

# Predição de Efetividade do tratamento de HIV

Giordano Bruno Olivetti Mattiello - 173056

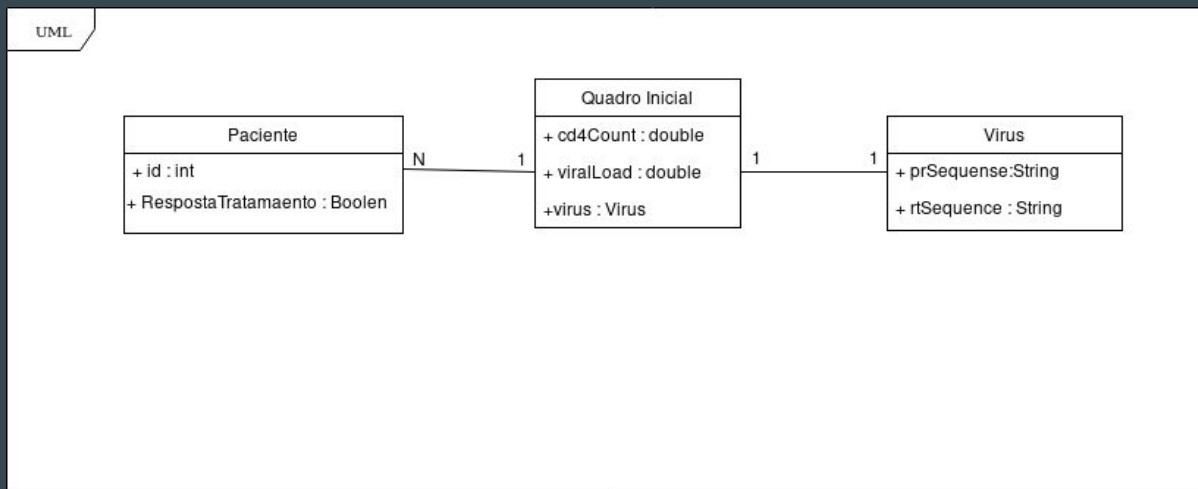
Daniela Marques de Moraes - 169562

# Vírus HIV

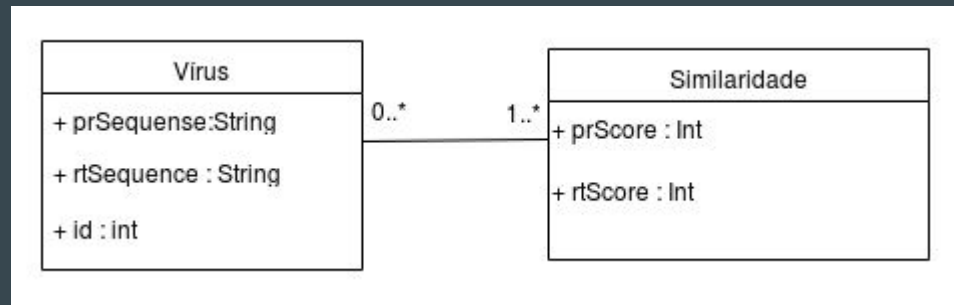
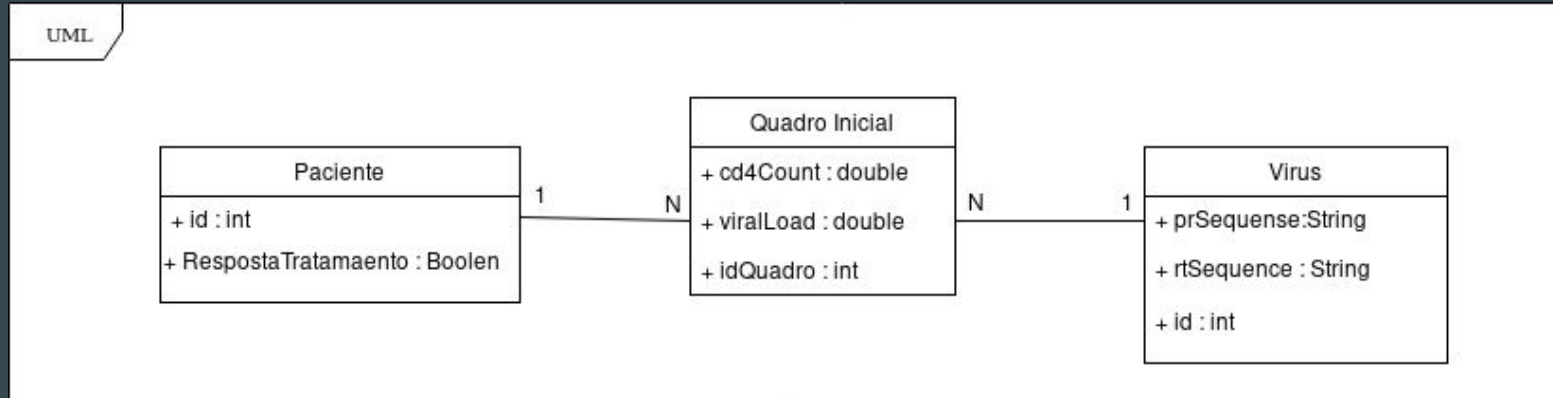
- Mata milhões de pessoas ao redor do mundo todo
- Reconhecido em 1981 pela primeira vez

Um passo para a compreensão do vírus é entender o seu projeto genético, neste estudo pretendemos identificar a relação entre a estrutura do vírus e a efetividade do seu tratamento usando machine learning para criar estes resultados.

# Primeiro Modelo Conceitual



# Modelo Conceitual Final



# Contagem de CD4

- Até que ponto o sistema imunológico pode proteger o corpo contra infecções?
- Qual é a gravidade dos danos causados pelo HIV?
- Saudável: 500 - 1000. Um número inferior a 50 células por microlitro de sangue é especialmente perigoso, podem surgir mais infecções oportunistas e causar perda grave de peso, cegueira ou morte.

# Carga viral

- O quanto a infecção é contagiosa
- Com que rapidez a contagem de CD4 tem probabilidade de diminuir
- Com que rapidez os sintomas têm probabilidade de surgir

# Etapa 2 - Relacional

# Modelo Lógico

paciente(id, respostaTratamento)

quadroInicial(idQuadro, idPaciente, idVirus, cd4Count, viralLoad)

As chaves estrangeiras são:

idPaciente -> paciente.id

idVirus -> virus.id

virus(id, prSequunce, rtSequence)



## Buscar por vírus que possuem rtSequence e prSequence iguais

```
SELECT * FROM (SELECT count(*) AS c FROM virus GROUP BY rtSequence, prSequence) n WHERE n.c >= 2
```

Resultado: Nem um vírus tem a mesma rtSequence e prSequence

## Buscar por pacientes que possuem cd4 menor que 50

```
SELECT COUNT(q.idPaciente) FROM QuadroInicial q WHERE cd4Count < 50
```

Resultado : 89 dentre 1000 pacientes tem cd4 menor que 50

## Buscar por pacientes que possuem cd4 menor que 50 que reagiram bem ao tratamento

```
SELECT COUNT(Q.IdPaciente) FROM QuadroInicial Q,Paciente P WHERE cd4Count < 50 AND P.Id= Q.IdPaciente AND RespostaTratamento = true
```

Resultado: 35 de 89 pacientes

## A carga viral máxima de uma pessoa que teve sucesso no tratamento

```
SELECT MAX(Q.viralLoad) FROM QuadroInicial Q, Paciente P WHERE Q.IdPaciente = P.Id AND P.RespostaTratamento = true
```

Resultado: 575.0

## A carga cd4 mínima de uma pessoa que teve sucesso no tratamento

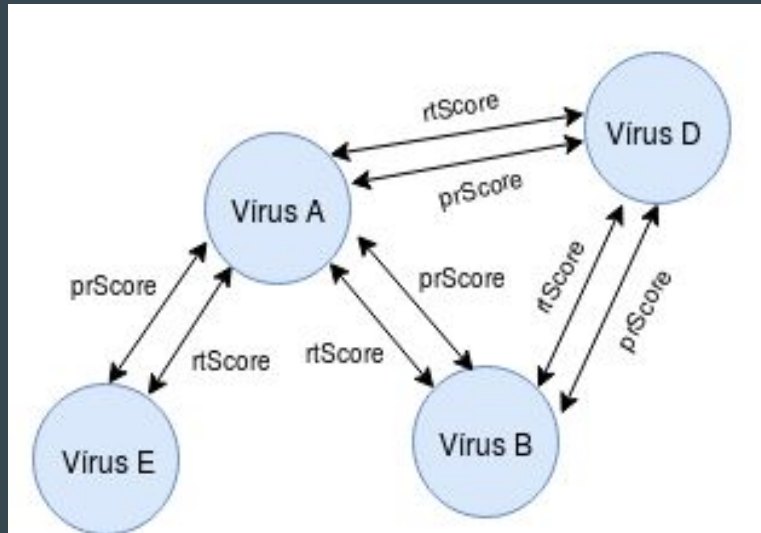
```
SELECT MIN(Q.viralLoad) FROM QuadroInicial Q, Paciente P WHERE Q.IdPaciente = P.Id AND P.RespostaTratamento = true;
```

Resultado: 2.9

# Etapa 3 - Não relacional

# Modelo Etapa 3

Para encontrar as similaridades genéticas, foi utilizado o algoritmo Needleman-Wunsch, o qual propõe uma matriz de scores para determinar o quão próximos são duas sequências.



Database Sequence

		G	C	A	T	G	C	U	
		0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5	
A	-2	0	0	1	0	-1	-2	-3	
T	-3	-1	-1	0	2	1	0	-1	
T	-4	-2	-2	-1	1	1	0	-1	
A	-5	-3	-3	-1	0	0	0	-1	
C	-6	-4	-2	-2	-1	-1	1	0	
A	-7	-5	-3	-1	-2	-2	0	0	

Query Sequence

Buscar os nós que possuem o mais alto score de RT

```
MATCH (v:Virus)-[s:Similaridade]->(v2:Virus) RETURN v,v2 ORDER BY toInteger(s.rtscore) DESC LIMIT 1
```

Buscar os nós que possuem o mais alto score de PR

```
MATCH (v:Virus)-[s:Similaridade]->(v2:Virus) RETURN v,v2 ORDER BY toInteger(s.prscore) DESC LIMIT 1
```

Buscar os nós que possuem o RT score acima da média

```
MATCH (v:Virus)-[s:Similaridade]->(v2:Virus) WHERE toInteger(s.rtscore) > 950 RETURN v,v2
```

Buscar os nós que possuem CD4 num nível crítico (abaixo de 50)

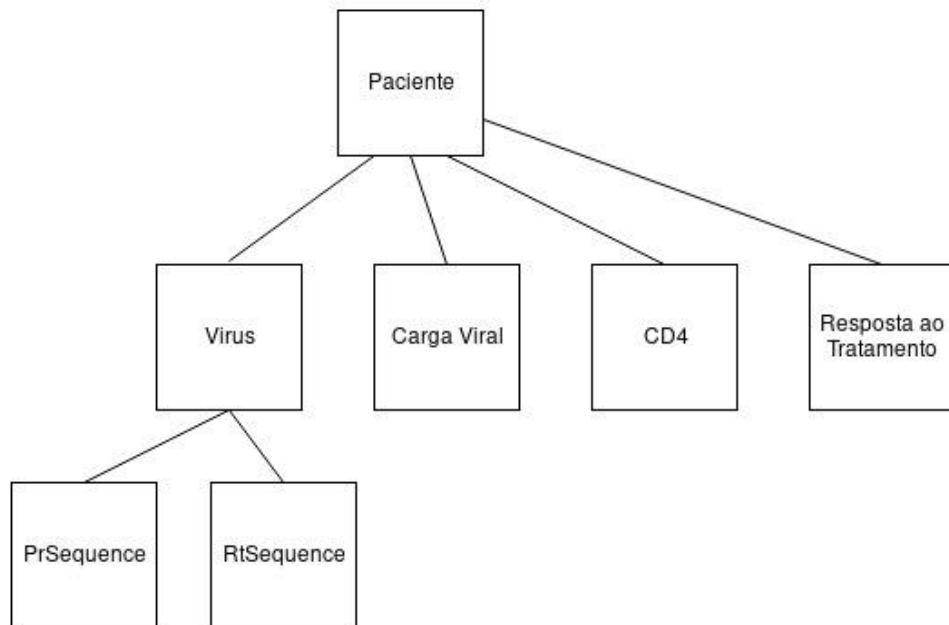
```
MATCH (v:Virus)-[s:Similaridade]->(v2:Virus) WHERE toInteger(v.cd4) < 50 AND toInteger(v2.cd4) < 50 RETURN v,v2
```

Buscar a média dos scores de RT e PR

```
MATCH (v:Virus)-[s:Similaridade]->(v2:Virus) RETURN avg(toInteger(s.rtscore)),avg(toInteger(s.prscore))
```

# Etapa 4 - XML e SPARQL

# Modelo Etapa 4



```

5  ## Obter todos os tipos de infecções HIV
6  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
7  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
8  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
9  PREFIX owl: <http://www.w3.org/2002/07/owl#>
10 PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>
11 PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
12 PREFIX mesh2015: <http://id.nlm.nih.gov/mesh/2015/>
13 PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
14 PREFIX mesh2017: <http://id.nlm.nih.gov/mesh/2017/>
15
16 SELECT DISTINCT ?descriptor ?label
17 FROM <http://id.nlm.nih.gov/mesh>
18
19 WHERE {
20     mesh:D015658 meshv:treeNumber ?treeNum .
21     ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
22     ?descriptor meshv:treeNumber ?childTreeNum .
23     ?descriptor rdfs:label ?label .
24 }
25
26 ORDER BY ?label
27 LIMIT 10

```

```

1  ## Obter todos os tipos de doenças sexualmente transmissíveis
2  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
5  PREFIX owl: <http://www.w3.org/2002/07/owl#>
6  PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>
7  PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
8  PREFIX mesh2015: <http://id.nlm.nih.gov/mesh/2015/>
9  PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
10 PREFIX mesh2017: <http://id.nlm.nih.gov/mesh/2017/>
11
12 SELECT DISTINCT ?descriptor ?label
13 FROM <http://id.nlm.nih.gov/mesh>
14
15 WHERE {
16     ?descriptor rdfs:label ?label .
17     ?descriptor meshv:broaderDescriptor mesh:D015229
18 }
19
20 ORDER BY ?label

```



```

1  ## Obter todas as annotations das infecções HIV
2  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
5  PREFIX owl: <http://www.w3.org/2002/07/owl#>
6  PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>
7  PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
8  PREFIX mesh2015: <http://id.nlm.nih.gov/mesh/2015/>
9  PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
10 PREFIX mesh2017: <http://id.nlm.nih.gov/mesh/2017/>
11
12 SELECT DISTINCT ?label ?annotation
13 FROM <http://id.nlm.nih.gov/mesh>
14
15 WHERE {
16     mesh:D015658 meshv:treeNumber ?treeNum .
17     ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
18     ?descriptor meshv:treeNumber ?childTreeNum .
19     ?descriptor rdfs:label ?label .
20     ?descriptor meshv:annotation ?annotation
21 }
22
23 ORDER BY ?label
24 LIMIT 10

```

```

26 ## Obter todas as doenças relacionadas a slow virus
27 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
28 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
29 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
30 PREFIX owl: <http://www.w3.org/2002/07/owl#>
31 PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>
32 PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
33 PREFIX mesh2015: <http://id.nlm.nih.gov/mesh/2015/>
34 PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
35 PREFIX mesh2017: <http://id.nlm.nih.gov/mesh/2017/>
36
37 SELECT DISTINCT ?descriptor ?label
38 FROM <http://id.nlm.nih.gov/mesh>
39
40 WHERE {
41     mesh:D012897 meshv:treeNumber ?treeNum .
42     ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
43     ?descriptor meshv:treeNumber ?childTreeNum .
44     ?descriptor rdfs:label ?label .
45 }
46
47 ORDER BY ?label
48 LIMIT 10

```

```

50  ## Obter todas as síndromes de deficiência imunológica
51  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
52  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
53  PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
54  PREFIX owl: <http://www.w3.org/2002/07/owl#>
55  PREFIX meshv: <http://id.nlm.nih.gov/mesh/vocab#>
56  PREFIX mesh: <http://id.nlm.nih.gov/mesh/>
57  PREFIX mesh2015: <http://id.nlm.nih.gov/mesh/2015/>
58  PREFIX mesh2016: <http://id.nlm.nih.gov/mesh/2016/>
59  PREFIX mesh2017: <http://id.nlm.nih.gov/mesh/2017/>
60
61  SELECT DISTINCT ?descriptor ?label
62  FROM <http://id.nlm.nih.gov/mesh>
63
64  WHERE {
65      mesh:D007153 meshv:treeNumber ?treeNum .
66      ?childTreeNum meshv:parentTreeNumber+ ?treeNum .
67      ?descriptor meshv:treeNumber ?childTreeNum .
68      ?descriptor rdfs:label ?label .
69  }
70
71  ORDER BY ?label
72  LIMIT 10

```

Quantos pacientes responderam bem ao tratamento

```
let $root := doc('mydoc.xml')
```

```
for $p in ($root)
```

```
return count($p//root//Paciente[Resp=0])
```

ID dos pacientes ainda doentes

```
let $root := doc('mydoc.xml')
```

```
for $p in ($root//Paciente)
```

```
where $p[Resp=0]
```

```
return <Doente>{data($p/PatientID)}</Doente>
```

Pessoas que a contagem de cd4 menor que 50

```
let $root := doc('mydoc.xml')
```

```
for $p in ($root//Paciente)
```

```
where $p[CD4-t0 <50]
```

```
order by $p/PatientID
```

```
return <cd4>{data($p/PatientID)}</cd4>
```

Média de contagem cd4

```
let $root := doc('mydoc.xml')
```

```
for $p in ($root)
```

```
return avg($p//Paciente/CD4-t0)
```

## Média de contagem VIRALLOAD

```
let $root := doc('mydoc.xml')
```

```
for $p in ($root)
```

```
return avg($p//Paciente/VL-t0)
```