin and DC-SIGN/DC-SIGNR with viral ligands: Differential pH dependence, internalization and virion binding. *Virology*, v. 373, n. 1, p. 189–201, 2008.

GRAMBERG, Thomas *et al.* LSEctin interacts with filovirus glycoproteins and the spike protein of SARS coronavirus. *Virology*, v. 340, n. 2, p. 224–236, 2005.

GRIFONI, Alba *et al.* Genetic diversity in Ebola virus: Phylogenetic and in silico structural studies of Ebola viral proteins. *Asian Pacific Journal of Tropical Medicine*, v. 9, n. 4, p. 337–343, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.apjtm.2016.03.016>>.

GROVE, Joe; MARSH, Mark. The cell biology of receptor-mediated virus entry. *Journal of Cell Biology*, v. 195, n. 7, p. 1071–1082, 2011.

HAMBESE, Hailu *et al.* Ebola Virus Disease in Domestic and Wild Animals: A Review. *Journal of Pharmacy and Alternative Medicine ISSN*, v. 10, p. 54–63, 2016.

HAN, Ziyang *et al.* Ebola virus mediated infectivity is restricted in canine and feline cells. *Veterinary Microbiology*, v. 182, p. 102–107, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.vetmic.2015.11.011>>.

HARTMAN, A. L. *et al.* Reverse Genetic Generation of Recombinant Zaire Ebola Viruses Containing Disrupted IRF-3 Inhibitory Domains Results in Attenuated Virus Growth In Vitro and Higher Levels of IRF-3 Activation without Inhibiting Viral Transcription or Replication. *Journal of Virology*, v. 80, n. 13, p. 6430–6440, 2006. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.00044-06>>.

HARTMAN, A. L. *et al.* Whole-Genome Expression Profiling Reveals That Inhibition of Host Innate Immune Response Pathways by Ebola Virus Can Be Reversed by a Single Amino Acid Change in the VP35 Protein. *Journal of Virology*, v. 82, n. 11, p. 5348–5358, 2008. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.00215-08>>.

HOENEN, Thomas *et al.* Ebola virus: unravelling pathogenesis to combat a deadly disease. *Trends in Molecular Medicine*, v. 12, n. 5, p. 206–215, 2006.

HOFFMANN, Markus *et al.* A Polymorphism within the Internal Fusion Loop of the Ebola Virus Glycoprotein Modulates Host Cell Entry. *Journal of Virology*, v. 91, n. 9, p. e00177-17, 2017. Disponível em: <<http://jvi.asm.org/lookup/doi/10.1128/JVI.00177-17>>.

HUBER, Wolfgang *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, v. 12, n. 2, p. 115–121, 2015. Disponível em: <<http://dx.doi.org/10.1038/nmeth.3252>>.

HUNT, Catherine L *et al.* The Tyro3 receptor kinase Axl enhances macropinocytosis of Zaire ebolavirus. *Journal of virology*, v. 85, n. 1, p. 334–347, 2011.

ILINYKH, Philipp A. *et al.* Role of protein phosphatase 1 in dephosphorylation of Ebola virus VP30 protein and its targeting for the inhibition of viral transcription. *Journal of Biological Chemistry*, v. 289, n. 33, p. 22723–22738, 2014.

JEMIELITY, Stephanie *et al.* TIM-family Proteins Promote Infection of Multiple Enveloped Viruses through Virion-associated Phosphatidylserine. *PLoS Pathogens*, v. 9, n. 3, 2013.

Jl, Xin *et al.* Mannose-binding lectin binds to Ebola and Marburg envelope glycoproteins, resulting in blocking of virus interaction with DC-SIGN and complement-mediated virus neutralization. *Journal of General Virology*, v. 86, n. 9, p. 2535–2542, 2005.

KALETSKY, R. L. *et al.* Tetherin-mediated restriction of filovirus budding is antagonized by the Ebola glycoprotein. *Proceedings of the National Academy of Sciences*, v. 106, n. 8, p. 2886–2891, 2009. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.0811014106>>.

KATZE, Michael G. *et al.* Innate immune modulation by RNA viruses: emerging insights from functional genomics. *Nature Reviews Immunology*, v. 8, n. 8, p. 644–654, 2008. Disponível em: <<http://www.nature.com/doi/10.1038/nri2377>>.

KIMBERLIN, C. R. *et al.* Ebolavirus VP35 uses a bimodal strategy to bind dsRNA for innate immune suppression. *Proceedings of the National Academy of Sciences*, v. 107, n. 1, p. 314–319, 2010. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.0910547107>>.

KONDRATOWICZ, Andrew S *et al.* T-cell immunoglobulin and mucin domain 1 (TIM-1) is a receptor for Zaire Ebolavirus and Lake Victoria Marburgvirus. *Proceedings of the National Academy of Sciences of the United States of America*, v. 108, n. 20, p. 8426–31, 2011. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3100998&tool=pmcentrez&endertype=abstract>>.

KORTEPETER, Mark G.; BAUSCH, Daniel G.; BRAY, Mike. Basic clinical and Laboratory features of filoviral hemorrhagic fever. *Journal of Infectious Diseases*, v. 204, n. SUPPL. 3, 2011.

KURODA, Makoto *et al.* Interaction between TIM-1 and NPC1 Is Important for Cellular Entry of Ebola Virus. *Journal of virology*, v. 89, n. 12, p. 6481–93, 2015. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4474285&tool=pmcentrez&endertype=abstract>>.

KUTMON, Martina *et al.* PathVisio 3: An Extendable Pathway Analysis Toolbox. *PLoS Computational Biology*, v. 11, n. 2, p. 1–13, 2015.

LAWRENCE, Philip *et al.* Human transmission of Ebola virus. *Current Opinion in Virology*, v. 22, p. 51–58, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.coviro.2016.11.013>>.

LEE, Michael S.; LEBEDA, Frank J.; OLSON, Mark A. Fold prediction of VP24 protein of Ebola and Marburg viruses using de novo fragment assembly. *Journal of Structural Biology*, v. 167, n. 2, p. 136–144, 2009. Disponível em: <<http://dx.doi.org/10.1016/j.jsb.2009.05.001>>.

LEENDERTZ, Siv Aina Jensen. Testing new hypotheses regarding ebolavirus reservoirs. *Viruses*, v. 8, n. 2, 2016.

LEROY, Eric M. *et al.* Fruit bats as reservoirs of Ebola virus. *Nature*, v. 438, n. 7068, p. 575–576, 2005.

LEUNG, Lawrence W *et al.* Ebolavirus VP35 suppresses IFN production from conventional but not plasmacytoid dendritic cells. *Immunology and cell biology*, v. 89, n. 7, p. 792–802, 2011. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4148147&tool=pmcentrez&endertype=abstract>>.

LI, M. *et al.* TIM-family proteins inhibit HIV-1 release. *Proceedings of the National Academy of Sciences*, v. 111, n. 35, p. E3699–E3707, 2014. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.1404851111>>.

LI, Xingguang *et al.* The 2014 Ebola virus outbreak in West Africa highlights no evidence of rapid evolution or adaptation to humans. *Scientific reports*, v. 6, n. October, p. 35822, 2016. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/27767073>> <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5073338>>.

LIU, Si Qing *et al.* Identifying the pattern of molecular evolution for Zaire ebolavirus in the 2014 outbreak in West Africa. *Infection, Genetics and Evolution*, v. 32, p. 51–59, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.meegid.2015.02.024>>.

LOVE, Michael I.; HUBER, Wolfgang; ANDERS, Simon. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, v. 15, n. 12, p. 1–21, 2014.

LUTHRA, Priya *et al.* Mutual antagonism between the ebola virus VP35 protein and the

RIG-I activator PACT determines infection outcome. *Cell Host and Microbe*, v. 14, n. 1, p. 74–84, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.chom.2013.06.010>>.

MARTIN, Baptiste; CANARD, Bruno; DECROLY, Etienne. Filovirus proteins for antiviral drug discovery: Structure/function bases of the replication cycle. *Antiviral Research*, v. 141, p. 48–61, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.antiviral.2017.02.004>>.

MARTINEZ, O. *et al.* A Mutation in the Ebola Virus Envelope Glycoprotein Restricts Viral Entry in a Host Species- and Cell-Type-Specific Manner. *Journal of Virology*, v. 87, n. 6, p. 3324–3334, 2013. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.01598-12>>.

MARTINEZ, Osvaldo; VALMAS, Charalampos; BASLER, Christopher F. Ebola virus-like particle-induced activation of NF- κ B and Erk signaling in human dendritic cells requires the glycoprotein mucin domain. *Virology*, v. 364, n. 2, p. 342–354, 2007.

MASON, Christopher P.; TARR, Alexer W. Human lectins and their roles in viral infections. *Molecules*, v. 20, n. 2, p. 2229–2271, 2015.

MEHEDI, M. *et al.* A New Ebola Virus Nonstructural Glycoprotein Expressed through RNA Editing. *Journal of Virology*, v. 85, n. 11, p. 5406–5414, 2011. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.02190-10>>.

MELANSON, Vanessa R; KALINA, Warren V; WILLIAMS, Priscilla. Ebola Virus Infection Induces Irregular Dendritic Cell Gene Expression. *Viral immunology*, v. 28, n. 1, p. 1–9, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25493356>>.

MERCER, Jason; GREBER, Urs F. Virus interactions with endocytic pathways in macrophages and dendritic cells. *Trends in Microbiology*, v. 21, n. 8, p. 380–388, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.tim.2013.06.001>>.

METSCHNIKOFF, Elias. Lecture on phagocytosis and immunity. *British Medical Journal*, v. 1, n. 1570, p. 213–217, 1891.

MIAO, Chunhui *et al.* Cell-cell contact promotes Ebola virus GP-mediated infection. *Virology*, v. 488, p. 202–215, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.virol.2015.11.019>>.

MILLER, Emily Happy *et al.* Ebola virus entry requires the host-programmed recognition of an intracellular receptor. *EMBO Journal*, v. 31, n. 8, p. 1947–1960, 2012. Disponível em: <<http://dx.doi.org/10.1038/emboj.2012.53>>.

MISASI, John *et al.* Filoviruses require endosomal cysteine proteases for entry but exhibit distinct protease preferences. *Journal of virology*, v. 86, n. 6, p. 3284–92, 2012. Disponível em: <<http://jvi.asm.org/content/86/6/3284.full>>.

MOHAN, Gopi S. *et al.* Less Is More: Ebola Virus Surface Glycoprotein Expression Levels Regulate Virus Production and Infectivity. *Journal of Virology*, v. 89, n. 2, p.

1205–1217, 2015. Disponível em: <<http://jvi.asm.org/lookup/doi/10.1128/JVI.01810-14>>.

MOLLER-TANK, Sven *et al.* Role of the phosphatidylserine receptor TIM-1 in enveloped-virus entry. *Journal of virology*, v. 87, n. 15, p. 8327–41, 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3719829&tool=pmcentrez&rendertype=abstract>>.

MOLLER-TANK, Sven; MAURY, Wendy. Phosphatidylserine receptors: Enhancers of enveloped virus entry and infection. *Virology*, v. 468, p. 565–580, 2014. Disponível em: <<http://dx.doi.org/10.1016/j.virol.2014.09.009>>.

MORIZONO, K; CHEN, I S. Role of phosphatidylserine receptors in enveloped virus infection. *J Virol*, v. 88, n. 8, p. 4275–4290, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3993771/pdf/zjv4275.pdf>>.

MORIZONO, Kouki *et al.* Mediate Viral Entry. v. 9, n. 4, p. 286–298, 2012.

MURRAY, James L *et al.* Rab9 GTPase Is Required for Replication of Human Immunodeficiency Virus Type 1 , Filoviruses , and Measles Virus. *Society*, v. 79, n. i, p. 11742–11751, 2005.

NDUNGO, Esther *et al.* A Single Residue in Ebola Virus Receptor NPC1 Influences Cellular Host Range in Reptiles. *mSphere*, v. 1, n. 2, p. 1–15, 2016. Disponível em: <<http://msphere.asm.org/content/1/2/e00007-16.abstract>>.

NG, Melinda *et al.* Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats. *eLife*, v. 4, n. DECEMBER2015, p. 1–22, 2015.

NORBURY, Christopher C. Drinking a lot is good for dendritic cells. *Immunology*, v. 117, n. 4, p. 443–451, 2006.

OKUMURA, A. *et al.* Interaction between Ebola Virus Glycoprotein and Host Toll-Like Receptor 4 Leads to Induction of Proinflammatory Cytokines and SOCS1. *Journal of Virology*, v. 84, n. 1, p. 27–33, 2010. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.01462-09>>.

OKUMURA, Atsushi *et al.* Suppressor of Cytokine Signaling 3 Is an Inducible Host Factor That Regulates Virus Egress during Ebola Virus Infection. *Journal of virology*, v. 89, n. 20, p. 10399–406, 2015. Disponível em: <<http://jvi.asm.org/content/89/20/10399.abstract>>.

OLIVAL, Kevin; HAYMAN, David. Filoviruses in Bats: Current Knowledge and Future Directions. *Viruses*, v. 6, n. 4, p. 1759–1788, 2014. Disponível em: <<http://www.mdpi.com/1999-4915/6/4/1759/>>.

PIPIROU, Zoi *et al.* Mouse LSECtin as a model for a human Ebola virus receptor. *Glycobiology*, v. 21, n. 6, p. 806–812, 2011.

POWLESLAND, Alex S *et al.* UKPMC Funders Group Author Manuscript UKPMC Funders Group Author Manuscript A novel mechanism for LSECtin binding to Ebola virus surface glycoprotein through truncated glycans *. v. 283, n. 1, p. 593–602, 2008.

PRICE, Morgan N.; DEHAL, Paramvir S.; ARKIN, Adam P. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, v. 26, n. 7, p. 1641–1650, 2009.

PRINS, K. C.; CARDENAS, W. B.; BASLER, C. F. Ebola Virus Protein VP35 Impairs the Function of Interferon Regulatory Factor-Activating Kinases IKK and TBK-1. *Journal of Virology*, v. 83, n. 7, p. 3069–3077, 2009. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.01875-08>>.

QUINN, K. *et al.* Rho GTPases Modulate Entry of Ebola Virus and Vesicular Stomatitis Virus Pseudotyped Vectors. *Journal of Virology*, v. 83, n. 19, p. 10176–10186, 2009. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.00422-09>>.

REID, S.P.a *et al.* Ebola virus VP24 binds kaiyopherin α 1 and blocks STAT1 nuclear accumulation. *Journal of Virology*, v. 80, n. 11, p. 5156–5167, 2006. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-33646754920&partnerID=40&md5=687ccfa293b224dcaef0ebfb40cb805a>>.

RITCHIE, Matthew E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, v. 43, n. 7, p. e47, 2015.

SAEED, Mohammad F. *et al.* Phosphoinositide-3 kinase-akt pathway controls cellular entry of ebola virus. *PLoS Pathogens*, v. 4, n. 8, 2008.

SAKURAI, Yasuteru *et al.* Drug Targets for Disease Treatment. v. 347, n. 6225, p. 995–998, 2016.

SANCHEZ, Anthony. Analysis of Filovirus Entry into Vero E6 Cells, Using Inhibitors of Endocytosis, Endosomal Acidification, Structural Integrity, and Cathepsin (B and L) Activity. *The Journal of Infectious Diseases*, v. 196, n. s2, p. S251–S258, 2007. Disponível em: <<http://jid.oxfordjournals.org/lookup/doi/10.1086/520597>>.

SCHUMANN, M.; GANTKE, T.; MUHLBERGER, E. Ebola Virus VP35 Antagonizes PKR Activity through Its C-Terminal Interferon Inhibitory Domain. *Journal of Virology*, v. 83, n. 17, p. 8993–8997, 2009. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.00523-09>>.

SHIMOJIMA, M. *et al.* Tyro3 Family-Mediated Cell Entry of Ebola and Marburg Viruses. *Journal of Virology*, v. 80, n. 20, p. 10109–10116, 2006. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.01157-06>>.

SHIMOJIMA, Masayuki; IKEDA, Yasuhiro; KAWAOKA, Yoshihiro. The Mechanism of Axl-Mediated Ebola Virus Infection. *The Journal of Infectious Diseases*, v. 196, n. s2, p. S259–S263, 2007. Disponível em: <<https://academic.oup.com/jid/article->

lookup/doi/10.1086/520594>.

SIMMONS, Graham *et al.* DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology*, v. 305, n. 1, p. 115–123, 2003.

SIMMONS, James A *et al.* Ebolavirus Glycoprotein Directs Fusion through NPC1 % Endolysosomes. v. 90, n. 1, p. 605–610, 2016.

SINN, P. L. *et al.* Lentivirus Vectors Pseudotyped with Filoviral Envelope Glycoproteins Transduce Airway Epithelia from the Apical Surface Independently of Folate Receptor Alpha. *Journal of Virology*, v. 77, n. 10, p. 5902–5910, 2003. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.77.10.5902-5910.2003>>.

SPENCE JS, KRAUSE TB, MITTLER E, JANGRA RK, Chandran K. Direct Visualization of Ebola Virus Fusion Triggering in the Endocytic Pathway. *MBio.*, v. 7, n. 1, p. 1–12, 2016.

STROM, Brian L. *et al.* Data Sharing, Year 1 — Access to Data from Industry-Sponsored Clinical Trials. *New England Journal of Medicine*, v. 371, n. 22, p. 2052–2054, 2014. Disponível em: <<http://www.nejm.org/doi/10.1056/NEJMp1411794>>.

SULLIVAN, Nancy J *et al.* Ebola Virus Glycoprotein Toxicity Is Mediated by a Dynamin-Dependent Protein-Trafficking Pathway Ebola Virus Glycoprotein Toxicity Is Mediated by a Dynamin-Dependent Protein-Trafficking Pathway. *Journal of virology*, v. 79, n. 1, p. 547–553, 2005.

TAKADA, A; KAWAOKA, Y. The pathogenesis of Ebola hemorrhagic fever. *Trends Microbiol*, v. 9, n. 10, p. 506–511, 2001.

TAKADA, Ayato *et al.* Human Macrophage C-Type Lectin Specific for Galactose and N - Acetylgalactosamine Promotes Filovirus Entry Human Macrophage C-Type Lectin Specific for Galactose and N -Acetylgalactosamine Promotes Filovirus Entry. v. 78, n. 6, p. 2943–2947, 2004.

TATTEVIN, Pierre; DURANTE-MANGONI, Emanuele; MASSAQUOI, Moses. Does this patient have Ebola virus disease? *Intensive Care Medicine*, v. 40, n. 11, p. 1738–1741, 2014.

TETHERIN, Cellular Protein Bst- *et al.* Antiviral Activity of the Interferon-Induced. *AIDS Research and Human Retroviruses*, v. 25, n. 12, 2009.

VANDE BURGT, Nathan H.; KALETSKY, Rachel L.; BATES, Paul. Requirements within the Ebola viral glycoprotein for tetherin antagonism. *Viruses*, v. 7, n. 10, p. 5587–5602, 2015.

VLIET, Sandra J Van *et al.* DC-SIGN – ICAM-2 interaction medi- ates dendritic cell trafficking. *America*, v. 1, n. 4, 2000.

WANG, Han *et al.* Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1. *Cell*, v. 164, n. 1–2, p. 258–268, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.cell.2015.12.044>>.

WATANABE, Shinji *et al.* Ebola Virus (EBOV) VP24 Inhibits Transcription and Replication of the EBOV Genome. *The Journal of Infectious Diseases*, v. 196, n. s2, p. S284–S290, 2007. Disponível em: <<https://academic.oup.com/jid/article-lookup/doi/10.1086/520582>>.

XUE, Qiong; MILLER-JENSEN, Kathryn. Systems biology of virus-host signaling network interactions. *BMB Reports*, v. 45, n. 4, p. 213–220, 2012.

YAN, Huimin *et al.* Targeting C-Type Lectin Receptors for Cancer Immunity. *Frontiers in immunology*, v. 6, n. August, p. 408, 2015. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fimmu.2015.00408/abstract>>.

YUAN, Shuai *et al.* TIM-1 acts a dual-attachment receptor for Ebolavirus by interacting directly with viral GP and the PS on the viral envelope. *Protein and Cell*, v. 6, n. 11, p. 814–824, 2015.

ZAMPIERI, C. A. *et al.* The ERK Mitogen-Activated Protein Kinase Pathway Contributes to Ebola Virus Glycoprotein-Induced Cytotoxicity. *Journal of Virology*, v. 81, n. 3, p. 1230–1240, 2007. Disponível em: <<http://jvi.asm.org/cgi/doi/10.1128/JVI.01586-06>>.

ZHANG, Adrianna P P *et al.* The ebola virus interferon antagonist VP24 directly binds STAT1 and has a novel, pyramidal fold. *PLoS Pathogens*, v. 8, n. 2, 2012.

ZHAO, Dianyuan *et al.* The Myeloid LSEctin Is a DAP12-Coupled Receptor That Is Crucial for Inflammatory Response Induced by Ebola Virus Glycoprotein. *PLoS Pathogens*, v. 12, n. 3, p. 1–20, 2016. Disponível em: <<http://dx.doi.org/10.1371/journal.ppat.1005487>>.

ZHENG, Hui *et al.* Ebolavirus Classification Based on Natural Vectors. *DNA and Cell Biology*, v. 34, n. 6, p. 150324073026001, 2015. Disponível em: <<http://online.liebertpub.com/doi/10.1089/dna.2014.2678>>.

8. MATERIAL SUPLEMENTAR

Tabela de Biointerações - T1

Term 1	Biointeractions	Term 2	PMID
CTSB	Cleavage of	Ebola's GP	22031933-16571833 -21866103-22238307
CTSL	Cleavage of	Ebola's GP	22031933-16571833 -21866103-22238307
NPC1	Fusion of	Ebola's GP	26771495-27238017-27303731 -26698106-26468524-21866103-21866101
NPC2	Associated	NPC1 and Ebola's GP	26771495-15071184
CDC42	Used to be a binder	EBOLA TRANS- INFECTION	10975523-23830563
RAB5	Required for the early endosome	Associated to Ebola	23342373
RAB7	Maturation of endosome	Associated to Ebola	26861015-23342373
RAC1	used by macropinocytosis	EBOLA ENTRY	23342373-18769720
ADAM17	Cleaves	GP	15103332
MFGE8	USED LIKE BRIDGE	PTDSER	24478428
ITGB3	BIND MFGE8	PTDSER	24478428
ITGAV	BIND MFGE8	PTDSER	24478428
ITGB1	Down-regulated	KPNA1,SOCS3	17718913
ITGA6	Down-regulated	KPNA1,SOCS3	26379663-26516900
ITGA1	Down-regulated	KPNA1,SOCS3	26379663-26516900
ITGA4	Down-regulated	KPNA1,SOCS3	26379663-26516900
ITGA5	Down-regulated	KPNA1,SOCS3	26379663-26516900
ITGA3	Down-regulated	KPNA1,SOCS3	26379663-26516900

ITGA2	Down-regulated	KPNA1,SOCS3	26379663-26516900
EGFR	bind in macropinocytosis	EBOLA ENTRY	26516900
FOLR1	bindo to GP	EBOLA ENTRY	11461707-14645601-12719583
TYRO3	EBOLA ENTRY	GAS6 BIND TO TYRO3	17005688-25277499-22156524-17940958-21047970-18040157
GAS6	EBOLA ENTRY	GAS6 BINDS PTDSER	17005688-21501828-17940958-21047970-18040157
UFO-AXL	EBOLA ENTRY	GAS6 BINDS UFO-AXL	17005688-25277499-22156524-21529875-24478428-17940958-21047970-18040157
MERTK	EBOLA ENTRY	GAS6 BINDS MERTK	17005688-25277499-17940958-21047970-18040157
ICAM3	Associated	EBOLA TRANS- INFECTION	25642836-11017109-21987776
ICAM2	Associated	EBOLA TRANS- INFECTION	1017109
CLEC10A	Binds GP	EBOLA TRANS- INFECTION	21866103-22031933-14990712
CD209	Bind GP	EBOLA ENTRY	12050398-12504546-11017109 -12502850-12186897-21855102-12223058-14990712-18083206
BST2(tetherin)	Antiviral Response	GP Antagonizes	25396298-26516900-27707924-19179289-19929170-20702620-20485522
CLEC4M	EBOLA ENTRY	Binds GP	12050398-12504546-14990712
TIMD4	EBOLA ENTRY	Binds PTDSER	23555248
HAVCR2	EBOLA ENTRY	Binds PTDSER	25136083-23555248
HAVCR1	EBOLA ENTRY	Binds PTDSER	26487564 -25277499-23555248-21536871 -26655238 -25855742-23698310
IKBKE	Associated to IFN response	vp35	19153231-19694547
TBK1	Associated to IFN response	vp35	19153231-19694547
KPNA1	VP24 Interferon supression	Binds STAT1	16698996-17928350-19447180
IL-4	Associated to expression of	CLEC4G	17339424-19111020
CLEC4G	BINDS GP	EBOLA ENTRY	16051304-17984090-18083206-21257728-26943817-17339424-19111020
IQGAP1	Associated to assembly and budding	VP40	23637409
IRF7	down-regulated	TBK1 and IKKE	25396298-21263462-19153231
MBL2	BINDS C1QBP	EBOLA ENTRY	23573288-16099912
C1QBP	BINDS MBL2	EBOLA ENTRY	23573288

CLEC6A	BINDS GP	EBOLA ENTRY	23573288
TPC2	Egress of RNA	Virus' Rna	26468524-25722412
TSG101	Associated to assembly and budding	VP40	14673115-12559917-18305167-12525615-17570046-12941909-11095724-17940959
VPS4A	Associated to assembly and budding	VP40	14673115-12559917-18305167-12525615-17570046-12941909-11095724-17940959
NEDD4	Associated to assembly and budding	VP40	14673115-12559917-18305167-12525615-17570046-12941909-11095724-17940959
SOCS3	Down-regulate	TLR4	26246577
TLR4	Associated Cytokine signal	EBOLA ENTRY	26246577-26003830-19846529
STAT1	Associated IFN response, down-regulated	KPNA1	22383882-19846529-19447180
RHOC	Associated to actin	EBOLA ENTRY	19625394
RHOB	Associated to actin	EBOLA ENTRY	19625394
IRF3	Associated IFN response, down-regulated	TBK1 and IKKE	25396298-21263462 -19153231 -16775331-18353943-20018655-19515768
EP300	IFN RESPONSE	IRF3 interacts with EP300	25396298-21263462 -19153231 -16775331-18353943-19515768
CREBBP	IFN RESPONSE	IRF3 interacts with EP300	25396298-21263462 -19153231 -16775331-18353943-19515768
MHC2	The virus' apresentation	EBOLA TRANS-INFECTION	25493356
MHC1	The virus' apresentation	EBOLA TRANS-INFECTION	15596847-25493356
EPS15	Associated clathrin-mediated endocytosis	ENDOCYTOSIS	21855102
DAB2IP	Associated clathrin-mediated endocytosis	ENDOCYTOSIS	21855102
TFAP2A	Associated clathrin-mediated endocytosis	ENDOCYTOSIS	21855102
ASGR1	Binds GP	EBOLA ENTRY	24473128-25877552
AKT1	Associated to cytokine signal	EBOLA ENTRY	18769720
PIK3R1	Associated to cytokine signal	EBOLA ENTRY	18769720
CD300A	BIND TO PTDSER	EBOLA ENTRY	24478428
PRKRA(pact)	VP35 downregulated	Associated to IFN response	23870315

DDX58(rig-1)	Binds PRKRA	Associated to IFN response	23870315-16698997
EIF2AK2	VP35 downregulated	Associated to dsRNA response	17065211
EIF2S1	Binds EIF2S1	Associated to dsRNA response	17065211
RAB9A	Structure of endosome	Associated to Ebola	16140752-15263003
ERK	Binds NFkB	Associated to cytokine signal	17434557-17108034
CD44	Binds GP	Associated Endocytosis by actin	Kegg pathway: 05205,05203
IGF1R	Binds GP	Associated to cytokine response	Kegg pathway: 05205,05203
PAK1	Binds RAC1 or CDC42	Associated Endocytosis by actin	Kegg pathway: 05205,05203
RASA2	Associated Endocytosis by actin	Binds CD44	Kegg pathway: 05205,05203
RHOA	Associated Endocytosis by actin	Binds CD44	19625394
TIAM1	Associated Endocytosis by actin	Binds CD44	Kegg pathway: 05205,05203
VAV2	Associated Endocytosis by actin	Binds CD44	Kegg pathway: 05205,05203
ssGP	Secreted of cell	Interacts of Adam17	21411529
PP1	Associated to transcription, translation and replication	Binds VP30	26937028-28192094-24936058
TOP1	Associated to transcription, translation and replication	Binds L	26937028-28192094-24936058
LC8	Associated to transcription, translation and replication	Binds VP35	26937028-28192094-24936058
CLATHRIN(complexo)	Involved the endocytosis's membrane	EBOLA ENTRY	21987776-17940957-20202662-21855102
COMPLEXO HOPS(VPS)	Endosome associated	Ebola Internalization	21866103-23302883-22395071
CAVEOLIN(complexo)	Endocytosis associated	EBOLA ENTRY	21047970-11967340-21855102-21987776

Tabela de Comparação de resultados entre os *softwares* - T2.

	LCA-SS	TAXONTREE	GENEAGES	KO
CTSB	2	2	Match	Invalidated
CTSL	2	3	Query_Recent	Query_Recent
NPC1	2	2	Match	Match
CDC42	2	3	Match	Query_Recent
RAB5A	2	2	Match	Match
RAB7	2	2	Match	Match
RAC1	2	2	Match	Match
RHOB	3	3	Query_Recent	Query_Old
ADAM17	3	6	Query_Recent	Query_Recent
NPC2	4	4	Query_Old	Query_Recent
ITGAV	5	6	Match	Query_Recent
ITGB1	4	6	Match	Query_Recent
ITGA6	6	6	Match	Match
ITGA1	2	13	Match	Match
ITGA4	6	13	Query_Recent	Query_Recent
ITGA5	5	6	Match	Query_Old
ITGA3	11	6	Query_Recent	Match
ITGA2	3	13	Query_Recent	Match
EGFR	3	6	Query_Recent	Query_Recent
FOLR1	2	4	Query_Old	Query_Recent
ITGB3	13	6	Query_Recent	Match
TYRO3	13	13	Query_Recent	Match
GAS6	13	13	Match	Match
ICAM3	18	21	Query_Recent	Query_Recent
ICAM2	13	20	Match	Query_Recent
CLEC10A	13	21	Match	Query_Recent
BST2	21	20	Match	Query_Recent
CLEC4M	13	27	Match	Query_Recent
CD209	21	26	Query_Recent	Query_Recent
HAVCR1	20	19	Match	Query_Recent
CLTC	2	2	Match	Match
CLTA	2	6	Match	Match
CLTB	3	4	Query_Recent	Query_Recent
CLTCL1	4	2	Match	Match
AXL	13	13	Match	Match
IKBKE	5	6	Match	Query_Old
TBK1	4	6	Match	Query_Recent

KPNA1	3	2	Match	Query_Old
CLEC4G	16	17	Match	Query_Recent
PP1	3	13	Query_Recent	Query_Recent
LC8	2	2	Match	Match
TOP1	1	6	Query_Recent	Invalidated
IQGAP1	3	3	Query_Recent	Match
MERTK	11	13	Match	Query_Recent
IRF7	13	13	Match	Match
MBL2	13	13	Query_Recent	Match
C1QBP	2	6	Query_Recent	Query_Recent
CLEC6A	21	21	Query_Recent	Match
TIMD4	16	20	Query_Recent	Invalidated
TPCN2	2	4	Query_Recent	Match
TSG101	2	2	Match	Match
NEDD4	2	6	Query_Recent	Query_Recent
SOCS3	8	7	Match	Query_Old
TLR4	13	13	Match	Query_Recent
STAT1	10	13	Match	Query_Recent
RHOC	3	3	Query_Recent	Query_Old
IRF3	11	13	Match	Query_Recent
ASGR1	20	21	Query_Recent	Query_Recent
CAV1	5	5	Match	Query_Old
CAV2	5	6	Match	Match
CAV3	5	4	Query_Old	Match
PIK3R1	4	6	Query_Recent	Query_Recent
PIK3CA	2	2	Match	Match
PIK3CB	2	2	Match	Match
PIK3CD	2	2	Match	Match
PIK3R2	3	6	Query_Old	Query_Recent
PIK3R3	4	4	Query_Old	Query_Recent
AKT1	2	2	Match	Invalidated
ACTB	1	3	Query_Recent	Query_Recent
ACTG1	1	3	Query_Recent	Query_Recent
ACTN1	2	3	Query_Recent	Query_Recent
ACTN4	2	3	Query_Recent	Query_Recent
CD300A	11	19	Query_Recent	Query_Recent
CD44	13	17	Query_Recent	Query_Recent
CREBBP	2	6	Query_Recent	Query_Recent
DAB21P	3	6	Query_Recent	Query_Recent
DDX58	3	6	Query_Recent	Match
EIF2AK2	2	13	Query_Recent	Query_Recent

EIF2S1	1	2	Query_Recent	Invalidated
EP300	2	6	Query_Recent	Query_Recent
EPS15	2	6	Query_Recent	Query_Recent
FLNA	2	4	Query_Recent	Query_Recent
FLNB	2	4	Query_Recent	Query_Recent
FLNC	2	4	Query_Old	Query_Recent
GSN	1	2	Match	Match
SCIN	2	2	Match	Match
HAVCR2	1	18	Query_Recent	Query_Old
IGF1R	6	6	Query_Recent	Query_Recent
IL4	13	17	Match	Query_Recent
MAPK1	2	3	Query_Recent	Query_Recent
MAPK3	2	3	Query_Recent	Query_Recent
MFGE8	6	7	Match	Query_Recent
NFKB1	4	4	Match	Match
NFKB2	13	4	Query_Recent	Match
REL	6	13	Query_Recent	Query_Recent
RELA	6	13	Query_Recent	Match
RELB	6	13	Query_Recent	Match
PAK1	2	3	Query_Recent	Query_Recent
PRKRA	6	6	Match	Invalidated
RAB9A	2	2	Match	Query_Old
RASA2	6	4	Query_Old	Query_Old
RHOA	3	3	Query_Recent	Query_Recent
TFAP2A	6	6	Match	Match
TIAM1	6	13	Query_Recent	Query_Recent
VAV2	3	4	Query_Old	Query_Recent
VPS11	2	2	Match	Match
VPS16	2	2	Match	Match
VPS18	2	2	Match	Match
VPS33A	2	2	Match	Match
VPS39	2	2	Match	Match
VPS41	2	2	Match	Match
VPS4A	2	2	Query_Recent	Invalidated
HLA-A	13	10	Match	Invalidated
HLA-B	13	10	Invalidated	Invalidated
HLA-C	13	10	Match	Invalidated
HLA-DMA	13	10	Match	Invalidated
HLA-DMB	13	10	Match	Query_Old
HLA-DOA	13	10	Match	Query_Old
HLA-DOB	13	10	Match	Query_Old

HLA-DPA1	13	10	Match	Query_Old
HLA-E	5	5	Match	Invalidated
HLA-F	13	10	Match	Invalidated
HLA-G	13	10	Match	Invalidated
HLA-DPB1	13	10	Match	Query_Old
HLA-DQA1	13	10	Match	Query_Old
HLA-DQA2	13	10	Match	Query_Old
HLA-DQB1	13	10	Match	Query_Old
HLA-DQB2	13	10	Match	Invalidated
HLA-DRA	13	10	Match	Query_Old
HLA-DRB1	13	10	Match	Invalidated
HLA-DRB3	13	10	Match	Query_Old
HLA-DRB4	13	10	Match	Query_Old
HLA-DRB5	13	10	Match	Query_Old

Figura Suplementar 1- WikiPathways.

Imagem representativa da disponibilização da via de interação criada.

Pathway:WP4217.

Ebola Virus Pathway on Host (Homo sapiens)

Elisson Lopes

Search:

Help: [Help](#), [About us](#), [Contact us](#), [Report a bug](#), [How to cite](#)

Download: [Download files](#), [Web service API](#), [WikiPathways RDF](#), [Embed code](#)

Activity: [Browse pathways](#), [Recent changes](#), [New pathways](#), [Edit pathways](#), [Create pathway](#), [Statistics](#)

Tools: [PathwayWidget](#), [Tissue expression](#), [Pathway Finder](#), [Software tools](#)

Community: [Quality control](#), [Development](#), [WikiPathways Blog](#), [AOP portal](#), [CIRM portal](#), [CPTAC portal](#), [Renal Genomics portal](#), [Disease portal](#), [ExRNA portal](#), [Micronutrient portal](#), [Nanomaterials portal](#), [NetPath portal](#), [Plants portal](#), [Reactome portal](#), [WormBase portal](#)

Launch Editor not working?

RESET

Download

Description

Study of Ebola interaction with human proteins. The course of the representation it occurred by left to right. The numbers on circles are related study of the evolutionary levels (provisional data).

Quality Tags

hide

No tags

Ontology Terms