Universidade de Lisboa
Faculdade de Ciências
Departamento de Informática

AIDS/SIDA

Home assignment 1

André Filipe Bernardes Oliveira, 45648
Mestrado em Informática (MI)

Relatório de **Bioinformática**

para o docente: André Falcão

2016/2017

Introduction

Acquired immunodeficiency syndrome (AIDS) is just an aggravated stage of another related disease: Human Immunodeficiency Virus (HIV) infection and is characterized by the onset of opportunistic infections or HIV-related cancers and a low $CD4^+$ T cell count. Hence, it is hard to analyze AIDS without exploring HIV infection as well, since that is what may lead to AIDS (patients infected with HIV will not always develop AIDS). HIV acts in such a way that it affects or destroy the host's immune system.

With that distinction in mind, this report will explore genetic characteristics more related to slow/rapid progression to AIDS than susceptibility or resistance to HIV, yet it is hard to draw the line. Despite this, it is important to understand how the virus works since that will reveal the susceptibility/resistance to AIDS. This work will focus on the type 1 virus (that is more common and virulent), within which exists strains with different tropism according to the type of the coreceptor used (chemokine receptor): CCR5 (M-tropic) or CXCR4 (T-tropic). CD4 is the primary receptor in both cases.

Besides the coreceptors and its ligands, that allow the entrance in $CD4^+$ T cells or macrophages, we should also consider the genes related to AIDS rate of progression (after HIV infection). Within this group, we find: CXCR1, CX3CR1, IFNG, IL4R and KIR3DL1. Only a brief overview (its purpose is to serve as a comparison) is going to be explored for the last three genes since they required a little bit more knowledge about immunology.

CCR5 and CCR2 are a β-chemokine involved in signal transduction (increase intracellular calcium levels). CXCR4 is also involved in signal transduction but is a α-chemokine. Even though CXCR1 belongs to the previous group of chemokines receptors, it recognizes an interleukin and transduces the signal via G-protein. Then there is a set of proteins that interact with these coreceptors: CL3L1 is a ligand for CCR5, inhibiting HIV infection; while CXCL12 is a chemoattractant for lymphocytes that activates CXCR4 (thus inhibiting CXCR4, it inhibits HIV entry). So, all these proteins are correlated in this way.

Another set of proteins may affect AIDS progression by other immune mechanisms. IFNG (Interferon-gamma) and KIR3DL1 (receptor on natural killer cells) are involved in innate and adaptive immunity against viral infections. IL4R is a cytokine that that regulates proliferation and differentiation of a several immune cells and chemokine production.

Both CXCR1 and IL4R are involved in slower progression rates of the disease and CXCL12 is usually associated with resistance to AIDS. On the other hand, CX3CR1 and IFNG are related to the rapid progression and CCL3L1 to susceptibility to AIDS. KIR3DL1 can either be associated with delayed or rapid progression rates.

Methods

The starting point for this research was OMIM database especially the "HUMAN IMMUNODEFICIENCY VIRUS TYPE 1, SUSCEPTIBILITY TO" entry (https://www.omim.org/entry/609423). From the available information, it was extracted relevant genes and proteins that could lead to the AIDS phenotype. Most of the information was obtained from the literature review and some from the Phenotype-Gene Relationships summary table. Before selecting the information, there was the need to understand the HIV infection cycle (using external sources like Wikipedia, World Health Organization or other heath institutions websites) to assess the most relevant biological sequences involved.

More information about the selected genes and proteins was retrieved from UniProt (mostly using the links available at OMIM). UniProt was used as the main hub for the information obtained like ID's, Gene Ontology Annotation, structure, among others.

A R script was developed to retrieve GO terms and count them. Despite that the information display in this report derive from ProteinON (http://www.lasige.di.fc.ul.pt/webtools/proteinon/index.html) since it provided a statistical measure to the results.

Ligands and other interacting proteins were obtained from reactome maps or from literature analysis. Comparisons of protein structure were performed considering tertiary structure only, since it is a higher level of complexity, of easier visualization and it can be inferred that similar structures have similar functions.

Most difficulties arise in the selection of the more relevant genes in OMIM because in the summary tables does not appear CCR5 and CXCR4 which seem to be critical to HIV's successful infection.

Discussion

Genes and proteins were divided in groups given the previous knowledge obtained from the literature review. Thus, the coreceptors (independent of the tropism) were grouped together, while its interacting proteins were in a different group, lastly the other three genes were in remaining group. This grouping (can be view in the table below) facilitates the comparisons between genes and consequent conclusions.

Table 1. Human genes involved in AIDS with respective proteins names and IDs.

Group	OMIM	Gene	Protein	UniProt
1	601373	*CCR5*	C-C chemokine receptor type 5	P51681
	162643	*CCR2*	C-C chemokine receptor type 2	P41597
	162643	*CXCR4*	C-X-C chemokine receptor type 4	P61073
	146929	*CXCR1*	C-X-C chemokine receptor type 1	P25024
2	601395	*CCL3L1*	C-C chemokine ligand 3-like 1	P16619
	600835	*CXCL12*	C-X-C Motif Chemokine Ligand 12 Stromal Cell-Derived Factor 1	P48061
3	147570	*IFNG*	Interferon-γ	P01579
	604946	*KIR3DL1*	Killer cell immunoglobulin-like receptor 3DL1 (Three Domains, Long cytoplasmic tail, 1)	P43629
	147781	*IL4R*	Interleukin 4 Receptor	P24394

Comparing protein structures (Figure 1), we can see that the coreceptors have a similar structure and differ from the rest, which makes sense since they all perform similar functions that allow the fusion and the entry of HIV. As for the two proteins that interact with receptors, they have a quite different structure between each other and comparing to the others, which is reasonable since they interact with different receptors and in different ways (one is ligand and the other is involved in chemotaxis).

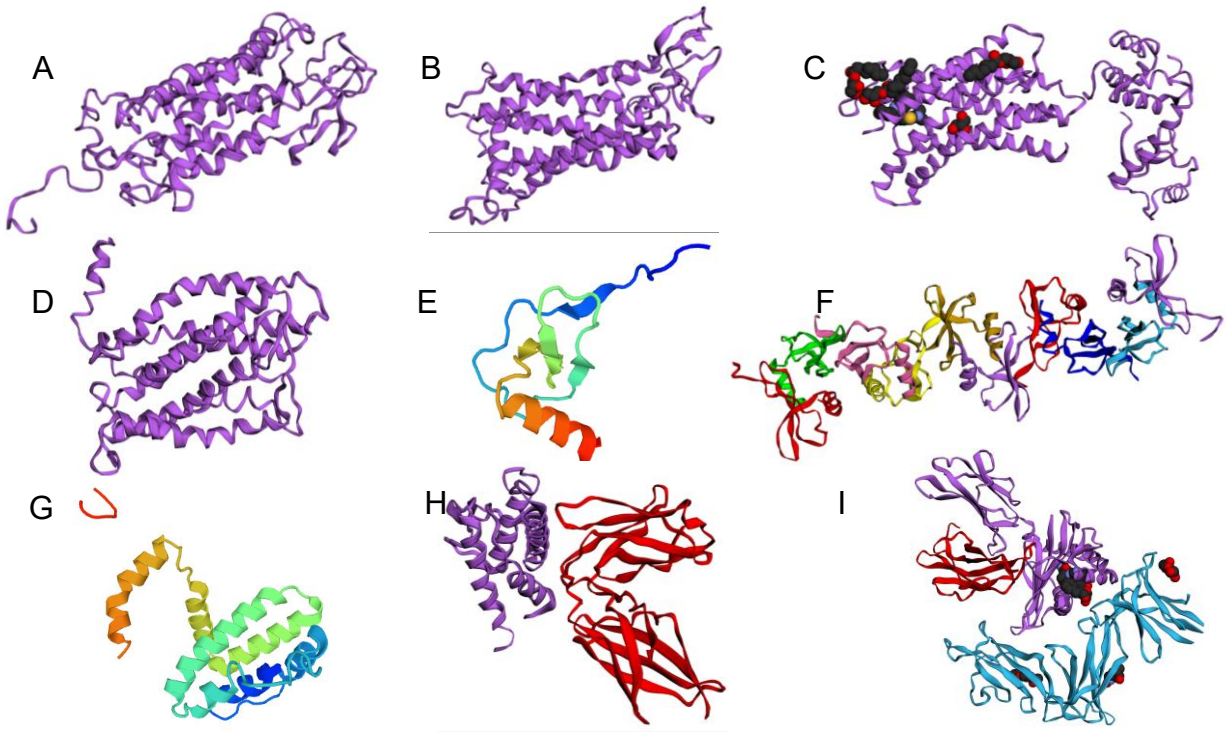

Figure 1. Tertiary Structure of the Proteins: A) CCR5, B) CCR2, C) CXCR4, D) CXCR1, E) CCL3L1, F) CXCL12, G) IFNG, H) IL4R, I) KIR3DL1 complex with HLA-B: KIR3DL1 is represented in blue.

Table 2. GO Annotations obtained from ProteinON for the following proteins: CCR5, CCR2, CXCR4 and CXCR1 (Group 1). Cutoff values were 1×10^{-7} and are represented in grey.

	Term Id	Term Name	# Proteins	%Occurence	e-value
molecular function	GO:0004950	chemokine receptor activity	4	100,0%	8,90E-16
	GO:0019956	chemokine binding	3	75,0%	6,00E-15
	GO:0019957	C-C chemokine binding	2	50,0%	4,60E-10
	GO:0015026	coreceptor activity	2	50,0%	7,60E-10
	GO:0016493	C-C chemokine receptor activity	2	50,0%	2,30E-08
	GO:0016494	C-X-C chemokine receptor activity	2	50,0%	3,70E-08
	GO:0005515	protein binding	4	100,0%	1,00E-06
	GO:0003779	actin binding	2	50,0%	1,50E-05
	GO:0019899	enzyme binding	2	50,0%	2,50E-05
biological process	GO:0002407	dendritic cell chemotaxis	4	100,0%	0,00E+00
	GO:0019221	cytokine-mediated signaling pathway	4	100,0%	4,50E-14
	GO:0006954	inflammatory response	4	100,0%	8,10E-14
	GO:0007204	elevation of cytosolic calcium ion concentration	3	75,0%	1,30E-11
	GO:0032101	regulation of response to external stimulus	3	75,0%	6,30E-10
	GO:0040012	regulation of locomotion	3	75,0%	1,40E-09
	GO:0006968	cellular defense response	2	50,0%	1,60E-09
	GO:0007186	G-protein coupled receptor signaling pathway	4	100,0%	6,00E-09
	GO:0022008	neurogenesis	3	75,0%	8,90E-09

	GO:0002437	inflammatory response to antigenic stimulus	2	50,0%	2,20E-08
	GO:0032680	regulation of tumor necrosis factor production	2	50,0%	6,40E-08
	GO:0019722	calcium-mediated signaling	2	50,0%	1,10E-07
	GO:0032103	positive regulation of response to external stimulus	2	50,0%	1,50E-07
	GO:0010604	positive regulation of macromolecule metabolic process	3	75,0%	2,40E-07
	GO:0031349	positive regulation of defense response	2	50,0%	2,50E-07
	GO:0042063	gliogenesis	2	50,0%	2,60E-07
	GO:0001819	positive regulation of cytokine production	2	50,0%	3,10E-07
	GO:0001666	response to hypoxia	2	50,0%	6,10E-07
	GO:0030334	regulation of cell migration	2	50,0%	1,40E-06
	GO:0000165	MAPK cascade	2	50,0%	3,60E-06
	GO:0007610	behavior	2	50,0%	4,30E-06
	GO:0007417	central nervous system development	2	50,0%	4,90E-06
	GO:0042127	regulation of cell proliferation	2	50,0%	1,10E-05
	GO:0035556	intracellular signal transduction	3	75,0%	6,20E-05
	GO:0016192	vesicle-mediated transport	2	50,0%	7,50E-05
	GO:0031325	positive regulation of cellular metabolic process	2	50,0%	9,40E-05
	GO:0019059	initiation of viral infection	2	50,0%	1,10E-04
	GO:0030260	entry into host cell	2	50,0%	1,10E-04
	GO:0009966	regulation of signal transduction	2	50,0%	1,40E-04
	GO:0050877	neurological system process	2	50,0%	1,40E-04
	GO:0006915	apoptotic process	2	50,0%	1,80E-04
	GO:0032940	secretion by cell	2	50,0%	2,40E-04
	GO:0051707	response to other organism	2	50,0%	2,60E-04
	GO:0022603	regulation of anatomical structure morphogenesis	2	50,0%	3,40E-04
	GO:0016043	cellular component organization	2	50,0%	8,40E-03
	GO:0006810	transport	3	75,0%	1,30E-02
	GO:0043170	macromolecule metabolic process	4	100,0%	1,50E-02
	GO:0019538	protein metabolic process	2	50,0%	1,00E-01
	GO:0044237	cellular metabolic process	3	75,0%	4,20E-01
cellular component	GO:0009986	cell surface	2	50,0%	2,30E-05
	GO:0005886	plasma membrane	4	100,0%	1,30E-04
	GO:0005887	integral to plasma membrane	2	50,0%	1,70E-04
	GO:0016021	integral to membrane	4	100,0%	9,90E-03
	GO:0044444	cytoplasmic part	3	75,0%	2,30E-02
	GO:0043231	intracellular membrane-bounded organelle	2	50,0%	1,40E-01

Table 3. GO Annotations obtained from ProteinON for the following proteins: CCL3L1 and CXCL12 (Group 2). Cutoff values were 1×10^{-7} and are represented in grey.

Term Id	Term Name	# Proteins	%Occurence	e-value

molecular function	GO:0008009	chemokine activity	2	100,0%	2,90E-08
biological process	GO:0042127	regulation of cell proliferation	2	100,0%	1,80E-06
	GO:0006935	chemotaxis	2	100,0%	1,90E-05
	GO:0006955	immune response	2	100,0%	3,00E-05
	GO:0048523	negative regulation of cellular process	2	100,0%	5,60E-05
	GO:0006950	response to stress	2	100,0%	1,70E-03
cellular component	GO:0005615	extracellular space	2	100,0%	9,60E-06

Table 4. GO Annotations obtained from ProteinON for the following proteins: IFNG, IL4R and KIR3DL1 (Group 3). Cutoff values were 1×10^{-7} and are represented in grey.

	Term Id	Term Name	# Proteins	%Occurence	e-value
molecular function	GO:0004888	transmembrane signaling receptor activity	2	66,7%	2,00E-04
	GO:0005515	protein binding	2	66,7%	3,00E-03
	GO:0003674	molecular_function	3	100,0%	1,00E+00
biological process	GO:0050776	regulation of immune response	3	100,0%	2,00E-10
	GO:0002293	alpha-beta T cell differentiation involved in immune response	2	66,7%	5,30E-09
	GO:0050715	positive regulation of cytokine secretion	2	66,7%	1,30E-08
	GO:0032642	regulation of chemokine production	2	66,7%	1,40E-08
	GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	2	66,7%	4,40E-08
	GO:0002699	positive regulation of immune effector process	2	66,7%	5,50E-08
	GO:0050870	positive regulation of T cell activation	2	66,7%	1,10E-07
	GO:0001819	positive regulation of cytokine production	2	66,7%	1,60E-07
	GO:0019221	cytokine-mediated signaling pathway	2	66,7%	6,40E-07
	GO:0045596	negative regulation of cell differentiation	2	66,7%	9,40E-07
	GO:0045597	positive regulation of cell differentiation	2	66,7%	9,60E-07
	GO:0042127	regulation of cell proliferation	2	66,7%	5,50E-06
	GO:0006952	defense response	2	66,7%	2,40E-04
cellular component	GO:0044459	plasma membrane part	3	100,0%	3,40E-06
	GO:0005615	extracellular space	2	66,7%	2,90E-05
	GO:0005887	integral to plasma membrane	2	66,7%	8,30E-05

In the above tables, we can see Gene Ontology terms count in each of the arranged groups. Terms in grey are below the defined cut-off value (1×10^{-7}). These results show that the arrangement was beneficial since the most related genes were grouped together. We can observe that the in the corepressors group (Table 2), the gene functions were related to chemokine

receptor activity that is a protein binding and that it is inflammatory response and related with calcium-mediated signaling. This confirms the information provided in the introduction.

The proteins that interact with the receptors (Table 3) display more homogenous results and its function is related with chemokine activity. The findings to the third group (Table 4) reveals that they are related to the regulation of immunologic processes which also confirms the information presented before.

Table 5. Selected genes and their interactors.

Genes	Interactors
CCR5	AFP, ARRB1, CCL3L1, CCL3, CCL5, CD4, CST9L, CTBP2, CXCR4, ETV5, IL24, JAK2, MIP-1-α, MIP-1-β, PIK3R2, PSMA5, PTK2, RANTES, STAT3, STAT5B
CCR2	POTEE
CXCR4	ADRBK2, ATP13A2, B2M, B4GAT1, CAV1, CCDC107, CCR5, CD4, CD79B, CXCL12, CXCR5, ENV, GCNT3, GNA13, GOLT1B, HLA-B, IPPK, ITCH, JAK2, JAK3, KCNK1, LPAR1, MYBL2, NEF, NT5E, NTRK3, P4HB, PTK2, PTPN6, PTPN11, SOCS3, ST13, STAM, TMEM63B, TMEM171, USP14, VIPR2
CXCR1	CDK1, GNAI2, PDCD6IP, TUBA1A, UBE2C
CCL3L1	CCR1, CCR3, CCR5
CXCL12	CXCR4, DAZAP2, FN1, PF4, VCAN
IFNG	GOPC, IFNGR1, IFNGR2, STAT6
KIR3DL1	ENV, FAM114A2, KIR2DL2, KIR2DS2, RPS6KB1
IL4R	ACVR2B, ANKLE2, ARIH2, ARL5B, BMPR1A, CD40, DCAKD, DHRS3, ERBB2, GNB2L1, GOLGA2, HGS, IL2RG, IL4, INPP5D, JAK1, JAK2, JAK3, KAT5, KCNT2, NCF1, NET1, PKMYT1, PLEKHH3, PRR11, PTPN6, PTPN11, RELT, RHBDF2, RHOBTB3, SHC1, SOCS5, STAT6, STX5, TNFRSF10A, TYK2, UTP6, ZDHHC9

Table 6. Genetic polymorphism with effects on HIV infection or AIDS progression phenotype.

Gene	Variation Name	Change	HIV/AIDS effect
CCR5	CCR5-del32	32-bp deletion in the CCR5	blocks HIV
	VAR_011841	R → S	HIV susceptibility
	VAR_012481	C → R	HIV resistance
CCR2	VAR_014339	V → I (64)	HIV resistance / delayed progression

The two previous tables summarize some keys aspects of this work: interactions between proteins and genetic variation that might affect the disease outcome. Table 6 results are important because if CCR5 is an HIV entry point, any change on that receptor might prevent HIV entrance (assuming that CCR5 was the only entry point), and this is what has been observed in different studies. Genetic variation in any of the coreceptors from group 1 will be important because it might compromise HIV entry and AIDS development.

Literature also confirm these findings as it was expected since the beginning of the work was based on a literature review to find and select the most relevant sequences. The main difficulties on this work was to understand the differences between AIDS and HIV infection, since those are two related diseases and often mistaken as the same thing. Even in the literature research it was hard to find papers solely on AIDS, because most were related to the opportunistic infections that rise after the immune system has been compromised. Literature confirm my

findings because my work was based on literature review provided in OMIM database, and not only on the summary provided in OMIM or UniProt.

According to a recent study (Liu *et al*., 2015), CCR5 and CXCR4 are essential for the HIV-1 infection since double negative cells (whose genes have been knocked-out) don't display an infected phenotype, which confirms the importance of the two receptors as the main HIV entry point. No further considerations about the literature will be made on this report since it is already longer than what was supposed to.

This work establishes some biological sequences related to AIDS and HIV infection, that could be potential therapy's targets. Most difficulties arise from not finding consistent or relevant information at the repositories (the displayed proteins wouldn't be my choice as relevant proteins); but also, because it was hard to separate between AIDS and HIV infection. Also, unlike cancer these disease is not so directly linked to genetic markets, but is caused by a virus, that can rapidly evolve and adapt to new entry targets, and there can be different strains of the virus, which difficult research process because a single disease can be related to multiple factors or genetic markets. Despite all that, this theme is also very rich in immunology concepts, which was hard for me to grasp since I lack some knowledge in that area.

References

- Bhatti, A. B., Usman, M., & Kandi, V. (2016). Current scenario of HIV/AIDS, treatment options, and major challenges with compliance to antiretroviral therapy. *Cureus*, *8*(3).
- Khalili, K., White, M. K., & Jacobson, J. M. (2017). Novel AIDS therapies based on gene editing. *Cellular and Molecular Life Sciences*, 1-12.
- Liu, C., Ma, X., Liu, B., Chen, C., & Zhang, H. (2015). HIV-1 functional cure: will the dream come true?. *BMC medicine*, *13*(1), 284.
- Mehlotra, R. K., Zimmerman, P. A., & Weinberg, A. (2016). Defensin gene variation and HIV/AIDS: a comprehensive perspective needed. *Journal of leukocyte biology*, *99*(5), 687-692.
- Miles, B., Miller, S. M., & Connick, E. (2016). CD4 T Follicular Helper and Regulatory Cell Dynamics and Function in HIV Infection. *Frontiers in Immunology*, *7*.
- Moukambi, F., Rodrigues, V., Fortier, Y., Rabezanahary, H., Borde, C., Krust, B., Estaquier, J. (2017). CD4 T follicular helper cells and HIV infection: friends or enemies?. *Frontiers in Immunology*, *8*.
- Mühle M, Lehmann M, Hoffmann K, Stern D, Kroniger T, et al. (2017) Antigenic and immunosuppressive properties of a trimeric recombinant transmembrane envelope protein gp41 of HIV-1. PLOS ONE 12(3): e0173454. doi: 10.1371/journal.pone.0173454
- Policicchio, B. B., Pandrea, I., & Apetrei, C. (2016). Animal models for HIV cure research. *Frontiers in immunology*, *7*.
- Qureshi, H., & Bhattacharya, J. (2017). Dual Immunity Concomitantly Suppresses HIV-1 Progression. *Trends in Microbiology*.
- Utay, N. S., & Hunt, P. W. (2016). Role of immune activation in progression to AIDS. *Current Opinion in HIV and AIDS*, *11*(2), 131-137.
- Valverde-Villegas, J. M., Matte, M. C. C., Medeiros, R. M. D., & Chies, J. A. B. (2015). New insights about Treg and Th17 cells in HIV infection and disease progression. *Journal of immunology research*, *2015*.
- Younas, M., Psomas, C., Reynes, J., & Corbeau, P. (2016). Immune activation in the course of HIV-1 infection: causes, phenotypes and persistence under therapy. *HIV medicine*, *17*(2), 89-105.