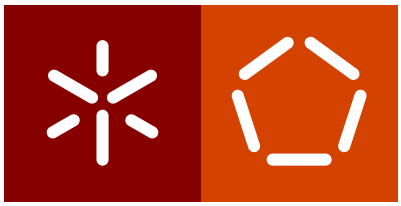
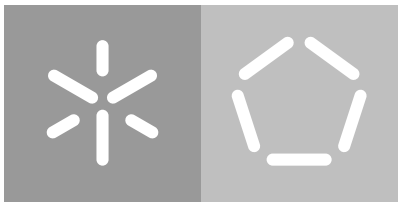


Universidade do Minho
Escola de Engenharia
Departamento de Informática

Daniel Carpinteiro Soares

In silico investigation of drugs for the treatment of dengue fever disease

July 2019



Universidade do Minho
Escola de Engenharia
Departamento de Informática

Daniel Carpinteiro Soares

In silico investigation of drugs for the treatment of dengue fever disease

Master dissertation
Master Degree in Bioinformatics

Dissertation supervised by
Luísa Pereira (Supervisor)
Miguel Rocha (Cosupervisor)

July 2019

ACKNOWLEDGEMENTS

I would like to begin by thanking everyone at the Genetic Diversity group at i3S Porto for receiving me in such a friendly way, you made my adaptation to living in Porto and working in a research group much easier. I'm sure there are few groups with such a good working dynamic such as ours.

A special thanks goes to Bruno Cavadas and Ana Magalhães. Bruno for all the patience and help developing my work, in two years of master's you were the person with whom I learnt more about bioinformatics. Ana for all the help, knowledge and dull work spent reviewing the more pharmaceutical part of my thesis.

A huge thanks to Luísa Pereira, for receiving me in your group and trusting in me even though you didn't know me. Thank you for all the guidance, not only in my thesis but in everything I needed, patience, joy and most of all the friendship. It was a pleasure working with you in such a stimulating research environment, I really learned a lot in this year.

I would also like to thank to everyone at Minho University, not only the friends I did there (Inês, Sofia and Ivan) but also all the professors I worked with, for all the knowledge and support throughout these two years. A special thanks to Miguel Rocha for co-supervising my thesis.

I couldn't finish without thanking my friends. Thanks to all my friends from Aveiro, for the friendship and happiness in all those Saturday night coffees, trips and parties. You accompany me since we all had baby teeth, I hope you guys stick around until we have no teeth at all. Thank you to my Biomedical Sciences friends, Sara, Zé, António, Brutt, Lua, Fonseca, Maia, Brás, Bia, Ferraz, Wilson, Santana, ... I could go on, but I wouldn't have enough space. Most of all thank you for all the friendship, beers and laughs, you guys are the best friends I could have.

To my Aveiro gang at i3S, thank you for all the friendship, lunches, coffees and support, it was much easier doing our theses with you guys around.

Thank you to my Rua Fernandes Tomás housemates, Lili and Zé, for all the crazy ideas, laughs, burnt food, unwashed dishes and sunset beers at Virtudes, living with you guys has truly been an amazing experience.

Finalmente gostava de agradecer aos meus pais e ao meu irmão. Aos meus pais por todo o apoio e esforço emocional e financeiro (principalmente este!) que depositaram em mim. Ao meu irmão um obrigado por ter sido sempre a definição de irmão mais velho (embora eu seja melhor em todos os aspectos!). Acima de tudo muito obrigado por me terem dado um tipo de educação que nenhuma universidade me podia ter dado.

ABSTRACT

The availability of *omics* data, being transcriptomics, proteomics or pathogen-host interactomics, is being explored for *in silico* testing of drugs. This *a priori* testing gives useful insights, avoiding wasting of time and money in the functional evaluation of a high amount of drugs.

In this work, we applied such a strategy to the dengue fever disease, which is caused by a virus transmitted to humans by mosquitoes. Published *omics* in dengue patient samples are available for total blood or specific blood cells, but no information was obtained for other tissues, which display symptoms in dengue disease and/or where the virus multiplies, such as liver, spleen and encephalon. Our reasoning was that an effective treatment of dengue disease should take into account these important tissues. So, we began by obtaining the complete human protein-coding gene expression profiles in 22 samples from liver, spleen and encephalon tissues from five Cuban patients that died due to dengue disease and from three deceased controls due to other causes. We applied a workflow to process the FASTAQ files, tested between local and global alignment, and after quality control, ended up with 18 samples (three dengue patients versus three controls for each tissue). We then estimated the differentially expressed genes, used gene set enrichment analysis to investigate which biological pathways were altered in the disease, and inputted the 150 upregulated and 150 downregulated genes into the CMap tool to investigate possible effective drugs in the three tissues and when comparing with published blood datasets.

As expected, due to tissue specialization, different pathways were altered in the different tissues: in blood, pathways related with the immune system, response to pathogens and cell cycle/repair mechanisms; in liver and spleen, the biggest differences were in metabolism, especially of lipids; in brain, alterations were mainly response to infection and changes related with neurotransmitters/cerebral diseases. CMap results, accordingly, indicated mainly antineoplastic drugs as potentially active in blood, anti-inflammatory and related with metabolism in liver and spleen, and anti-inflammatory and interfering with receptors in encephalon. Although these mechanisms seem to be dissimilar, there are several drugs that overlap in different parts of the complex networks and should be careful evaluated in treating dengue disease. The workflow used in this work is publicly available at GitHub (https://github.com/danielcarpsoares/Drug-Repurposing-Dengue) so that it can be used in other tissues or other diseases.

RESUMO

A vasta quantidade de dados *omic* disponíveis, sejam de transcriptómica, proteómica ou interatómica agente patogénico-hospedeiro, tem vindo a ser explorada com o intuito de testar fármacos *in silico*. Este método permite reduzir, *a priori*, a quantidade de recursos despendidos no teste da eficácia de um fármaco numa determinada doença.

Neste trabalho, aplicamos esta estratégia à Febre da Dengue, uma doença viral que é transmitida por mosquitos a humanos. Atualmente existem dados de transcriptómica disponíveis para esta doença em amostras de sangue, mas não para outros tecidos onde o vírus se replica e/ou apresentam sintomas, como é o caso do fígado, baço e encéfalo. A nossa motivação é a de que um tratamento eficaz desta doença deveria ter em conta todos estes tecidos. Desta forma, inicialmente foi obtida a expressão completa de todos os genes que codificam proteínas em 22 amostras de fígado, baço e encéfalo recolhidas de cinco Cubanos que morreram devido à Dengue e de três controlos que morreram por outras causas. Uma *pipeline* foi aplicada aos dados da sequenciação destes tecidos, nomeadamente aos ficheiros FASTAQ obtidos, seguida de comparação da qualidade entre alinhamento local ou global e controlo de qualidade das amostras. No final, 18 amostras (três pacientes de Dengue e três controlos para cada tecido) foram analisadas quanto aos genes diferencialmente expressos, vias biológicas significativamente alteradas, e potenciais fármacos eficazes a contrariar a Dengue (pela inserção dos 150 genes mais sobre- e sub-expressos na ferramenta CMap) em cada tecido e por comparação com o sangue.

Tal como esperado, devido à especialização dos tecidos, as vias alteradas nos vários tecidos eram variáveis: em sangue verificou-se que estavam mais relacionadas com o sistema imune, resposta a agente patogénicos e ciclo celular; em fígado e baço, as maiores diferenças verificaram-se naquelas relacionadas com o metabolismo, principalmente de lípidos; em encéfalo, verificou-se maioritariamente resposta à infeção, e relacionados com neurotransmissores e doenças neurológicas. Em concordância, os resultados do CMap, indicaram diferentes potenciais fármacos benéficos: antineoplásicos em sangue; anti-inflamatórios e relacionados com metabolismo em fígado e baço; e anti-inflamatório e relacionados com recetores em encéfalo. Embora estes mecanismos de ação aparentem ser muito diferentes, alguns fármacos podem ser simultaneamente efetivos em diversos pontos das complexas redes biológicas e deveriam, por isso, ser avaliados como potencialmente interessantes no tratamento da Dengue. A *pipeline*, assim como todo o código utilizado está publicamente disponível na plataforma GitHub (https://github.com/danielcarpsoares/Drug-Repurposing-Dengue).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

A

AKT Protein Kinase B.

ALS Aspartate Aminotransferase.

ALT Alanine Aminotransferase.

B

BP Biological Process.

BWA Burrows-Wheeler Alignment.

C

CC Cellular Components.

CDC Centers for Disease Control and Prevention.

D

DENV Dengue Virus.

DF Dengue Fever.

DHF Dengue Haemorrhagic Fever.

DSS Dengue Shock Syndrome.

E

EGFR Epidermal Growth Factor Receptor.

F

FDA Food and Drug Administration.

FDR False Discovery Rate.

FLT3 FMS-like Tyrosine Kinase.

G

GEO Gene Expression Omnibus.

GO Gene Ontology.

GSEA Gene Set Enrichment Analysis.

I

IEDB Immune Epitope Database and Analysis Resource.

K

KEGG Kyoto Encyclopedia of Genes and Genomes.

L

LDL Low Density Lipoprotein.

LXR-RXR Liver X Receptor-Retinoid X Receptor.

M

MF Molecular Function.

MHC Major Histocompatibility Complex.

N

NES Normalized Enrichment Scores.

NGS Next Generation Sequencing.

P

PAHO Pan American Health Organization.

PBMCS Peripheral Blood Mononuclear Cells.

PI3K Phosphoinositide 3-Kinase.

Q

QSAR Quantitative Structure Activity Relationship.

R

RCP Representative Concentration Pathway.

S

SEARO South East Asia Region.

W

WHO World Health Organization.

WPRO Western Pacific Region.

1

STATE OF THE ART

1.1 DENGUE FEVER DISEASE

1.1.1 *Virus, Vector and Transmission*

Dengue Virus (DENV) belongs to the family of viruses Flaviviridae, genus *Flavivirus*. This virus has four distinguishable serotypes, DENV-1, DENV-2, DENV-3 and DENV-4. All of these four serotypes share epidemiological similarities, being responsible for the development of Dengue disease after infecting humans, but are genetically distinguishable (Chen and Vasilakis, 2011). Dengue is, according to the *World Health Organization (WHO)*, the most rapidly spreading mosquito-borne viral disease in the world (Stehman et al., 2002), being transmitted to humans through the bite of an infected female mosquito.

The main and most effective vector in Dengue transmission is the mosquito of the species *Aedes aegypti*. This species is also responsible for spreading other arboviruses like the Yellow Fever, Zika and Chikungunya viruses (Kraemer et al., 2015). Although this mosquito is thought to be original from Africa it has spread through many other regions of the globe, like Asia, Australia, the Americas and the South Pacific (Higa, 2011). Though less effective and with a limited ability to transmit DENV when compared to *Aedes aegypti*, other species among the *Aedes* genus like *Aedes albopictus* are also important vectors, responsible for the transmission of dengue virus to humans. *Aedes albopictus* is original from Asia and particularly common in countries like Japan and China and is mainly found in urban areas with nearby green spaces like gardens or parks, where they reproduce and feed on humans, transmitting the virus (Higa, 2011).

Both species optimally fixate in areas with high urbanization and population density, bad sewers and places nearby reservoirs of stagnant water combined with green spaces like parks (Higa, 2011). Besides these factors, temperature and humidity are also key in the ability of the mosquito to survive and reproduce.

The cycle of virus transmission occurs when a DENV infected mosquito bites a previously non-infected human, and when a previously non-infected mosquito bites an infected human becoming able to spread the virus. Because of this and since there is no intermedi-

ate host between mosquitoes and humans nor another mechanism of transmitting DENV, worldwide occurrence of Dengue cases is a direct cause of the global distribution of *Aedes aegypti* and *Aedes albopictus* mosquitoes (Higa, 2011). This can be seen in the Figures below where after overlapping the global distribution of these two mosquitoes (Figures 1 and 2) results in the Dengue risk areas worldwide (Figure 3).

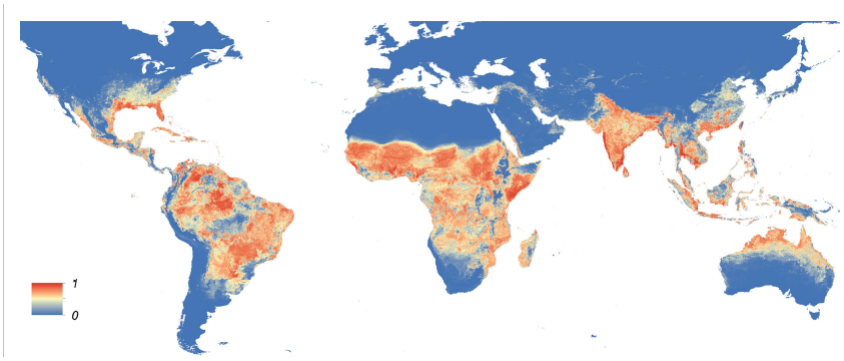


Figure 1.: Global distribution of *Aedes aegypti*, adapted from (Kraemer et al., 2015).

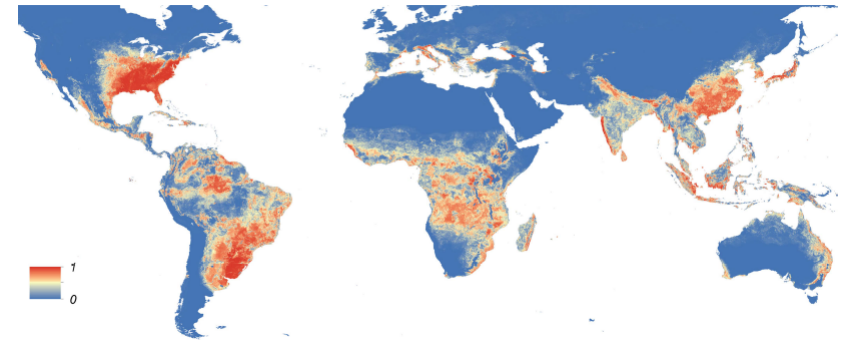


Figure 2.: Global distribution of *Aedes albopictus*, adapted from (Kraemer et al., 2015).

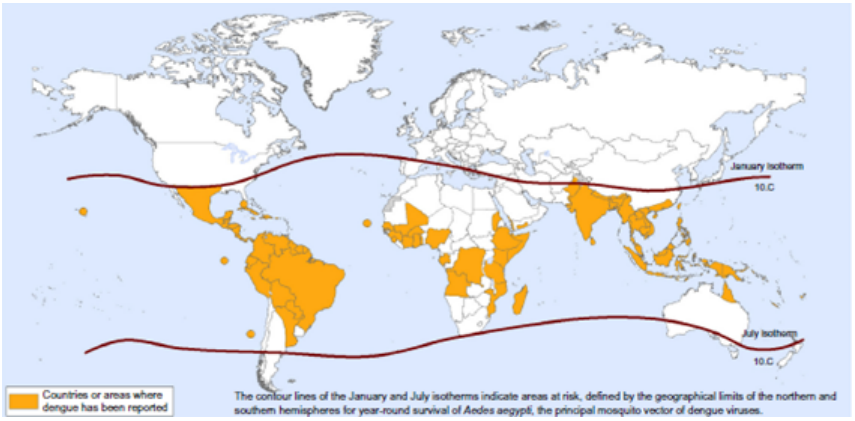


Figure 3.: Areas of Dengue disease risk, adapted from (Wilder-Smith et al., 2013).

1.1.2 *Epidemiology*

Dengue is an infectious disease present in tropical and subtropical regions of the world. Because of this it is estimated that up to 3.6 billion people (about 40% of the world population) are at risk of getting Dengue infected sometime in their lives, making it the most common mosquito-borne viral disease (Brady et al., 2012). Although Dengue was only reported as endemic in nine countries before 1970, according to WHO, nowadays it is considered as such in more than 100 countries throughout African and American continents, Eastern Mediterranean, South-East Asia and Western Pacific regions. Of these regions, *Pan American Health Organization (PAHO)*, *South East Asia Region (SEARO)* and *Western Pacific Region (WPRO)* are the most seriously affected (Figure 4).

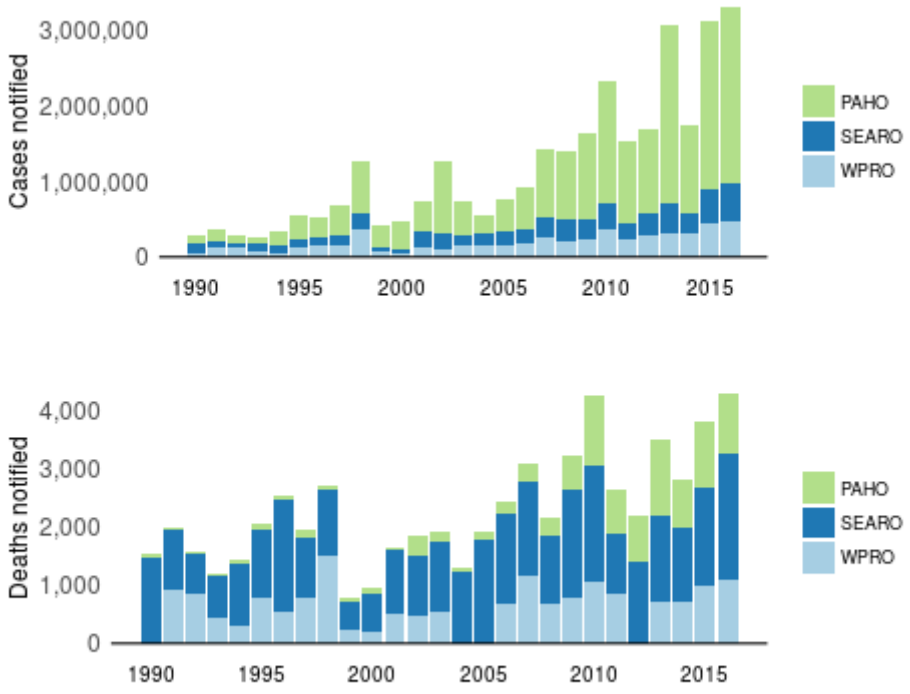


Figure 4.: Cases and deaths in Pan American Health Organization (PAHO), South East Asia Region (SEARO) and Western Pacific Region (WPRO), adapted from (World Health Organization).

Epidemiologic studies regarding this disease are difficult since there are some limitations in diagnostic methods and most Dengue cases happen in very poor countries with a poor health system, which leads to misdiagnosis or even no diagnosis at all. Misdiagnosis happens because diagnosis of Dengue based on clinical symptoms is not trustworthy and there is a need of confirmatory laboratory diagnostic tests. These laboratory tests are not widely

used in the regions where Dengue happens and, additionally, there is a high similarity of symptoms between Dengue and other viral diseases, being often misinterpreted as common flu or Yellow Fever cases, which are also common in these zones (Atkins et al., 2004).

Besides these factors, there is a deficiency in reporting known cases to organizations like WHO or the *Centers for Disease Control and Prevention (CDC)*, which keep track of these situations. Despite this, it is estimated that Dengue has an incidence of 50-200 million infections per year and more than 20,000 deaths occur every year as a direct cause of this disease (Brady et al., 2012). Of all the regions referred above, Southeast Asia is the zone of the globe that suffers the most with Dengue, with children being the most affected by it, as the adults that did not die from it acquired resistance (Hussain et al., 2015).

Further studies also show the heavy socio-economic burden Dengue brings to countries, health care systems and people where this disease is endemic (Gubler, 2012). Estimations show that yearly Dengue costs go around 46.45 million dollars in Latin America and 950 million dollars in Southeast Asia (Halasa et al., 2012). Individually, each hospitalized patient is thought to cost about 1394 dollars (Suaya et al., 2009).

1.1.3 *Symptoms and Presentation*

After being exposed to the virus through a vector, patients can either go asymptomatic (about 75% of the cases) or develop symptoms that range from a flu to more severe cases in which coagulopathy happens. Later cases where the disease does not go asymptomatic usually happen after an incubation period that lasts between 4 to 10 days after contact with the virus (Sheperd, 2017).

The milder form of the disease in which only a flu-like syndrome occurs is known as *Dengue Fever (DF)*. DF is characterized by a rapid onset of fever, which is accompanied by severe headaches, myalgias, arthralgias and gastrointestinal discomfort. Although rare, haemorrhagic manifestations might happen. Typically, these symptoms last between 7 and 14 days after which homeostasis is normally restored (Diamond and Pierson, 2015).

The more severe cases are characterized by the symptoms previously referred in DF combined with coagulopathy, spontaneous bleeding, low to moderate liver injury and increased vascular fragility and permeability (Diamond and Pierson, 2015). These cases are known as DHF.

DHF can eventually progress to an even more severe case, known as *Dengue Shock Syndrome (DSS)*. Besides all the other symptoms already referred, DSS is characterized by rapid fluid loss that leads to severe hypotension and haemorrhagic episodes, mainly bleedings in the skin and the gastrointestinal tract. These symptoms can sometimes be deadly (Martina et al., 2009).

Several organs like liver, spleen and encephalon have known altered functions after Dengue onset. Although liver function and its involvement in Dengue pathogenesis is still not completely known, this is one of the main organs whose metabolism is known to be altered and compromised after onset of Dengue disease. DENV affects the normal functions and pathways of hepatocytes, leading to auto-immune response on liver cells, circulatory compromise, hypoxia and hypotension caused by vascular leakage (Trung et al., 2010). Histological changes comprise hepatocyte necrosis, alterations in fat metabolism, hyperplasia, death of Kupffer cells, Councilman Bodies and infiltration of cells at the portal tract. Hepatic dysfunction happens as a consequence of this infection and usually presents itself with hepatomegaly and a rise, although it can be slight, of transaminase levels (Samanta and Sharma, 2015).

Clinical findings show an increase in *Aspartate Aminotransferase (ALS)* and *Alanine Aminotransferase (ALT)*, being both correlated with Dengue severity, especially with the amount and occurrence of bleeding events (Samanta and Sharma, 2015; Trung et al., 2010). However, some doubts arise from the high values of both ALS and ALT and its direct relation with liver disfunction. Even though it is known that the liver can be the main source of ALT and ALS, it is theorized that both substances can come from a different tissue/organ. Although in reduced quantities when comparing to liver, ALT is also found in cardiac and skeletal muscle and ALS can be found in erythrocytes, cardiac and skeletal muscle, kidneys and brain. More evidence of alteration in these tissues comes when looking at Dengue symptoms, mainly regarding musculoskeletal injury and pain (Samanta and Sharma, 2015; Trung et al., 2010).

Encephalon alterations due to DENV infection result as a consequence of the virus ability to cross the blood-brain barrier and enter the central nervous system. The main manifestations that occur as a result of this phenomenon are Encephalitis, Encephalopathy and less commonly Meningitis. Besides this, and combined with the alterations in the circulatory system caused by DENV, stroke episodes in patients are also common (Gupta et al., 2015).

As for spleen alterations, although the exact mechanism through which it occurs is not precisely known, cases of sudden spleen rupture have been reported. Despite being more common in the early acute phase of the disease, these cases have also been reported in patients at the convalescent phase. If not treated immediately through surgery these cases can be fatal (De Silva and Gunasekera, 2015).

1.1.4 *Physiopathology and Risk Factors*

Although the exact mechanisms by which DENV infects human cells and develops into the sometimes deadly disease are still not completely known, there are some hypotheses that

try to explain how this occurs and how the disease has such different ways of manifesting as described above.

One of the observations already made is that when a patient is infected for the first time with DENV, he becomes immunized against the serotype he was infected with and usually develops a milder form of the disease (DF). After this primary infection, if a second infection by a different serotype happens, then the patient usually develops the disease with a more aggressive spectre of symptoms (DHF or even DSS) (Martina et al., 2009). This is thought to happen due to a phenomenon known as cross-reactive T-cell response. In this situation, instead of granting protection against the viral agent, memory T-cells cause the immunopathology (Farrell, 2013). It is believed that this happens after the first infection with DENV, at the convalescent phase, when there is the creation of memory T-cells for the specific DENV-serotype the individual was primarily infected with. When the second infection with a different DENV serotype happens the previously formed memory T-cells are activated, being this known as cross-reactive T-cell response. The newly activated memory T-cells, ineffective in combating the infection with the new serotype, are responsible for the production of high concentrations of inflammatory cytokines like IFN-γ, TFN but low concentration of anti-inflammatory ones like IL-10 and IL-13. This explains the symptoms experienced in extreme cases of Dengue infection, being responsible for the exaggerated immunopathology referred above (Martina et al., 2009).

Besides cross-reactive T-cell response, other factors have been indicated to explain the different ways the disease manifests and its diverse clinical severity. Virulence is related to the strain of DENV (Martina et al., 2009; Sessions et al., 2013), as exemplified in the 1981 outbreak in South America, when disease evolution and outcome was distinct between the less virulent and dangerous DENV-2 indigenous genotype and the more aggressive DENV-2 Southeast Asian genotype (Martina et al., 2009). Other studies showed that age (being more dangerous in children than adults) (Hoang et al., 2010), the serotype sequence of infection (primary infection with DENV-1 followed by DENV-2 or DENV-3 serotypes are more dangerous) and the time span between the first and second infection (longer time between infections translates into more dangerous outcomes), influence the way the disease develops (Kwissa et al., 2014).

Another very important factor affecting disease outcome is human host genetic diversity. The presence of certain gene *loci* have been linked with a higher propensity to develop more aggressive manifestations of Dengue. Some of these *loci* are located at genes related with the immunologic system, more specifically, HLA-I and HLA-II, TNF-α, Fc receptor, TAP and DC-SIGN (Cahill et al., 2018). Additionally, comparisons between DF and DSS phenotypes considering the ancestry of the patients were made. Protective and susceptible haplotypes were identified in genes *MICB*, *PLCE1*, *PLCB4*, *CHST10*, *AHRR*, *PPP2R5E* and *GRIP1*. Results show that the populations with the higher risk of developing both the

more severe and milder Dengue phenotypes are the Northeast and Southeast Asian populations. Regarding DSS phenotype, Europeans and populations from North-America also have a greater genetic predisposition to develop it. Contrastingly Africans have higher genetic protection against DF and DSS (Sierra et al., 2017), whereas European and American populations only have protection against DF (Oliveira et al., 2018).

Interestingly, a genome-wide association study demonstrated that African ancestry is protective against Dengue showing that differentially expressed haplotypes between populations in the genes *OSBPL10* and *RXRA* are favourable in the way the African population responds to the disease. These genes play an important role in the *Liver X Receptor-Retinoid X Receptor (LXR-RXR)* activation pathway that is linked with lipid metabolism and immune function, more specifically cytokine production, viral replication and their entrance in cells (Sierra et al., 2017).

Retinoids are a class of lipids, normally in low body concentrations, that are linked with important organism functions such as vision, embryonic development, gene expression, tissue differentiation, immune function and activation of tumour suppressor genes (Kiser et al., 2014). Despite this, when in higher concentrations, specially of vitamin A, they are dangerous for the body, being prooxidant, cytotoxic, mutagenic and teratogenic (Mawson, 2013). Dengue infection has been associated with higher levels of these substances by a mechanism in which the damaging effects of the virus cause death of hepatocytes and consequentially the emission of retinoids (Mawson, 2013). It is proposed that this is also responsible for viral replication and disease pathogenesis. More evidence for this arises when comparing the similar symptoms of Dengue and retinoid intoxication.

1.1.5 *Increasing Risk in Dengue Infection*

Most recent statistics show not only an increase in the number of Dengue cases reported in the last years, but also an increase on the propensity one might get infected in the future. Hereafter it is expected an increase not only in the global incidence of Dengue cases worldwide but also in its geographical expansion, which will lead to an escalation in the resulting problems of the disease. It is important to state that Dengue transmission is mainly affected by the spatial distribution of the vectors *Aedes aegypti* and *Aedes albopictus*. Factors responsible for this distribution are temperature, rainfall, highly densely populated zones and bad sewer systems (Kraemer et al., 2015). The main reasons for these previsions are:

- The increase in global temperatures due to climatic changes. Higher temperatures allow a more suitable environment for the vector, increasing its survival capability. On the other hand, increase in global temperatures will make previous non-endemic regions

into endemic ones, since it will allow the migration of the *Aedes* mosquitoes outside of the tropics into new suitable regions (Liu-Helmersson et al., 2016; Wilder-Smith et al., 2013).

- The increase of more virulent strains of DENV in humans and mosquitoes when comparing to less virulent ones. According to epidemiological and phylogenetic studies, overall percentage of more aggressive strains, with an higher capability of infecting hosts, is increasing when compared to milder ones (Wilder-Smith et al., 2013).

- Globalization and the increase of worldwide travels. The increase of travelling not only into but also outside of endemic regions is a risk factor regarding the increase of Dengue cases since it enables the migration of vectors outside of their natural habitat, reaching non-endemic zones. Additionally, it increases the number of people in contact with the vector in endemic zones (Wilder-Smith et al., 2013).

- Socioeconomical factors, such as the increase and rapid growth of highly densely populated urban areas, as well as the economical gaps between countries, mark the difference on how countries manage the incidence of the disease, despite the identical environmental factors shared within the endemic zone. The most important factors that help and grant the ideal habitat for vector survival and growth in these zones are: high human population density and urbanization, bad health systems, poor sanitation and the presence of water reservoirs with stagnant water (Wilder-Smith et al., 2013).

Recent outbreaks in current non-endemic zones, like the one that happened in 2012 in Madeira island, Portugal, and in other zones of South Europe show the reality and danger of these predictions (Liu-Helmersson et al., 2016).

Figure 5 illustrates vectoral capacity for *Aedes aegypti* in different cities of Europe in the past, present and compares it with future predictions. *Representative Concentration Pathway (RCP)* is a projection of our future climate considering greenhouse gas emission. Higher values of RCP mean more extreme climates with a higher degree of climate changes. As it can be seen in the image, predictions show an extreme increase in vectoral capacity for *Aedes aegypti*. This will probably reflect in Dengue disease to emerge in Europe, mainly in South Europe, a posterior non-endemic zone for this disease.

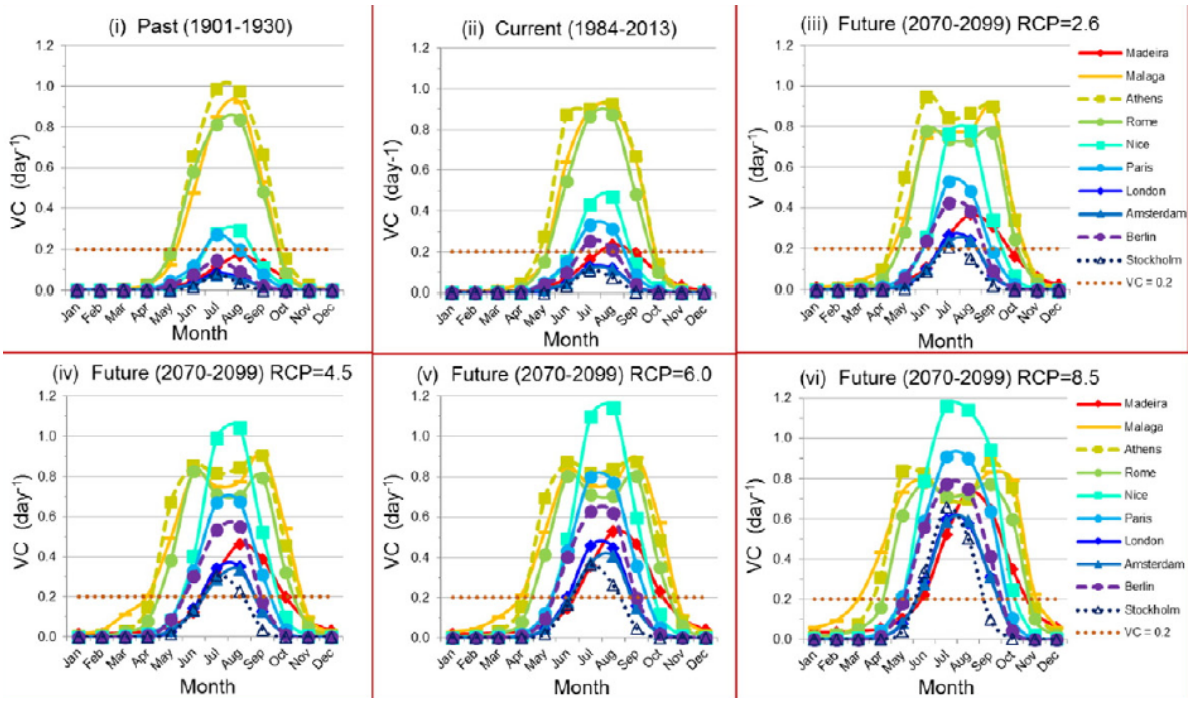


Figure 5.: Past and present occurrence of Dengue, and future predictions for the disease in Europe considering Representative Concentration Pathway (RCP), adapted from (Liu-Helmersson et al., 2016).

1.1.6 *Antiviral Drugs and Vaccines*

Due to the lack of knowledge on the complex physiopathology regarding Dengue infection there are still no antiviral drugs to treat it, neither an effective preventive vaccine. Many efforts have been made to change this since much of the worldwide population has already been or is at risk of being infected.

Regarding vaccines, two big challenges arise. Firstly there is no good animal model that mimics the way Dengue infects and develops in humans, making the study of the pathogenic mechanisms more difficult and restricted to *in vitro* studies and field patient observation. On the other hand, evidence shows that being already immune to one of Dengue's serotypes increases the risk of developing Haemorrhagic Fever and more aggressive phenotypes of the disease when infected by another of the four different serotypes. This way, and in order to develop safe and effective vaccines, they must grant immunization to all Dengue serotypes at once. To date there are two different vaccines with the characteristics previously referenced: one by GlaxoSmithKline (already on clinical trials) and another by Sanofi Pasteur (already introduced in some countries) (Rajapakse et al., 2012). However, recent fatalities of children with severe dengue in Philippines have halted the vaccination with the Sanofi Pasteur vaccine. Because of this, and according to WHO, further tests are

needed to assess this vaccine safety (Iacobucci, 2018). More recently (May, 2019), *Food and Drug Administration (FDA)* approved the reintroduction of this vaccine. Although this is an important step in combating Dengue disease, there are still some limitations on this vaccine, since it will only be available for those in endemic areas with ages between nine and 16 years old and who have been laboratory confirmed previously infected.

Concerning Dengue treatment, there is still no actual specific antiviral medication for this infection. Instead, infected patients receive treatment to handle and try to minimize the symptoms (Rajapakse et al., 2012). This way patients experiencing symptoms usually receive antipyretics to reduce pain and fever, which is one of the main problems common to all patients infected with Dengue, especially in the early stage of the disease (Rajapakse et al., 2012).

In more serious cases of patients who developed Haemorrhagic Fever or even Shock Syndrome, additional treatments are required besides fever and pain control, so that vascular integrity and fluid volume is maintained. Thus, crystalloids and volume expanders are administered. The first are responsible for the maintenance of intravascular volume, blood pressure and normal urine formation. The second are given to patients when there is a deficit of intravascular volume and are responsible for the restauration of intravascular volume, blood pressure and tissue perfusion (Sheperd, 2017).

Although, as previously referred, the immune system plays an important role in Dengue disease physiopathology, treatment with immunomodulators, corticosteroids and other nonsteroidal anti-inflammatory drugs show no results when treating the disease and should even be avoided. Besides this, close monitoring of the patient should be maintained during the course of the infection, mainly regarding heart, hepatic and renal function (Rajapakse et al., 2012; Weerakoon et al., 2011).

Statins belong to a group of drugs known as HMG-CoA reductase inhibitors. These drugs are administrated to patients with hyperlipidaemia for their lipid-lowering properties, mainly *Low Density Lipoprotein (LDL)* cholesterol. Additionally, they improve endothelial normal function, which has been hypothesised could bring beneficial effects in the stabilization of Dengue's vasculopathy (Martinez-Gutierrez et al., 2014). *In vitro* studies showed that statins, because of their reduction in Dengue's virion assembly, might have antiviral properties (Martínez-Gutierrez et al., 2011). In an *in vivo* study, mice were infected with Dengue and posteriorly treated with Lovastatin. Results showed that the usage of this drug in infected mice increased their survival rate and decreased their overall DENV viremia (Martinez-Gutierrez et al., 2014). In spite of these findings, clinical trials where statins were administrated in patients infected with Dengue showed that although statin usage was well tolerated by infected patients, they had no beneficial effects in reducing Dengue's symptoms neither significantly reducing viremia (Chia et al., 2018).

1.2 HIGH-THROUGHPUT INFORMATION

1.2.1 *Transcriptomic Analysis*

Transcriptomics is one of the many *omics* technologies allowing, in this case, the profiling of the whole set of RNA transcripts, known as transcriptome. Transcriptomics is very helpful since it allows the quantification of the organism gene expression at a given moment in specific conditions, enabling the comparison of normal *versus* altered tissues, for example (Lowe et al., 2017).

Nowadays, there are two main transcriptomic technologies: microarrays and RNA-Seq. Microarrays use known and pre-synthetized nucleic acid sequences, named as probes, which are attached to an array plate. Following a first step of hybridization between samples and probes, it is possible to perform relative quantification of mRNA transcripts, indicative of gene expression levels. Because of its ability to quantify nucleic acids, microarrays have been widely used in gene expression analysis and genotyping (Bumgarner, 2013).

After the implementation of high-throughput next-generation sequencing technologies there was a major progress in transcriptomics through the development of RNA-Seq, a revolutionary technology which has many advantages over microarrays. RNA-Seq has better quality when quantifying gene expression, enabling not only the quantification of known transcripts but also of novel ones (Kukurba and Montgomery, 2015). Another advantage of RNA-Seq when comparing to microarrays is that this technology has increased quantifying sensitivity for genes expressed at very low concentrations (Wang et al., 2009).

Transcriptomic analysis is very helpful in many diverse fields such as diagnostics and research, more specifically in understanding how the human and pathogen interaction affects metabolism, in annotating gene function and in clarifying the physiopathology of diseases. This kind of technology is especially helpful after quantification, since it enables the extraction, after normalization and statistical analysis of the data, of which genes are differentially expressed (Lowe et al., 2017).

Regarding RNA-Seq, after sample preparation and sequencing is performed, output undergoes a quality control and differential expression workflow (Figure 6). FASTQ file is a common file format used as output for RNA-Seq technologies, similarly used by different companies that perform sequencing, such as Illumina. It consists of a file with a header containing information relative to the sequence ID and its length. After the header, it is presented the sequence *per se*, containing the coding letters for RNA or DNA. Finally comes the quality string which contains a numerical score for the sequencing quality (Cock et al., 2009).

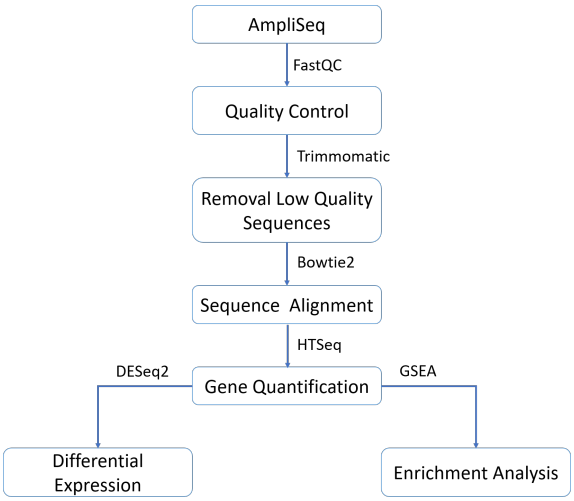


Figure 6.: Workflow for RNA-Seq analysis.

These files are posteriorly used in quality control to assess if they have enough sequencing quality to be used in later analysis. A widely used software to perform this is FASTQC. FASTQC uses the previously referred score present in the quality string of the FASTQ file. Additionally, this software analyses the sequence GC content as well as over or under represented sequences, which can indicate if the samples were contaminated. From the output information a decision whether the samples have enough quality for further analysis or not can be taken (Leggett et al., 2013). Besides these quality control steps, trimming of the sequences is also performed. Trimming consists of removing the adapter or continuously low quality reads that are below a defined threshold. An example of a software that performs this and is Trimmomatic (Bolger et al., 2014).

From here, alignment of the sequences needs to occur. Bowtie and *Burrows-Wheeler Alignment (BWA)* both perform alignment of the reads to a reference sequence. Both algorithms are based on the Burrows-Wheeler Transform, a method for character compression that saves computational resources. They work by building indexes of the reference sequence, that will later serve as guides for the alignment (Langmead, 2011). Alignments can either be global or local, being the main difference between both that local alignments try to match portions of the reads to the reference sequence, whereas in global alignments matching is done for the whole read (Polyanovsky et al., 2011). Finally, Bowtie and BWA give as output a BAM or SAM file.

To conclude, these aligned reads contained in the BAM files need to be quantified to be used in later expression analysis. HTSeq is an open-source Python bioinformatic tool developed to manipulate biological. This tool is designed for performing many analysis such as indexing, variant calling and quantification. Quantification is especially important in this kind of analysis since it enables the association between the reads and the genes.

HTSeq gives as output a txt file with the counts *per* gene across the studied individuals (Anders et al., 2015).

Differentially expressed genes, either overexpressed or underexpressed, are particularly interesting when comparing diseased and healthy tissues, helping to understand the pathology and to identify potential biomarkers, therapeutic targets or gene expression patterns in disease that can be used in diagnostics (Rodriguez-Esteban and Jiang, 2017). Discovery of these differentially expressed genes starts on the file with the quantified reads, and several tools are widely used and available as packages in R Bioconductor. One of these packages, limma, enables the analysis of both RNA-Seq and microarrays, containing tools to read, normalize and obtain differentially expressed genes (Ritchie et al., 2015). Other tools like limma used for this kind of analysis on count data are edgeR and DESeq2, although not being available for microarrays and being both limited for RNA-Seq technologies. These packages allow normalization of the reads, after which statistical tests such as the Wald test are done. This test calculates a p-value for the expression of each read, of which the ones that are statistically significant can be hypothesized as differentially expressed (Love et al., 2014; Robinson et al., 2009).

From this set of interesting genes further analyses are done to try to better comprehend the functional significance of such expression values. This important step in the *omics* field is commonly known as enrichment or pathway analysis. This is done by associating the differentially expressed gene names to their biological annotations and grouping in molecular pathways. Examples of these annotations are the *Gene Ontology (GO)* (Carbon et al., 2017) terms (BP, CC and MF) and KEGG (Kanehisa et al., 2018), which have information regarding gene name, function, or the pathways the gene is a part of. Afterwards, enrichment analysis algorithms identify which annotations are over or under represented in the set of genes in pairwise comparisons (Tipney and Hunter, 2010). One of these softwares is *Gene Set Enrichment Analysis (GSEA)*, a tool that allows enrichment analysis in gene expression data that shares same biological function or regulation (Subramanian et al., 2005). When working with transcriptomics, usually with large quantities of data (approximately 20,000 genes), this kind of procedure narrows the large list of initial genes to a smaller group of genes whose function is probably altered.

1.2.2 *Transcriptomics in Dengue*

Regarding Dengue, several transcriptomic analyses have already been done in different populations, such as Thai and Vietnamese, and considering different phenotypes (healthy, DF, DHF and DSS) (Kwissa et al., 2014; Loke et al., 2010; Long et al., 2015; Nascimento et al., 2009; Popper et al., 2012; Sessions et al., 2013).

Attending to the diverse phenotypes, it was shown that the phenotype groups clustered separately, especially the acute phase patients who had high heterogeneity, proving that there is different gene expression patterns considering disease state (Long et al., 2015).

Another variable studied is how transcriptional profiles behave considering the amount of viral load present on the infected person. Data showed that patients with high viremia had a positive correlation with the NS-1 antigen (antigen used in the diagnostic of Dengue). Besides this, concentration of NS-1 had an inverse correlation with the duration of the acute phase of the disease (Kwissa et al., 2014).

Genes correlated with high viremia and acute phase of Dengue were associated with upregulation of immunological pathways such as sensing of viruses and production of IFN-I. Genes linked with low viremia and a later stage of the disease were mainly associated with pathways regarding cell cycle, proliferation, cell metabolism and translational control, typical of a more normal, healthy cell condition (Long et al., 2015).

Another investigation took into account the infection with different strains of DENV and how it affects transcription. Transcriptomic behaviour was different amongst different DENV serotypes. This demonstrates that cells have different ways of responding to DENV infection, having some common pathways activated among strains but having others that are serotype specific (Sessions et al., 2013).

Studies like these not only help to increase Dengue knowledge but also enable the identification of potential biomarkers that can, in the future, be helpful in developing new diagnostic tests. This is achievable by analysing which genes are differentially expressed between the different stages of the disease (DF, DHF and DSS) and healthy controls. This way, models with specific transcript profiles can be used to predict patient disease state (Nascimento et al., 2009).

Analysis of available literature shows that the study of the transcriptome in Dengue disease cases has always been done from blood samples and mostly using microarrays. As the transcriptome information obtained is always a snapshot of the sample tissue at the precise moment it was collected (Lowe et al., 2017), these studies provide a limited insight into alterations occurring in the body of dengue patients. It would be very informative to analyse other tissues since the transcriptome profile differs with tissues/organs specialization.

1.3 *In silico* investigation of drugs

1.3.1 *Current Methods*

With recent developments in technology and the capability of sequencing biologic data much easier when compared to the past there has been a great increase in available biological data. The mining of such information could revolutionize the way research facilities and

pharmacological companies address the discovery and development of new drugs (Loging et al., 2007). This kind of technology is especially helpful in the initial stages of drug research, being able to narrow the identification of potential genes or pathways the future drug will affect. Since there is a significant reduction of targets, benefits of using such pipelines are the reduction of the time it takes to develop a drug as well as the financial cost associated with it, once there is a decrease of trial-error research. Besides this, such information can be used in personalized medicine, since patients respond differently to disease and this information can help adjust therapy, maximizing its response (Agren et al., 2014).

Examples of *in silico* methods for the investigation of new drugs are: homology modelling, which predicts and generates the 3D structure of the protein of interest based on its sequence; molecular docking, predicts binding affinity between molecules based on molecular interactions; virtual high-throughput screening, is based on the evaluation of compounds and their potential to bind to specific sites inside a molecule of interest; *Quantitative Structure Activity Relationship (QSAR)*, enables the association between structural properties of molecules and their biological activity; conformational analysis, considers the bonding between a molecule and its receptor site and calculates the most favourable conformational energy; *omics* analysis, where through the analysis of transcriptomics, genomics or proteomics it is possible to predict which genes and pathways are altered, giving helpful information when trying to discover new drug targets or biomarkers (Ferrero et al., 2017).

For the purpose of this study *omics* analyses will be focused on. In this field there is a specially interesting software, CMap. It contains genetic expression profiles from tissues that were treated with many different molecules creating the bridge between transcriptomics analysis and *in silico* drug discovery. This tool uses as input the previously determined up or downregulated genes, returning the drugs related with that gene expression pattern. This is performed through data mining and pattern searching algorithms. Because of this, CMap is especially helpful in drug repurposing, using existing drugs as new candidates for diseases with no therapy (Lamb et al., 2006).

Successful examples of combining transcriptomics and drug administration were obtained in cancer. In one of these studies (Uhlen et al., 2017), the analysis of transcriptomics from patients with different types of tumours showed the relation between identified differentially expressed genes, referred to as candidate prognostic genes, and clinical outcome, helping in predicting patient survival as well as in adapting their therapy to have better results. In another study (Agren et al., 2014), specific for hepatocellular cancer, transcriptomics from patients were analysed and with the help of *in silico* algorithms identification of potential new anticancer drugs was possible and specific for each patient, increasing positive clinical outcomes.

1.3.2 *Published Data on Dengue Disease*

Although there is not much literature available regarding *in silico* drug investigation on Dengue disease, there are some studies focused on the discovery of antiviral drugs for this virus.

One of these studies emphasizes on the discovery of new antiviral drugs that target DENV's NS2B/NS3 protease complex, an important enzyme in the virus capability of replication. In order to do this, an *in silico* approach was performed so that small molecules with the ability of inhibiting this complex were retrieved. From the set of molecules only those with high affinity for binding with NS2B/NS3 were kept and after this, five of these molecules, that were commercially available, were studied *in vitro*. When tested *in vitro* with DENV, two of the molecules exhibited an inhibitory effect on DENV's replication, being possible new drug candidates for Dengue disease treatment (Cabarcas-Montalvo et al., 2016).

In another study computational methods were used as a mean to better predict an effective vaccine formula for Dengue disease. With this purpose epitopes (parts of the antigen that are recognized by the immune system) from Dengue were selected using available databases such as the *Immune Epitope Database and Analysis Resource (IEDB)*. After this, affinity between epitopes and the *Major Histocompatibility Complex (MHC)* was predicted using a specific tool from IEDB, from which only those with lower affinity were selected since they are less prone to develop dangerous secondary effects when administrated in the vaccine. Subsequently, epitopes that had identical HLA binding profiles were removed leaving only one epitope *per* HLA. The final proposed vaccine formula is composed of 15 epitopes, having a predicted efficacy of 92.49% (Murphy et al., 2018).

More recently a new study was developed which detected new drug candidates for Dengue Disease. Here a different strategy was implemented where information from transcriptomics, proteomics and protein-protein interactions was used. From the transcriptomic analysis performed in three different datasets with DHF and control individuals, 3.892 significant genes were obtained. These genes were then used in CMap, a tool used to relate gene expression patterns with drugs and diseases, and 85 drug candidates were obtained. Proteomic information was gathered from available literature, returning proteins with significant differences in expression, between infected and normal individuals. Interactions between these proteins and drug candidates was established by STITCH, returning 548 drug candidates. In protein-protein interactions, a search for human and dengue proteins interaction was performed returning 221 proteins. Similarly to what was done in proteomics, these proteins were analysed in STITCH, returning 415 drug candidates. Finally, all drug candidates from the three different analysis were evaluated together, reaching eight

common drug candidates. These drugs thus be potentially repurposed and tested for the treatment of Dengue disease (Amemiya et al., 2019).

2

AIMS

Due to the lack of knowledge on the complex physiopathology regarding Dengue infection there are still no antiviral drugs to treat it. Many efforts have been made to change this since epidemiology shows a steady increase in the incidence of Dengue and much of the worldwide population is either at risk or has already been infected with this virus.

In this work, *in silico* models of the impact of Dengue virus in cell metabolism from diverse tissues were developed in order to identify possible interfering drugs which could be used for treatment of Dengue disease. First, we analysed the transcriptome from liver, encephalon and spleen in deceased patients due to dengue disease versus controls (our own data; no similar datasets available so far in literature), as well as in blood (publicly available data). We checked for significantly differentially expressed genes in the pairwise comparisons and evaluated which molecular pathways were significantly changed. Secondly, we evaluated through Cmap which drugs could interfere with the expression profiles from each of the tissues in the Dengue context.

The workflow used in this work will be available with the R code implemented at GitHub, so that it can be easily applied by other researchers in dengue or other disease contexts.

3

MATERIALS AND METHODS

3.1 TESTING IN BLOOD SAMPLES FROM PUBLISHED DATASETS

3.1.1 *Datasets*

Three datasets from microarrays with Dengue infected patients and healthy controls were retrieved from the database *Gene Expression Omnibus (GEO)* (Barrett et al., 2013), identified by IDs GSE18090 (Nascimento et al., 2009), GSE38246 (Popper et al., 2012) and GSE51808 (Kwissa et al., 2014). GSE18090 and GSE38246 were both collected from *Peripheral Blood Mononuclear Cells (PBMCs)*, while GSE51808 was collected from whole blood samples. GSE18090 is from Brazilian adults, while GSE38246 is from children of Nicaragua and GSE51808 from children and adults of Thailand. The chips used for GSE18090, GSE38246 and GSE51808 were Affymetrix Human Genome U133 Plus 2.0 Array, SMD Print1430 hr1 and Affymetrix HT HG-U133+ PM Array Plate, respectively.

Within each dataset, samples were divided in two groups considering disease status: DHF patients and healthy controls. Individuals with DSS and DF were excluded from further analysis. DFs were removed because they have the milder form of the disease, and are more difficult to identify/diagnose. As for the DSS cases, although this is the most severe form of the disease, only GSE38246 had individuals with this phenotype. This way, it was decided to remove them from the analysis once it was not possible to perform a consistent comparison across the three datasets. Table 1 summarizes the final distribution of DHF and healthy individuals used in the analysis across the three microarrays.

Table 1.: Distribution of Control and DHF individuals across GSE18090, GSE38246 and GSE51808.

	GSE18090	**GSE38246**	**GSE51808**
DHF	**10**	**32**	**10**
Control	**8**	**8**	**9**

Since data was obtained from microarrays, occurrence of multiple probes *per* gene is present. As we need one expression value *per* gene, we used the "genefilter" package from R, more specifically the featureFilter function, which keeps only the probe with the highest expression variance for each gene.

Additionally, for GSE38246 after a first analysis of the data we realized it contained probes specific for viruses, bacteria, miRNA, shRNA or ncRNA. Since the purpose of this study is confined to the analysis of human transcriptomics of genes coding proteins, those probes were removed.

3.1.2 *Differential Expression Evaluation in Array Data*

Gene differential expression was obtained using the pipeline from the "limma" package in Bioconductor R, specific for the analysis of microarray data Ritchie et al. (2015). Firstly the *lmFit* function was used with the purpose of fitting the expression values from each probe of the microarray into a linear model. Then the linear model was used in the *makeContrasts* function which calculates coefficients and standard deviations between two groups (DHF patients and healthy controls). This output is then used in the *eBayes* function, which estimates p-values and logarithmic fold changes for each gene using standard errors towards a common value of expression. Since multiple analyses were performed, p-values were adjusted through the *topTable* function, which applies the Benjamini-Hochberg procedure for *False Discovery Rate (FDR)* to handle multiple tests. The threshold used for admitting a gene as differentially expressed was an adjusted p-value <0.05.

Volcano plots for the expression profiles of the microarrays were created using the R packages "ggplot2" and "ggrepel", giving as inputs the Gene symbol, adjusted p-value and the logarithmic fold change of each gene. The common upregulated and downregulated genes between the three datasets were easily identified in the Venn diagrams built using the "VennDiagram" package of R.

3.2 Testing in other tissue samples from our own datasets

3.2.1 *Biological samples and NGS-expression kit laboratorial processing*

Tissue samples were *post-mortem* collected from Cuban individuals. Both Dengue infected and control samples (accidental death) were retrieved from liver, spleen and encephalon. Individuals were identified with an ID: those infected are 875, 900, 37478, 39538 and 39539 (although 875 and 900 only have spleen and hepatic samples). Controls are identified with F1, F2 and F4.

Gene expression evaluation was done by *Next Generation Sequencing (NGS)* at the i3S Scientific Platform, using the *Ion AmpliSeq Transcriptome Human Gene Expression Kit from Thermo Fisher Scientific* (Waltham, Massachusetts, USA), which was sequenced in the Ion 550™ Chip kit and Ion S5™ XL System (Thermo Fisher Scientific, Waltham, Massachusetts, USA). This kit contains one amplicon per protein-coding gene, which is amplified and sequenced, allowing to infer expression levels of the targeted 20,812 genes. This is not a proper RNA-Seq, which sequences all transcripts present in the cells, implying higher level of raw reads and more expensive processing of samples. As we were only interested in global expression levels and not in identifying alternative transcripts, the AmpliSeq strategy is as informative as the traditional RNA-Seq. The total raw reads allowed in the chip for each sample were around 10.8 million reads.

3.2.2 *AmpliSeq bioinformatic analysis*

From the sequencing, FASTQ files for all samples were obtained. A first quality control check was performed by uploading FASTQ files to the software FASTQC, which retrieves the overall sequence quality scores, the base sequence content, the sequence GC content and the presence of duplicated or overrepresented sequences. From the analysis of these parameters it can be concluded if sequencing went according to plan, if the samples have enough quality for further analysis or if they had contaminants.

Samples that passed this first step of quality control were posteriorly used in trimmomatic to filter those reads that had a low sequencing quality score. FASTQ files were inputted in the software and filtering parameters were set as: sliding window 4:15 (the sequence is scanned in a four nucleotide span and reads with an average sequencing quality score below 15 are removed); and minlen 40 (reads shorter than 40 bases are removed).

Output FASTQ files from trimmomatic were then uploaded to the i3S local server, and global and local alignments were performed on each sample using Bowtie2 with the hg19 human genome as reference. The aligned files have BAM extension. Afterwards, reads per gene were counted and the matrix with the expression values was created in the python package HTSeq. This software uses as input the BAM files from the alignments and matches the reads to the genes in the chip. The numbers in the matrix represent the number of reads of a given gene in a sample, and as the AmpliSeq kit contains one amplicon per gene (same size), the expression normalization by gene length is not required. This file is therefore ready to be used in differential expression.

Pearson correlation coefficients between local and global alignments inside the same tissue were calculated using the "Correl" function from Excel for the control samples. Plots relating the two alignments were also done in Excel using the "Scatterplot" function.

3.2.3 *Differential Expression Evaluation in AmpliSeq Data*

Gene differential expression was estimated for each tissue matrix (resulting from the local alignment) using the pipeline from the "DESeq2" package in Bioconductor, specific for the analysis of RNA-Seq data (Love et al., 2014). Information whether individuals were infected or controls was used as the variable in the *DESeqDataSetFromMatrix* function which creates a *DESeqDataSeq* object. The *estimateSizeFactors* function was then applied to the *DESeqDataSeq* object, which normalizes the data considering the median value of each gene expression across samples.

Afterwards the statistical testing of differential expression occurs. First, DESeq performs automated independent filtering, in which by removing the weakly–expressed genes from the input, more genes can be found to be significant among those kept for analysis, improving the power of the test. The Wald test was applied in the *nbinomWaldTest* function, elucidating if a given variable (in this case, disease state) significantly affects the model (in this case gene expression). Results consist of individual p-values for differential expression for each gene, which are adjusted by using the FDR procedure.

Similarly to what was done in the microarray data, volcano plots for the expression profiles across the three tissues were created using the R packages "ggplot2" and "ggrepel".

3.3 Enrichment Analysis

Enrichment analysis was achieved using GSEA (Subramanian et al., 2005). Initially, expression matrices with all genes and samples for each microarray and tissue were uploaded in GSEA. Gene set enrichment analysis was performed for GO (CC, BP and MF) and the KEGG pathway databases for each of the uploaded datasets. To perform such, default parameters from GSEA were used, except the "Min size: exclude smaller sets" parameter that was reduced from 15 to 10 which will include all annotations that have a minimum of 10 genes.

With the results from the Gene set enrichment analysis a leading edge analysis from GSEA was done with a posterior GSEApreranked. The leading edge analysis enables the observation of the distribution of the genes across the annotations. GSEApreranked runs a gene set enrichment analysis taking into account an ordered list of genes, usually using the p-value for differential expression as a parameter.

3.4 Drug Repurposing

For the drug repurposing step, the CMap Query tool from clue.io (https://clue.io/query) was used. For the blood information, the first 150 (this number is a limitation from the

tool) genes that were significantly up and down regulated (FDR values) between patients and controls were uploaded in the Query tool. Only genes that appeared commonly in at least two of the three datasets were inputted and were ordered accordingly to the highest absolute mean fold change value across the three datasets. In the case of the AmpliSeq data, Query analyses were made independently for each tissue. We inputted the 150 upregulated and downregulated genes ranked by the lowest p-values (even if they were not statistically significant).

The human cell lines that are contained in the Cmap Query tool are: A375 (malignant melanoma), A549 (non-small cell lung carcinoma), HCC515 (non-small cell lung adenocarcinoma), HEPG2 (hepatocellular carcinoma cell line), MCF7 (breast adenocarcinoma), PC3 (prostate adenocarcinoma), VCAP (metastatic prostate cancer), HT29 (colorectal adenocarcinoma), and HA1E (kidney epithelial immortalized).

Cmap query output consists in a list of perturbagens rank-ordered by the similarity of differentially expressed gene sets to the query gene set. A positive score indicates there is similarity between a given perturbagen's signature and that of the query, while a negative score indicates that the two signatures are opposing. So, we considered scores of below -90 as the drugs that could potentially treat dengue disease. The drugs are complemented with information for their molecular mechanism of action. We then classified these drugs into seven main groups of actions: antineoplastic, antibiotic, antiparasitic, immunosuppressant, cardiovascular, anti-inflammatory, antiviral and other. This classification was based on information contained in several databases: Inxight: Drugs (https://drugs.ncats.io/); PubChem (Kim et al., 2019); Drugbank (Wishart et al., 2018); FDA (https://www.fda.gov/); EMA (https://www.ema.europa.eu/en); ClinicalTrials (https://clinicaltrials.gov/). Some molecular mechanisms of actions can be affiliated in more than one main group of action. The bar plots created for demonstrating the distribution of drugs among different groups of action were created using the "ggplot2" package from R.

4

RESULTS

4.1 EXPRESSION PROFILES AND DRUG DISCOVERY IN THE BLOOD SAMPLES FROM DENGUE COHORTS

The dendrogram analyses performed in the blood dengue cohorts allowed to confirm that most of the samples group in the respective cluster of infected or control expression profiles (Figure 7). We decided to keep all of them in the following analyses, maintaining the variance that can occur within groups.

When illustrating the results of the differential expression analyses in volcano plots (Figure 8; full information provided in Supplementary Tables S1-S3), it is clear that a considerable number of genes are significantly differentially expressed in each dataset. This is especially so in GSE51808, whose higher variability can be explained by it being made of several types of blood cells in comparison with the two other datasets made of only one type of blood cell. The statistically differentially expressed genes amounted in: 637 for GSE18090, 804 for GSE38246 and 7185 for GSE51808.

Applying Venn diagrams to the upregulated and the downregulated sets of these genes (Figure 9), it was possible to confirm the sharing of a number of genes between datasets: for the upregulated, 66 in the three datasets, 71 in GSE18090-GSE38246, 190 in GSE38246-GSE51808 and 400 in GSE18090-GSE51808; for the downregulated, 14 in all three datasets, 92 in GSE18090-GSE51808 and 172 in GSE38246-GSE51808.

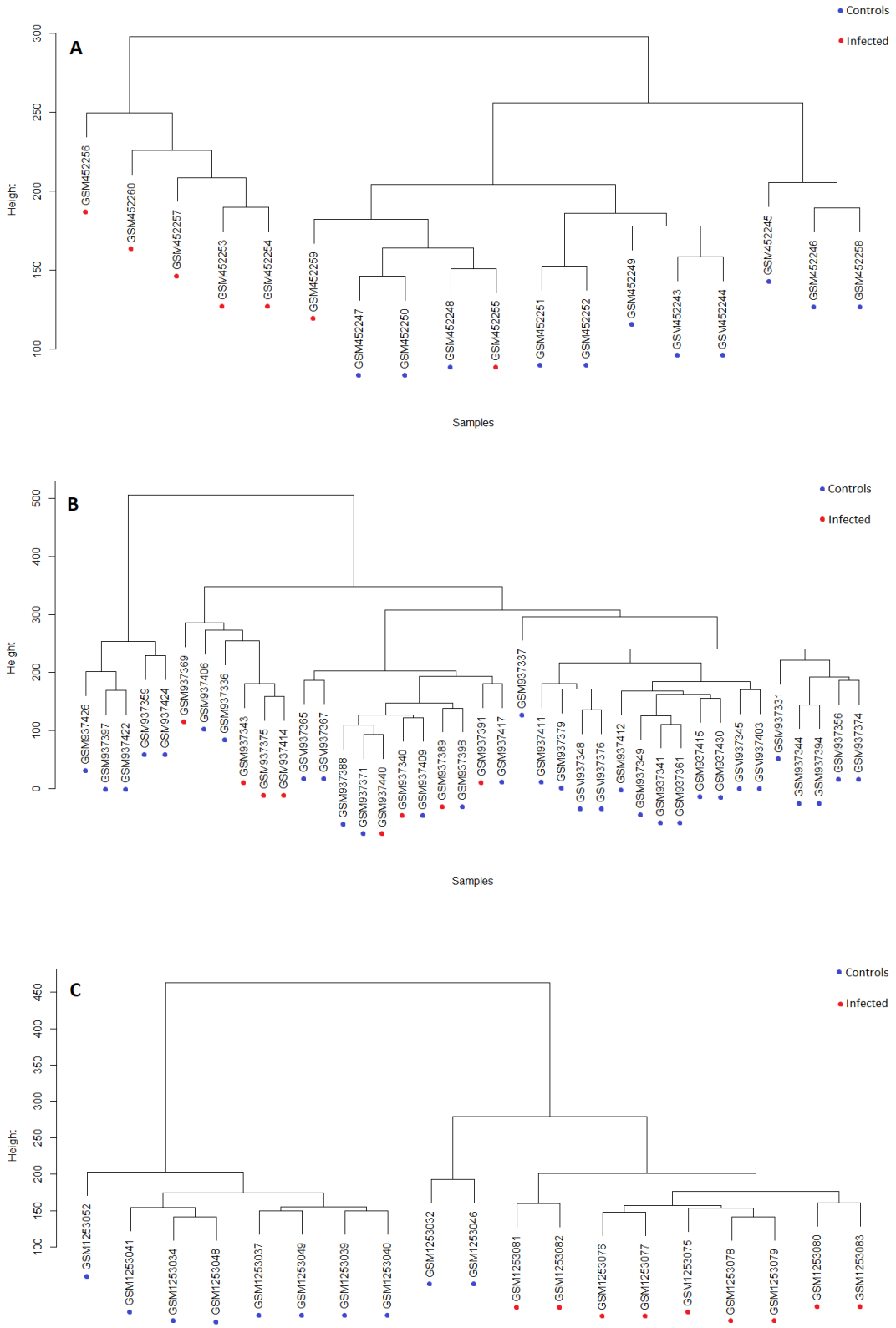


Figure 7.: Dendrograms for the gene expression profiles in the three blood datasets: (A) GSE18090, (B) GSE38246 and (C) GSE51808. The red points indicate the infected individuals, while the blue points indicate control samples.

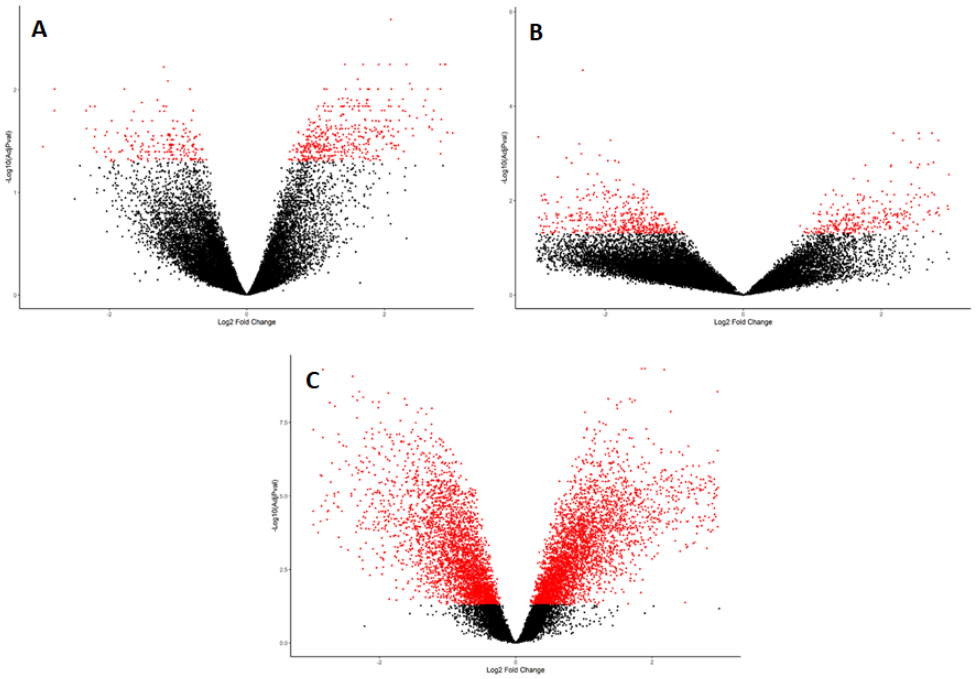


Figure 8.: Volcano plots for the gene differential expression in the three blood datasets: (A) GSE18090, (B) GSE38246 and (C) GSE51808. The points represented in red are the statistically significant genes. The negative side of the graph refers to downregulated genes in patients when compared to controls; the positive side of the graph refers to upregulated genes in patients when compared to controls.

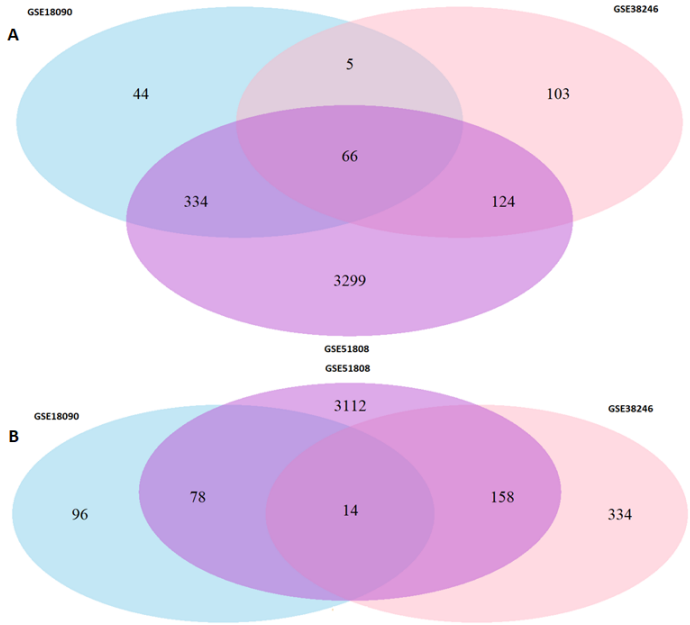


Figure 9.: Venn diagrams of the upregulated (A) and downregulated (B) genes in the three blood datasets.

In order to investigate which type of genes were upregulated and downregulated, we conducted gene set enrichment analysis in each of the three datasets. Figure 10 represents the top most significantly enriched pathways, when using the database KEGG (results with GO databases are reported in Supplementary Figures S1-S3 and Supplementary Tables S7-S18). As can be observed in the figure, biological pathways related with the immune system and response to pathogens are downregulated in the infected group, while biological pathways related with the cell cycle and repair mechanisms are upregulated in the infected group.

The CMap tool results for the drug discovery indicated a total of (Supplementary Table S31) 148 known compounds that were inferred (CMap score inferior to -90) as having potential impact in dengue haemorrhagic fever treatment. Figure 11 summarizes the amount of compounds taking into account its mechanism of action (some compounds have more than one mechanism of action). These drugs are spread across seven main different mechanisms of action being them antineoplastic (120), antibiotic (12), antiparasitic (10), immunosuppressant (eight), cardiovascular (seven), anti-inflammatory (six) and antiviral drugs (five).

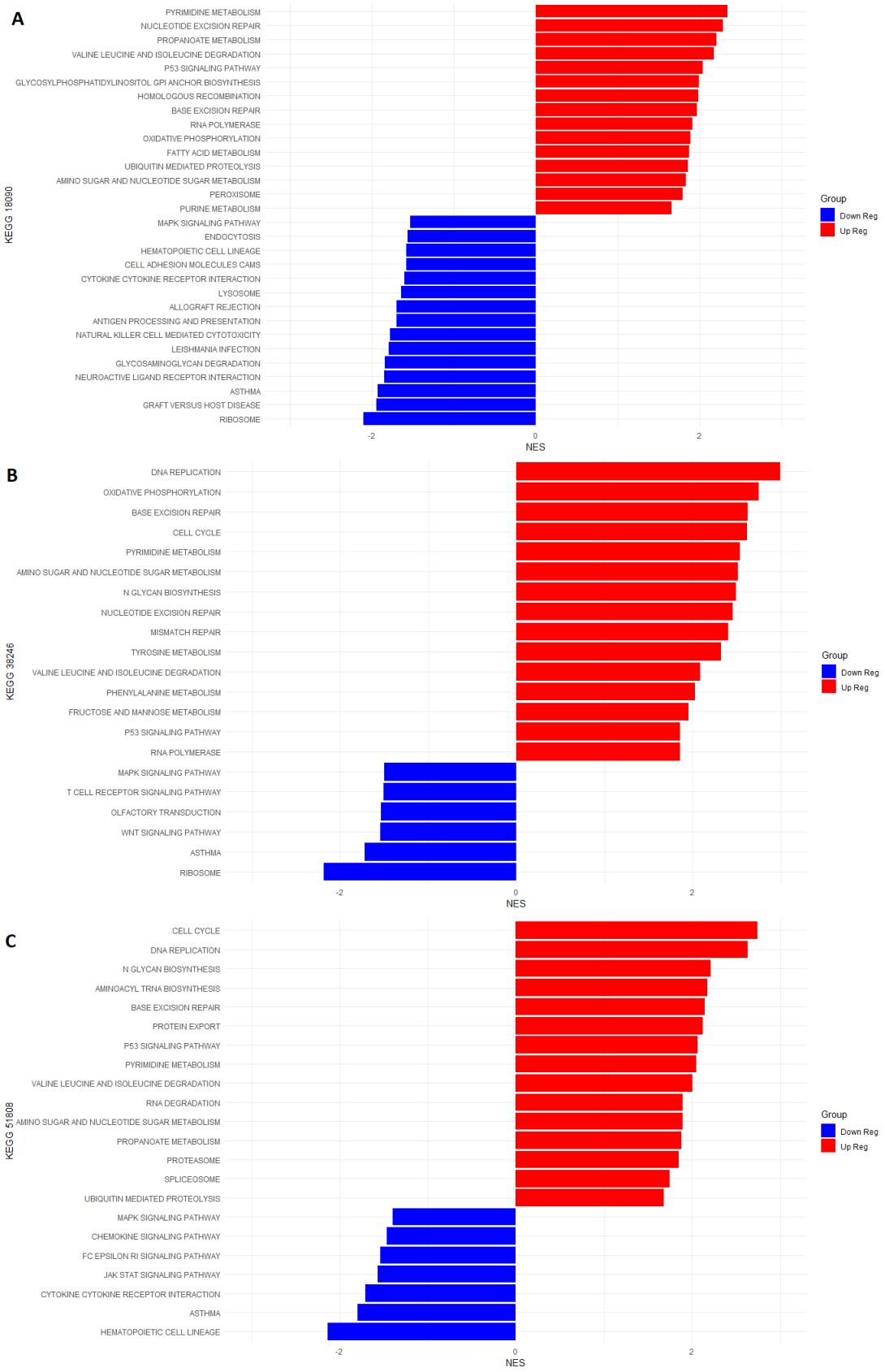


Figure 10.: Top KEGG pathways (cutoff value; -log p $\geq$ 2) in the three blood datasets: (A) GSE18090, (B) GSE38246, (C) GSE51808. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls; while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.

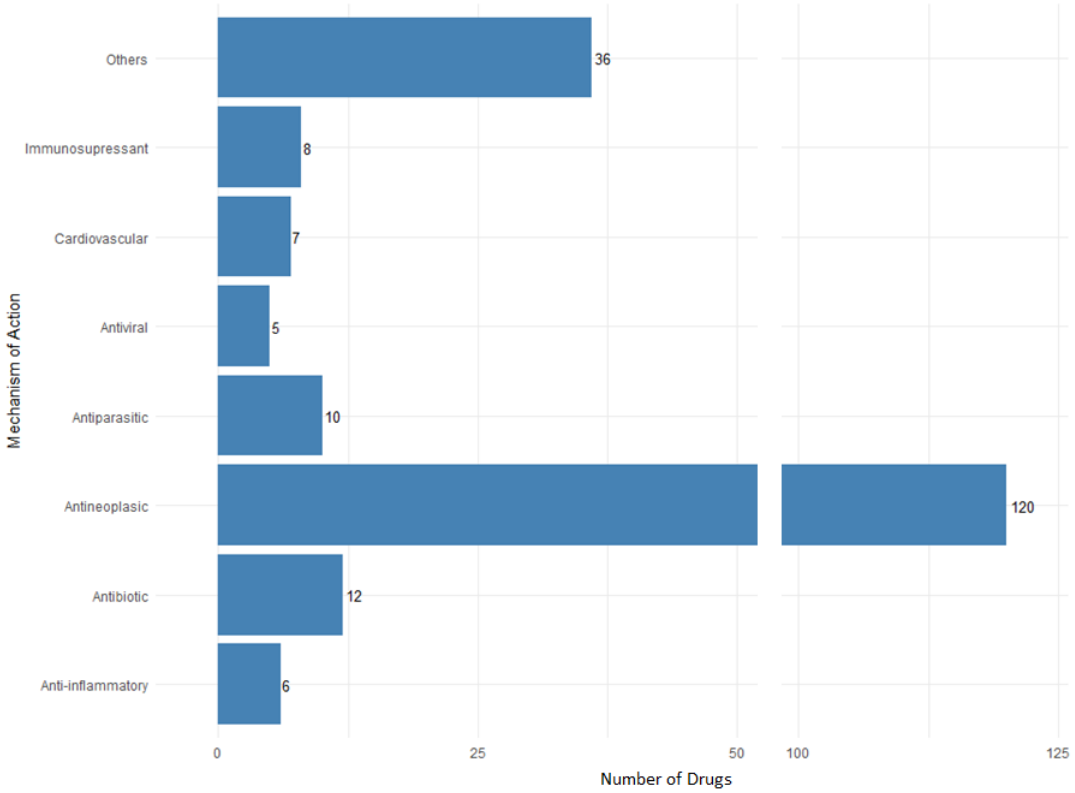


Figure 11.: CMap (score inferior to -90) identified compounds that potentially impact dengue haemorrhagic fever treatment, according to their mechanism of action.

4.2 LIVER, SPLEEN AND ENCEPHALON SAMPLES FROM DENGUE COHORTS

4.2.1 *AmpliSeq data Processing*

The AmpliSeq characterization aimed at approximately 10 million reads *per* sample. These samples were collected in deceased people, after the decaying processes occurring after death began (Ferreira et al., 2018). So, it is not surprising that the initial amount of reads obtained was lower, varying between 1,410,487 and 9,239,028, being in mean 5,412,203. There was no effect of the tissue analysed, as the mean of initial reads was similar between tissues: 5,987,967 in spleen; 5,136,132 in encephalon; and 5,077,912 in liver.

The quality control done through the GC content showed that the available tissue samples (liver and spleen) from cases 900 and 875 did not follow the theoretical unimodal distribution, presenting bimodal distributions, an indication of some form of contamination. Although we processed these samples for trimming and alignment, we ended up removing them from further analyses. Table 2 summarizes samples in terms of the number of initial reads, reads after trimming, and globally and locally aligned reads.

The trimming process reduced further the number of reads to continue under analyses, ending up with around 3.5 million reads. We then tested both local and global alignments, which lead to a mean number of reads of 3.4 and 2.8 millions, respectively. By performing linear regression analyses (Figure 12) in the controls between the number of reads per gene between local and global alignments, we confirmed the high correlation in all tissues (r2=0.871 in liver; r2=0.760 in spleen; and r2=0.992 in encephalon), so we decided to use the local aligned reads (higher amounts) for further analyses.

Table 2.: Amount of reads obtained in the various samples along the processing.

Case/Control	Tissue	Sample ID	Init Reads	Trim Reads	Loc Reads	Glob Reads
	Spleen	37478-B2	5912869	3468999	3087301	2424472
Case 37478	Enceph.	37478-ENC	3265359	1977898	1890646	1629243
	Liver	37478-HEP	3599850	2077442	1953011	1555946
	Spleen	39539-B2	7487948	4578828	4056040	3182181
Case 39539	Enceph.	39539-ENC	2555179	1558651	1491919	1307360
	Liver	39539-HEP	2920431	1708452	1588579	1263524
	Spleen	39538-B2	5748724	3425399	3057317	2438116
Case 39538	Enceph.	39538-ENC	3181759	1790796	1679569	1369419
	Liver	39538-HEP	1410487	744213	641014	539616
Case 900	Spleen	900-B2	4123839	2798065	2709844	2382675
	Liver	900-HEP	4561069	2648159	2485305	2132127
Case 875	Spleen	875-B2	2865948	1751492	807150	696681
	Liver	875-HEP	3377876	1937712	1709855	1485398
	Spleen	F1-B2	4971485	3171544	2862997	2292589
Control F1	Enceph.	F1-ENC	5730768	3398798	3225538	2723449
	Liver	F1-HEP	2021538	1167422	1047472	872742
	Spleen	F2-B2	7594924	5403008	5143529	4189310
Control F2	Enceph.	F2-ENC	8515179	6519338	6383762	5821274
	Liver	F2-HEP	8881841	6069089	5745221	4648833
	Spleen	F4-B2	7964950	5568306	5268896	4219780
Control F4	Enceph.	F4-ENC	7568552	5909657	5856624	5635984
	Liver	F4-HEP	9172251	6498742	6169857	4939151

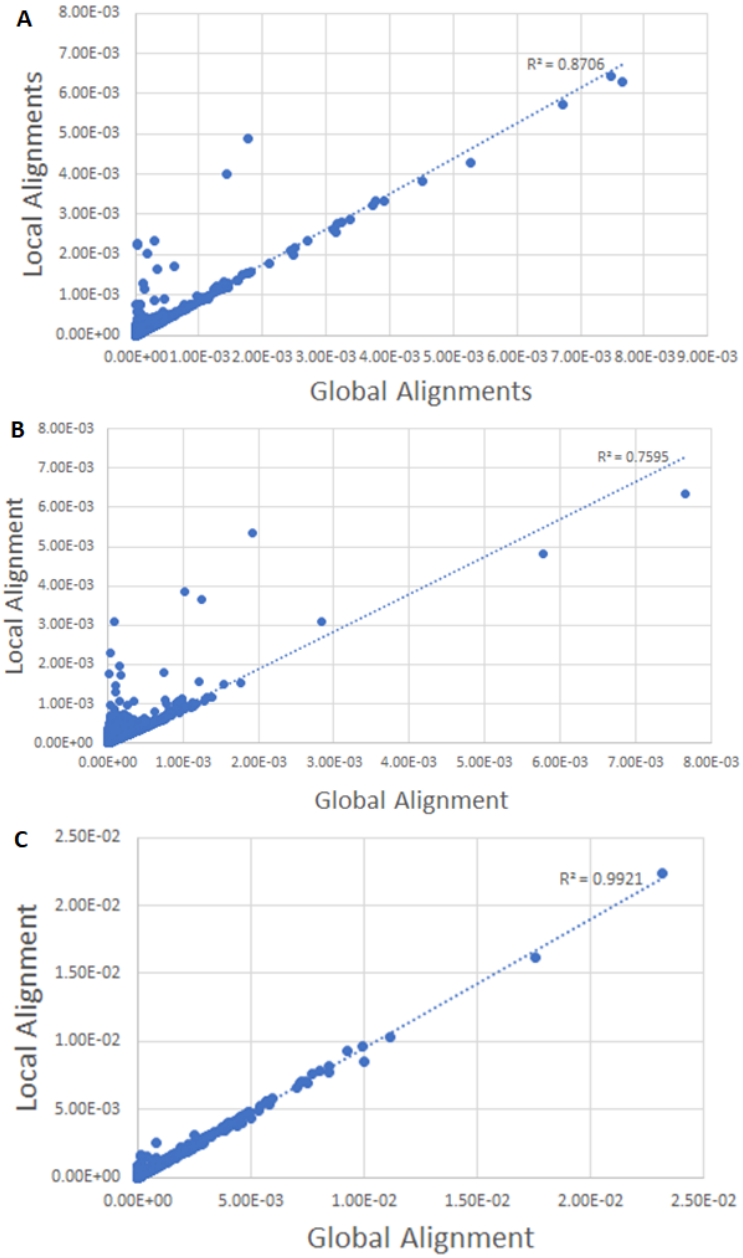


Figure 12.: Linear regression between the local and global aligned number of reads in the control samples in the various tissues (A - liver; B - spleen; C – encephalon).

4.2.2 *Expression profiles and drug discovery in the liver, spleen and encephalon samples from dengue cohorts*

The dendrogram analyses performed in the three tissues allowed to confirm that most of the samples group in the respective cluster of infected or control expression profiles (Figure 13).

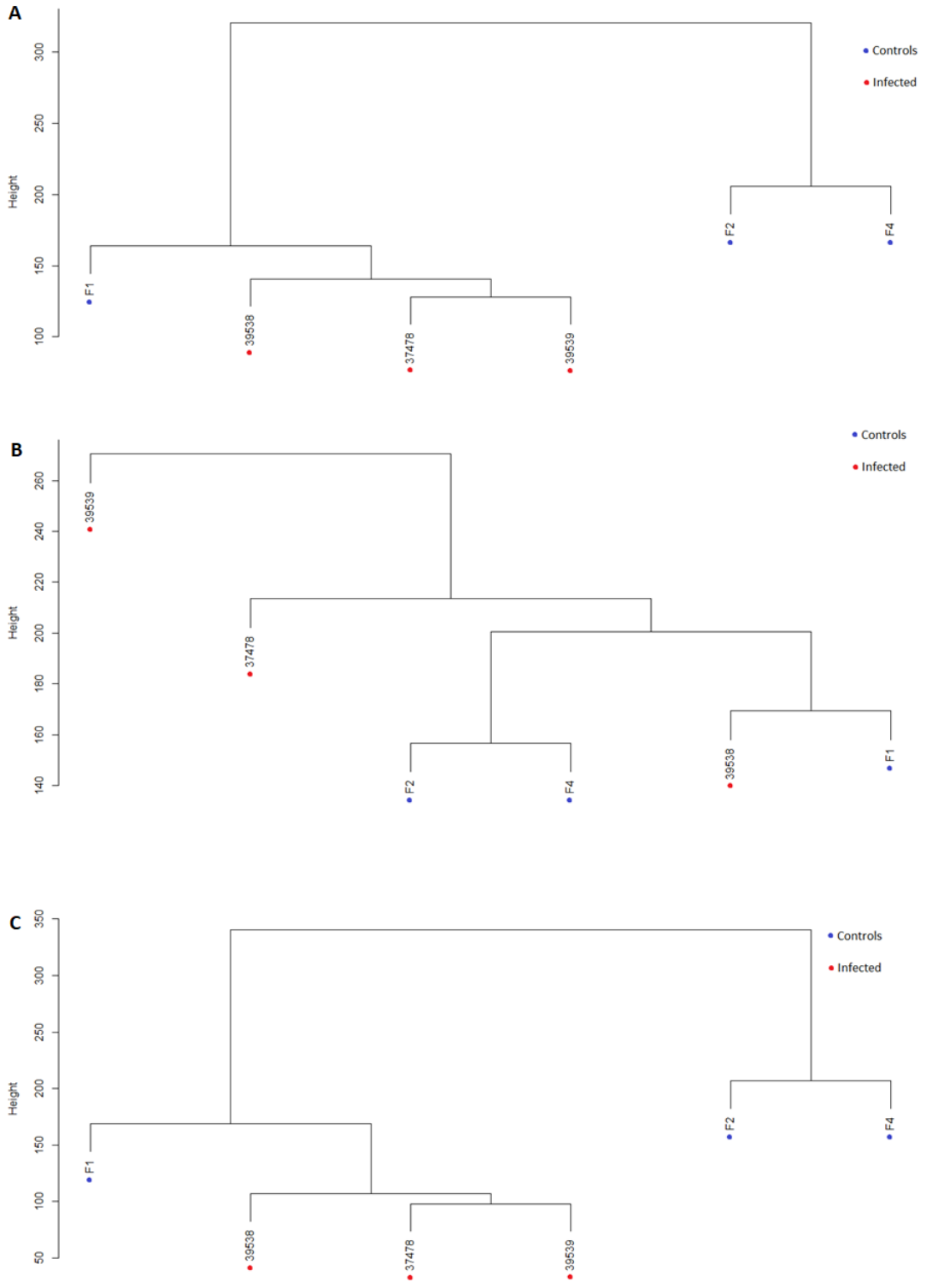


Figure 13.: Dendrograms for the gene expression profiles in the liver (A), spleen (B) and encephalon (C) datasets. The red points indicate the infected individuals, while the blue points indicate control samples.

The profiles of expression obtained in the various tissues indicated that the amount of genes with at least one read was between 75-88% of total genes evaluated: 15,675 in liver, 16,627 in spleen and 18,301 in encephalon. The number of genes maintained for each tissue after the DESeq2 independent filtering was 15,605 in liver, 5,169 in spleen and 11,956 in encephalon. The volcano plots for the differential expression (Figure 14; full information provided in Supplementary Tables S4-S6) reveal that the following numbers of significantly upregulated and downregulated genes were found in each tissue: 72 and 88 in liver; 518 and 102 in spleen; and 2362 and 168 in encephalon.

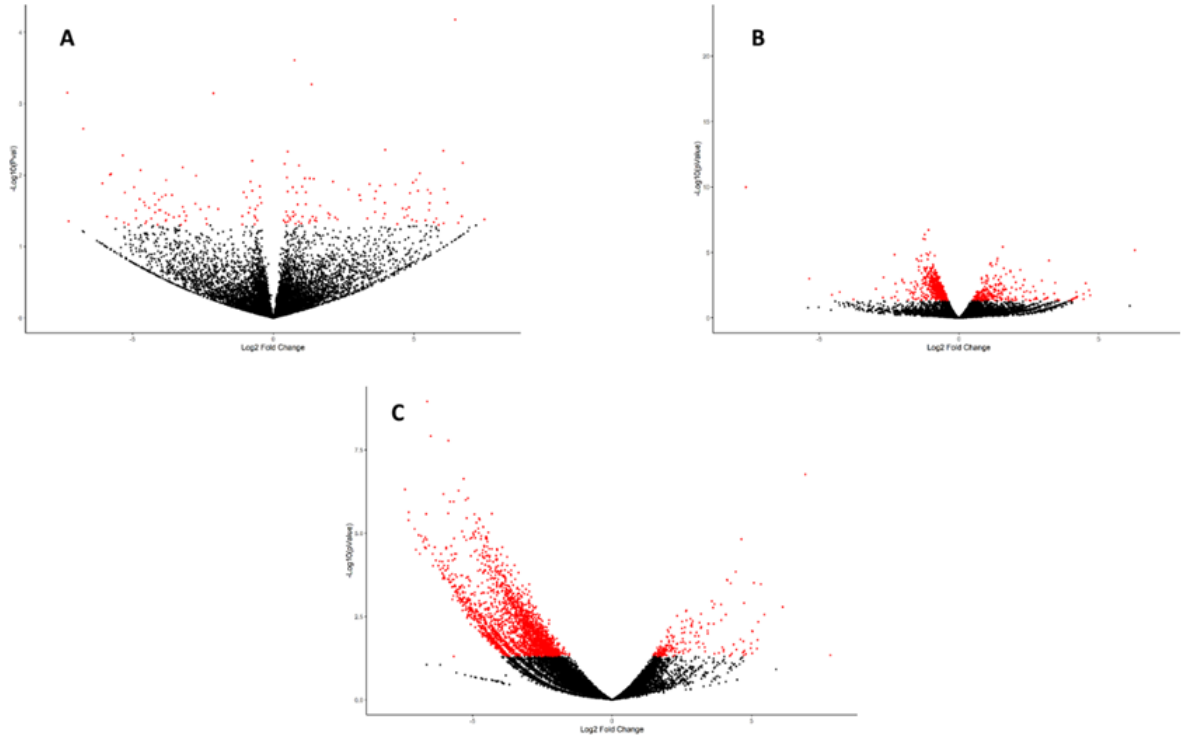


Figure 14.: Volcano plots for the gene differential expression in the liver (A), spleen (B) and encephalon (C) datasets. The points represented on red are the statistically significant genes. The negative side of the graph refers to downregulated genes in patients when compared to controls; the positive side of the graph refers to upregulated genes in patients when compared to controls.

Figure 15 illustrates the top most significantly downregulated pathways in the patients compared with controls for the KEGG database (results with GO databases are reported in Supplementary Figures S4-S6 and Supplementary Tables S7-S18). Results show a strong relation with metabolism, namely of lipids (propanoate, retinol, glycosphingolipid, fatty acid) in liver and spleen, while for encephalon, significant changes were observed in the downregulation in patients cohort of pathways related with response to infection and with neurotransmitters/cerebral diseases.

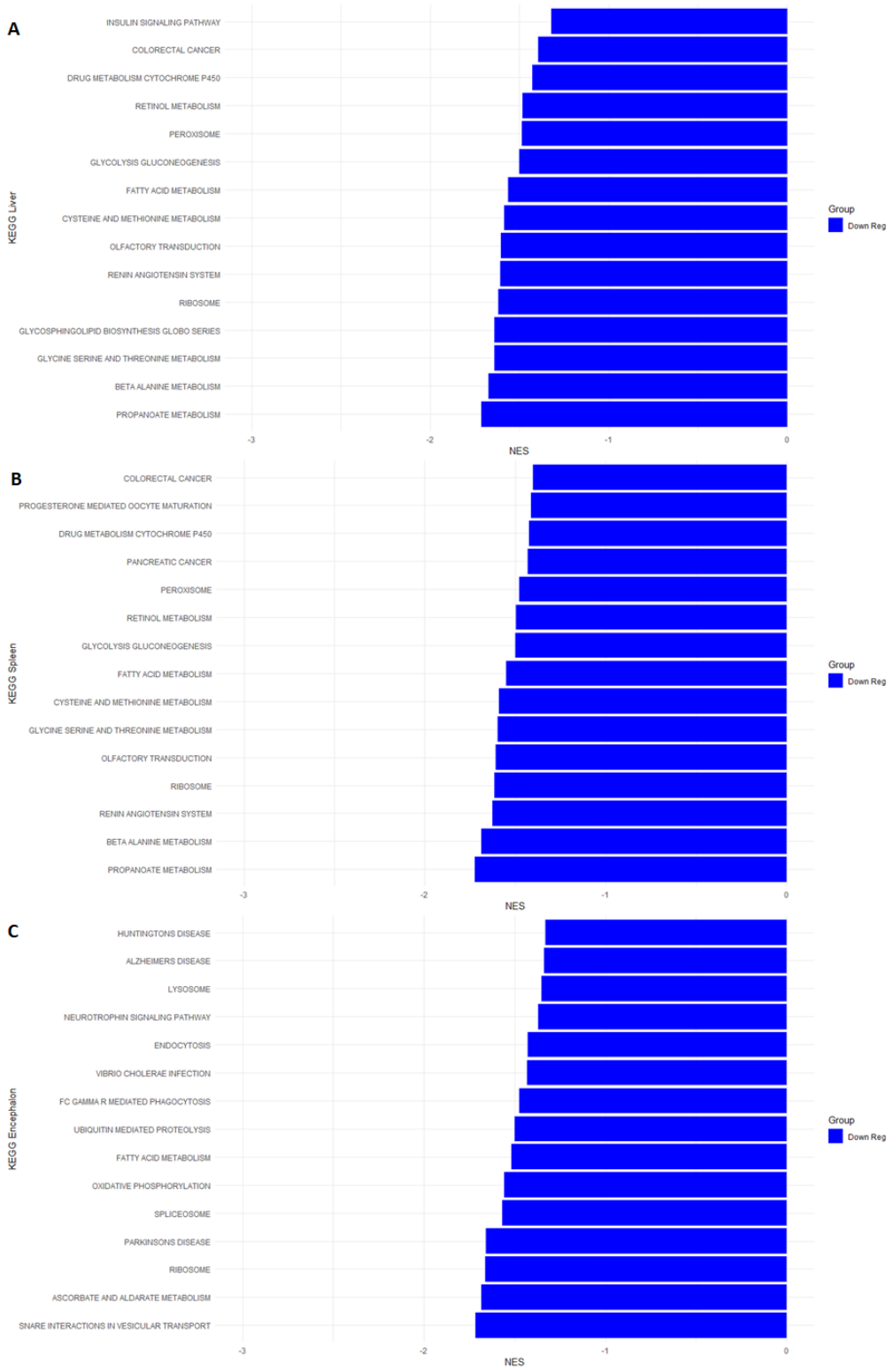


Figure 15.: Top KEGG pathways (cutoff value; -log p $\geq$ 2) in the (A) Liver, (B) Spleen and (C) Encephalon. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls, while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.

The CMap tool results, when considering all the cell lines included in the database (Figure 16A; Supplementary Tables S32; S34; S36) identified a diverse pattern from the one in blood, with more drugs of the anti-inflammatory class, cardiovascular (possibly related with lipids) and other (which includes the drugs for obesity and diabetes, thus metabolism acting drugs). As one of the cell lines used is derived from a hepatocellular carcinoma, we also obtained information on CMap when limiting to this cell line. This analysis has the advantage of avoiding averaging CMap scores amongst perturbagen's signatures from diverse cell lines, leading to a higher number of potentially effective drugs (Figure 16B; Supplementary Tables S33; S34; S37).

The comparison of CMap identified drugs between tissues does not indicate one drug effective in all tissues, at most, a few drugs (n=14) were identified in two tissues. Given the high stochasticity in identifying a drug from a connectivity map including 1,000,000 profiles, we instead based the comparison between tissues in the CMap molecular mechanism of action (Supplementary Table S38). There was only one mechanism of action common to all four tissues, the "histamine receptor antagonist", supporting the further testing of applying this kind of drugs in dengue (Malavige et al., 2018)."*Phosphoinositide 3-Kinase (PI3K)*" mechanism was identified in liver, spleen and blood, while the related "*Protein Kinase B (AKT)* inhibitor" and "*FMS-like Tyrosine Kinase (FLT3)* inhibitor" mechanisms were identified in encephalon, liver and blood, calling attention to a more careful evaluation of these mechanisms in dengue (Chen et al., 2017). Indications were somewhat contradicting for neurotransmitter receptors, such as dopamine and serotonin, as well as adenosine and adrenergic receptors, for which agonist, antagonist and inhibitor mechanisms were simultaneously identified (in some cases, even in the same tissue). These opposing insights call for careful functional evaluation of these drugs in dengue (Smith et al., 2014), especially as they can have several side effects affecting personality.

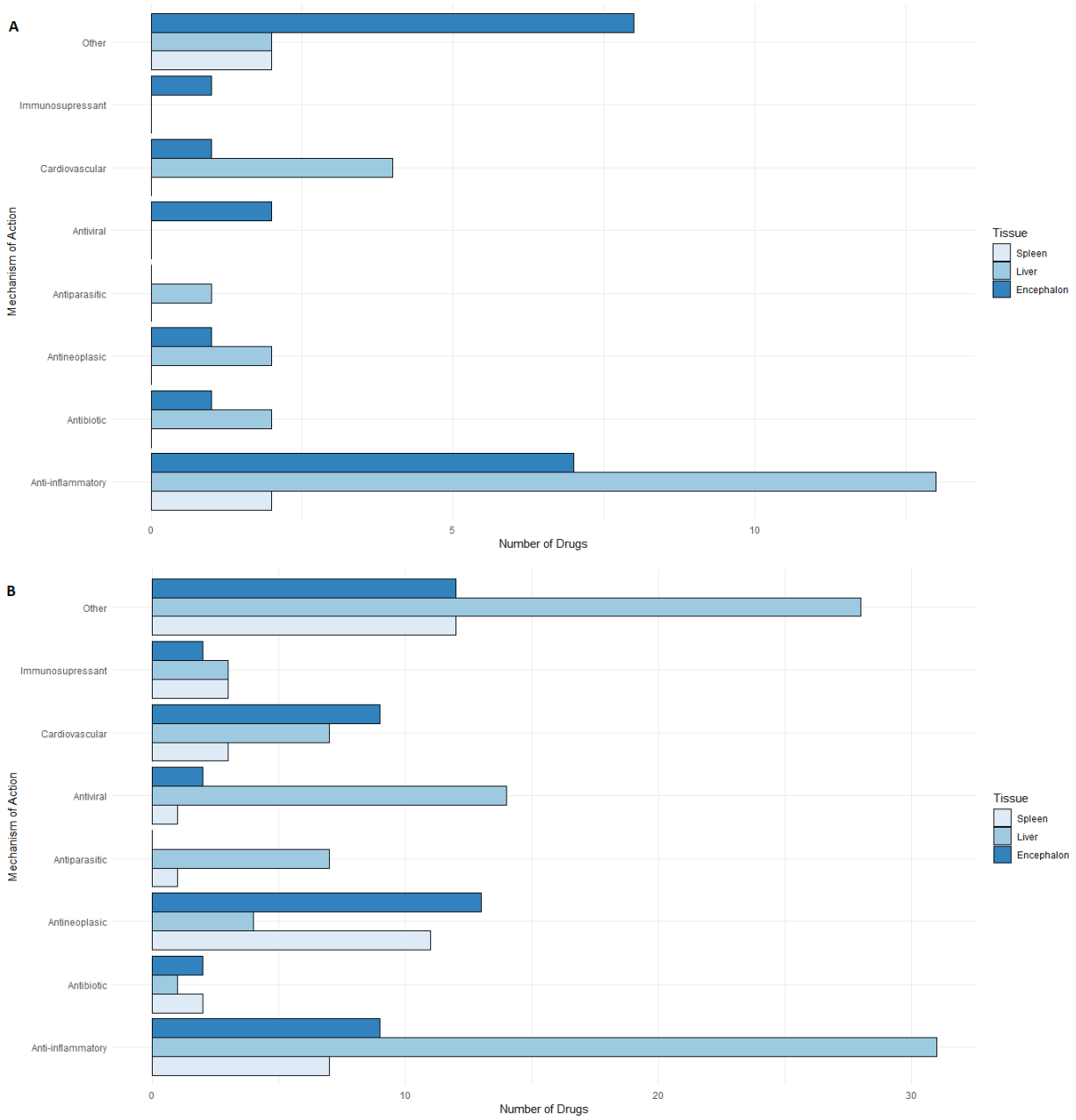


Figure 16.: CMap (score inferior to -90) identified compounds that potentially impact dengue hemorrhagic fever treatment, according to their mechanism of action, in liver, spleen and encephalon. A- When considering all cell lines; B- Comparison versus hepatic cell line.

5

DISCUSSION

The decreasing cost of *omics* profiling has render high-throughput technologies available for understanding the molecular basis of diseases. Compendia of gene expression profiles which have been perturbed pharmacologically have been obtained, allowing to glean gene function (Hughes et al., 2000). In humans, the second phase of CMap, called L1000 (Subramanian et al., 2017), is one of the most complete compendia, including 1,319,138 profiles from 42,080 perturbagens (19,811 small molecule compounds, 18,493 shRNAs, 3,462 cDNAs, and 314 biologics), corresponding to 25,200 biological entities (19,811 compounds, shRNA and/or cDNA against 5,075 genes, and 314 biologics) for a total of 473,647 signatures (consolidating replicates). Some authors already compared blood expression profiles of dengue patients against CMap (Amemiya et al., 2019), but a broader picture for other affected tissues was in demand.

Post-mortem human tissue samples are crucial for studying the patterns of gene expression underlying tissue specificity, as sampling many tissues, especially from encephalon, from living individuals would be impossible. A drawback of the *post-mortem* human tissue samples is the significant reduction of RNA, which is degraded fast upon death (Ferreira et al., 2018). Thus, *post-mortem* RNA-Seq analysis poses similar alignment issues to DNA samples retrieved from ancient or historical specimens, in being typically highly degraded into small fragments and having contaminant molecules hampering amplification. A commonly strategy used in the analysis of ancient sequence data is to perform local alignment (Taron et al., 2018), where sub-portions of the query sequence are tested in the alignment to the reference sequence. By contrast, in the global alignment, the entire query sequence is aligned to the reference target sequence, this way it is also referred to as end to end alignment. Thus local alignments are best adjusted to degraded samples, leading to a higher number of aligned reads. In our samples, although we aimed at an AmpliSeq characterization of around 10.8 million reads *per* sample, the values obtained varied widely between samples, attaining a mean of 5.4 million of initial reads, concordant with some degradation of the samples. After local alignment, a mean of 3.4 million aligned reads was obtained. Linear regression analyses confirmed that the general profile of gene expression between local and global alignments was highly concordant, being higher in the local than global

alignment, as expected in theory. This observation supported the use of the local alignment inferred expression levels in this work.

Thus, the analysed samples in this work allowed to characterize the main gene expression changes occurring in blood, liver, spleen and encephalon of dengue patients, and to identify candidate drugs directed to those changes. In blood, pathways related with the immune system and response to pathogens were downregulated, demonstrating the failure of the immune system in responding to the disease in these deceased patients, while those related with cell cycle/repair mechanisms were upregulated as was already described (Kwissa et al., 2014). Concordantly, CMap identified 58.8% of antineoplasic drugs amid the 204 identified drugs-mechanism of action in blood. In liver and in spleen there was a significant downregulation of metabolic pathways, especially of lipids, and many anti-inflammatory and related metabolism drugs were pointed out. In brain, pathways of response to infection and related with neurotransmitters/cerebral diseases were increased, concordant with CMap identification of drugs acting in neurotransmitters, adenosine and adrenergic receptors.

The dissimilarity in the results between tissues, which reflects tissue specificity, could at first instance indicate that a cocktail-based treatment would be more suitable to address this diversity. However, given the high connectivity between molecular networks in the body, and the multi-effects of drugs, some candidate drugs can be effective in the various tissues in dengue disease. Some important insights can already be summarized from our results, which coupled with previous independent *in vitro* and *in vivo* testing, inform us on how pursuit further on drug repurposing for treatment of later phases of dengue disease. The only signal common to all four tissues was for the "histamine receptor antagonist". Four drugs were identified in the four tissues, but these are highly related with rupatadine, an orally available second-generation antihistamine that has been already tested in dengue (Malavige et al., 2018). Rupatadine has long acting dual histamine-1-receptor blocking activities and platelet activating factor receptor blocking activities, useful for the treatment of allergic disease and chronic urticarial, while in dengue it also significantly reduced endothelial permeability by dengue *in vitro*, and significantly inhibited the increased haematocrit in dengue-infected mice with dose-dependency (Malavige et al., 2018). These authors conducted a randomised, placebo-controlled trial in 183 adult patients in Sri Lanka with acute dengue, which showed that: (1) rupatadine up to 40 mg daily appeared safe and well-tolerated, (2) higher platelet counts and lower aspartate-aminotransferase levels on day 7 in the rupatadine group compared to the placebo group, and (3) smaller effusions on day 8 in the subgroup of patients with pleural effusions. A larger sample size and range of recruitment time in the future will allow to evaluate more securely the potential beneficial effects of rupatadine in dengue.

Host cell kinases are important for the replication of some hemorrhagic fever viruses, and panels of kinases inhibitors have been tested for their ability to block these viruses' replication (Mohr et al., 2015). One signalling pathway, the PI3K/AKT pathway, was reported to be essential for the propagation of the viruses, and accordingly, the CMap results indicated "PI3K inhibitor" in liver, spleen and blood, while the related "AKT inhibitor" was identified in encephalon, liver and blood. Also, "FLT3 inhibitor" mechanism was detected in encephalon, liver and blood, which may reflect the fact that constitutive activation of FLT3 is involved in leukemogenesis, partially through phosphorylation/activation of the serine-threonine kinase (Chen et al., 2010). Again, all this mechanisms are related with "MTOR inhibitor" (detected in spleen and blood), as the mammalian target of rapamycin (mTOR), an important downstream effector of AKT and a master regulator of cell growth and metabolism, seems to have its signalling activated by FLT3 kinase (Chen et al., 2010).

There were many signals for the involvement of various receptors, as of dopamine, serotonin, adenosine and adrenergic, for which agonist, antagonist and inhibitor mechanisms were simultaneously identified. Previously, (Smith et al., 2014) identified a group of small molecules that inhibited infection with dengue and other flaviviruses, and interestingly those molecules were similar in structure to tricyclic antipsychotic compounds that act as antagonists of serotonin and dopamine receptors. Further knockdown of dopamine receptor D4 reduced DENV replication, via inhibition of *Epidermal Growth Factor Receptor (EGFR)*-related kinase (ERK) phosphorylation. Although it seems strange the involvement of receptors that are mainly located in neurons, it has been shown that dopamine receptors are expressed on rodent and human macrophages, a primary target cell of DENV infection. These macrophage related dopamine receptors may be the ones being influenced by DENV. The link with EGFR was also detected in CMap related in liver and blood, as "EGFR inhibitor".

At first instance, there were not many indications of lipid-related drugs in the CMap results. This may be so because of the late stage of the disease analysed in this work, while lipids may be more important in the first stages of entrance of the virus into the cells and in their replication in the acute phase (Sierra et al., 2017). Even so, a "LXR agonist" mechanism was identified in liver, "Vitamin D receptor agonist" and "Vitamin K antagonist" were detected in encephalon, and statins (lovastatin and simvastatin, respectively) were highlighted in spleen and blood. These results call for a continuation on trials with statins (bigger sample sizes; different concentrations of intake; different times of intake), as at least the few clinical trials already realized (Whitehorn et al., 2015; Chia et al., 2018) showed that treatment or continuation of statin intake has no negative effects on dengue patients.

6

CONCLUDING REMARKS AND FUTURE WORK

The work conducted here testifies the gain in information on drug repurposing through the performance of *omics* analyses in the various tissues affected in a disease. In this way, the *in silico* evaluation of drugs can take into consideration the effects of tissue specialization for the design of a more efficient treatment strategy, through the selection of a compound effective in several tissues or a cocktail of drugs able to deal with this diversity. This strategy is especially important in discovering innovative drug treatments for neglected diseases, as the funding for the study of these diseases is scarce.

Specifically, the work developed by us sheds more light in the complicate physiopathology of Dengue, giving more insight in the alterations in specific tissues (liver, spleen and encephalon), pathways and genes. Additionally, the relation of these alterations with drugs, also performed by us, could also contribute in the discovery of treatments for this disease which is still untreated.

Several paths can now be followed departing from these first results. In terms of bioinformatics, machine learning could be applied to the transcriptomic data in the various tissues, as an alternative methodology to identify effective drugs. These methods should be trained in another database different from the one used in CMap, so that biases due to the limited number and type of cell lines and of tested drugs can be avoided. Also, given that *post-mortem* samples are prone to degradation, alignment algorithms that were specifically designed for degraded DNA samples could be applied and compared with our results.

In biological terms, an increase in the number and quality (decreasing time since death for the extraction of RNA) of samples *per* tissue should be procured, in order to replicate the study. Despite this, the identification of a few drugs biologically meaningful in the context of dengue disease, could already support the pursuit of functional evaluation tests, first *in vitro* (infection assays in diverse types of cells) and then *in vivo* (in mice models of dengue disease).

BIBLIOGRAPHY

Rasmus Agren, Adil Mardinoglu, Anna Asplund, Caroline Kampf, Mathias Uhlen, and Jens Nielsen. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Molecular Systems Biology*, 10(3):1–13, 2014. ISSN 17444292. doi: 10.1002/msb.145122.

Takayuki Amemiya, M. Michael Gromiha, Katsuhisa Horimoto, and Kazuhiko Fukui. Drug repositioning for dengue haemorrhagic fever by integrating multiple omics analyses. *Scientific Reports*, 9(1):523, 2019. ISSN 2045-2322. doi: 10.1038/s41598-018-36636-1. URL http://www.nature.com/articles/s41598-018-36636-1.

Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, 2015. ISSN 14602059. doi: 10.1093/bioinformatics/btu638.

Simon D Atkins, Ian M Clark, Marcela Echavarria, Jose L. Sanchez, Shellie A. Kolavic-Gray, Christina S. Polyak, Felicia Mitchell-Raymundo, Bruce L. Innis, David Vaughn, Richard Reynolds, Leonard N. Binn, A S Fauci, Elodie Ghedin, Naomi a Sengamalay, Martin Shumway, Jennifer Zaborsky, Tamara Feldblyum, Vik Subbu, David J Spiro, Jeff Sitz, Hean Koo, Pavel Bolotov, Dmitry Dernovoy, Tatiana Tatusova, Yiming Bao, Kirsten St George, Jill Taylor, David J Lipman, Claire M Fraser, Jeffery K Taubenberger, Steven L Salzberg, Illumina, Iruka N Okeke, Ramanan Laxminarayan, Zulfiqar a Bhutta, Adriano G Duse, Philip Jenkins, Thomas F O'Brien, Ariel Pablos-Mendez, Keith P Klugman, Lília Perfeito, Lisete Fernandes, Catarina Mota, Isabel Gordo, Alan D. Radford, David Chapman, Linda Dixon, Julian Chantrey, Alistair C. Darby, Neil Hall, Pei-yun Shu, Jyh-hsiung Huang, Jane D. Siegel, Emily Rhinehart, Mary Marguerite Jackson, Linda Chiarello, Samuel Yang, Richard E. Rothman, Céire Costelloe, Chris Metcalfe, Andrew Lovering, David Mant, Alastair D Hay, E Domingo, Laura J. Frost, Michael R. Reich, Rachel L Hallett, Colin J Sutherland, Neal Alexander, Rosalynn Ord, Musa Jawara, Chris J Drakeley, Gijs Walraven, Geoffrey a T Targett, Margaret Pinder, Ali Alloueche, Colin R Parrish, Edward C Holmes, David M Morens, Eun-chung Park, Donald S Burke, Charles H Calisher, Catherine A Laughlin, Linda J Saif, Peter Daszak, Vania Rosas-Magallanes, Gustave Stadthagen-Gomez, Jean Rauzier, Luis B. Barreiro, Ludovic Tailleux, Frédéric Boudou, Ruth Griffin, Jérome Nigou, Mary Marguerite Jackson, Brigitte Gicquel, Olivier Neyrolles, Pei-yun Shu, Jyh-hsiung Huang, C Lee Ventola, Mark Woolhouse, Wan Yang, Alicia Karspeck, and Jeffrey Shaman. Current Advances in Dengue.

Clinical and Diagnostic Laboratory Immunology, 11(4):642–650, 2004. ISSN 1071-412X. doi: 10.1128/CDLI.11.4.642. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3130384{&}tool=pmcentrez{&}rendertype=abstract{%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25859123{%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4378521{%}5Cnhttp://www.ncbi.nlm.nih.gov/pubm.

Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, and Alexandra Soboleva. NCBI GEO: Archive for functional genomics data sets - Update. *Nucleic Acids Research*, 41(D1):991–995, 2013. ISSN 03051048. doi: 10.1093/nar/gks1193.

Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btu170.

Oliver J. Brady, Peter W. Gething, Samir Bhatt, Jane P. Messina, John S. Brownstein, Anne G. Hoen, Catherine L. Moyes, Andrew W. Farlow, Thomas W. Scott, and Simon I. Hay. Refining the Global Spatial Limits of Dengue Virus Transmission by Evidence-Based Consensus. *PLoS Neglected Tropical Diseases*, 6(8), 2012. ISSN 19352735. doi: 10.1371/journal.pntd.0001760.

Roger Bumgarner. Overview of DNA microarrays: types, applications, and their future. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, Chapter 22:Unit 22.1., jan 2013. ISSN 1934-3647. doi: 10.1002/0471142727.mb2201s101. URL http://www.ncbi.nlm.nih.gov/pubmed/23288464http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4011503.

Maria Cabarcas-Montalvo, Wilson Maldonado-Rojas, Diana Montes-Grajales, Angela Bertel-Sevilla, Irene Wagner-Döbler, Helena Sztajer, Michael Reck, Maria Flechas-Alarcon, Raquel Ocazionez, and Jesus Olivero-Verbel. Discovery of antiviral molecules for dengue: In silico search and biological evaluation. *European Journal of Medicinal Chemistry*, 110: 87–97, mar 2016. ISSN 0223-5234. doi: 10.1016/J.EJMECH.2015.12.030. URL https://www.sciencedirect.com/science/article/pii/S0223523415304165?via{%}3Dihub.

Megan E Cahill, Samantha Conley, Andrew T Dewan, and Ruth R Montgomery. Identification of genetic variants associated with dengue or West Nile virus disease : a systematic review and meta-analysis. pages 1–18, 2018.

S. Carbon, H. Dietze, S. E. Lewis, C. J. Mungall, M. C. Munoz-Torres, S. Basu, R. L. Chisholm, R. J. Dodson, P. Fey, Paul D. Thomas, H. Mi, A. Muruganujan, X. Huang,

S. Poudel, J. C. Hu, S. A. Aleksander, B. K. McIntosh, D. P. Renfro, D. A. Siegele, G. Antonazzo, H. Attrill, N. H. Brown, S. J. Marygold, P. Mc-Quilton, L. Ponting, G. H. Millburn, A. J. Rey, R. Stefancsik, S. Tweedie, K. Falls, A. J. Schroeder, M. Courtot, D. Osumi-Sutherland, H. Parkinson, P. Roncaglia, R. C. Lovering, R. E. Foulger, R. P. Huntley, P. Denny, N. H. Campbell, B. Kramarz, S. Patel, J. L. Buxton, Z. Umrao, A. T. Deng, H. Alrohaif, K. Mitchell, F. Ratnaraj, W. Omer, M. Rodríguez-López, M. C. Chibucos, M. Giglio, S. Nadendla, M. J. Duesbury, M. Koch, B. H.M. Meldal, A. Melidoni, P. Porras, S. Orchard, A. Shrivastava, H. Y. Chang, R. D. Finn, M. Fraser, A. L. Mitchell, G. Nuka, S. Potter, N. D. Rawlings, L. Richardson, A. Sangrador-Vegas, S. Y. Young, J. A. Blake, K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, M. A. Harris, J. Hayles, S. G. Oliver, K. Rutherford, V. Wood, J. Bahler, A. Lock, J. De Pons, M. Dwinell, M. Shimoyama, S. Laulederkind, G. T. Hayman, M. Tutaj, S. J. Wang, P. D'Eustachio, L. Matthews, J. P. Balhoff, R. Balakrishnan, G. Binkley, J. M. Cherry, M. C. Costanzo, S. R. Engel, S. R. Miyasato, R. S. Nash, M. Simison, M. S. Skrzypek, S. Weng, E. D. Wong, M. Feuermann, P. Gaudet, T. Z. Berardini, D. Li, B. Muller, L. Reiser, E. Huala, J. Argasinska, C. Arighi, A. Auchincloss, K. Axelsen, G. Argoud-Puy, A. Bateman, B. Bely, M. C. Blatter, C. Bonilla, L. Bougueleret, E. Boutet, L. Breuza, A. Bridge, R. Britto, H. Hye-A-Bye, C. Casals, E. Cibrian-Uhalte, E. Coudert, I. Cusin, P. Duek-Roggli, A. Estreicher, L. Famiglietti, P. Gane, P. Garmiri, G. Georghiou, A. Gos, N. Gruaz-Gumowski, E. Hatton-Ellis, U. Hinz, A. Holmes, C. Hulo, F. Jungo, G. Keller, K. Laiho, P. Lemercier, D. Lieberherr, A. Mac- Dougall, M. Magrane, M. J. Martin, P. Masson, D. A. Natale, C. O'Donovan, I. Pedruzzi, K. Pichler, D. Poggioli, S. Poux, C. Rivoire, B. Roechert, T. Sawford, M. Schneider, E. Speretta, A. Shypitsyna, A. Stutz, S. Sundaram, M. Tognolli, C. Wu, I. Xenarios, L. S. Yeh, J. Chan, S. Gao, K. Howe, R. Kishore, R. Lee, Y. Li, J. Lomax, H. M. Muller, D. Raciti, K. Van Auken, M. Berriman, L. Stein, Paul Kersey, P. W. Sternberg, D. Howe, and M. Westerfield. Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. *Nucleic Acids Research*, 45(D1):D331–D338, 2017. ISSN 13624962. doi: 10.1093/nar/gkw1108.

Hsin Hsin Chen, Chien Chin Chen, Yee Shin Lin, Po Chun Chang, Zi Yi Lu, Chiou Feng Lin, Chia Ling Chen, and Chih Peng Chang. AR-12 suppresses dengue virus replication by down-regulation of PI3K/AKT and GRP78. *Antiviral Research*, 142:158–168, 2017. ISSN 18729096. doi: 10.1016/j.antiviral.2017.02.015.

Rubing Chen and Nikos Vasilakis. Dengue-Quo tu et quo vadis? *Viruses*, 3(9):1562–1608, 2011. ISSN 19994915. doi: 10.3390/v3091562.

Weina Chen, Elias Drakos, Ioannis Grammatikakis, Ellen J. Schlette, Jiang Li, Vasiliki Leventaki, Efi Staikou-Drakopoulou, Efstratios Patsouris, Panayiotis Panayiotidis, L. Jeffrey Medeiros, and George Z. Rassidakis. MTOR signaling is activated by FLT3 ki-

nase and promotes survival of FLT3-mutated acute myeloid leukemia cells. *Molecular Cancer*, 9(1):292, 2010. ISSN 14764598. doi: 10.1186/1476-4598-9-292. URL http://www.molecular-cancer.com/content/9/1/292.

Po Ying Chia, Htet Lin Htun, Wei Ping Ling, Yee Sin Leo, Tsin Wen Yeo, and David Chien Boon Lye. Hyperlipidemia, statin use and dengue severity. *Scientific Reports*, 8 (1):1–7, 2018. ISSN 20452322. doi: 10.1038/s41598-018-35334-2. URL http://dx.doi.org/10.1038/s41598-018-35334-2.

Peter J A Cock, Christopher J. Fields, Naohisa Goto, Michael L. Heuer, and Peter M. Rice. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6):1767–1771, 2009. ISSN 03051048. doi: 10.1093/nar/gkp1137.

W. T.T. De Silva and M. Gunasekera. Spontaneous splenic rupture during the recovery phase of dengue fever. *BMC Research Notes*, 8(1):5–8, 2015. ISSN 17560500. doi: 10.1186/s13104-015-1234-5.

Michael S. Diamond and Theodore C. Pierson. Molecular Insight into Dengue Virus Pathogenesis and Its Implications for Disease Control. *Cell*, 162(3):488–492, 2015. ISSN 10974172. doi: 10.1016/j.cell.2015.07.005. URL http://dx.doi.org/10.1016/j.cell.2015.07.005.

Alison Farrell. Cross-reactive memory. *Nature Medicine*, 19(4):417–417, 2013. ISSN 1078-8956. doi: 10.1038/nm.3172. URL http://www.nature.com/articles/nm.3172.

Pedro G. Ferreira, Manuel Muñoz-Aguirre, Ferran Reverter, Caio P. Sá Godinho, Abel Sousa, Alicia Amadoz, Reza Sodaei, Marta R. Hidalgo, Dmitri Pervouchine, Jose Carbonell-Caballero, Ramil Nurtdinov, Alessandra Breschi, Raziel Amador, Patrícia Oliveira, Cankut Çubuk, João Curado, François Aguet, Carla Oliveira, Joaquin Dopazo, Michael Sammeth, Kristin G. Ardlie, and Roderic Guigó. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nature Communications*, 9(1), 2018. ISSN 20411723. doi: 10.1038/s41467-017-02772-x.

Enrico Ferrero, Ian Dunham, and Philippe Sanseau. In silico prediction of novel therapeutic targets using gene-disease association data. *Journal of Translational Medicine*, 15(1):1–16, 2017. ISSN 14795876. doi: 10.1186/s12967-017-1285-6.

Duane J. Gubler. The economic burden of dengue. *American Journal of Tropical Medicine and Hygiene*, 86(5):743–744, 2012. ISSN 00029637. doi: 10.4269/ajtmh.2012.12-0157.

VishalKumar Gupta, Arati Lalchandani, and Kanhaiya Agrawal. Neurological manifestations of dengue infection. *Annals of Tropical Medicine and Public Health*, 8(4):117, 2015.

ISSN 1755-6783. doi: 10.4103/1755-6783.162370. URL http://www.atmph.org/text.asp?2015/8/4/117/162370.

Yara A. Halasa, Donald S. Shepard, and Wu Zeng. Economic cost of dengue in Puerto Rico. *American Journal of Tropical Medicine and Hygiene*, 86(5):745–752, 2012. ISSN 00029637. doi: 10.4269/ajtmh.2012.11-0784.

Yukiko Higa. Dengue Vectors and their Spatial Distribution. *Tropical Medicine and Health*, 39(4SUPPLEMENT):S17–S27, 2011. ISSN 1349-4147. doi: 10.2149/tmh.2011-S04. URL http://joi.jlc.jst.go.jp/JST.JSTAGE/tmh/2011-S04?from=CrossRef.

L. T. Hoang, D. J. Lynn, M. Henn, B. W. Birren, N. J. Lennon, P. T. Le, K. T. H. Duong, T. T. H. Nguyen, L. N. Mai, J. J. Farrar, M. L. Hibberd, and C. P. Simmons. The Early Whole-Blood Transcriptional Signature of Dengue Virus and Features Associated with Progression to Dengue Shock Syndrome in Vietnamese Children and Young Adults. *Journal of Virology*, 84(24):12982–12994, 2010. ISSN 0022-538X. doi: 10.1128/JVI.01224-10. URL http://jvi.asm.org/cgi/doi/10.1128/JVI.01224-10.

https://clinicaltrials.gov/. ClinicalTrials. URL https://clinicaltrials.gov/.

https://drugs.ncats.io/. Inxight: Drugs. URL https://drugs.ncats.io/.

https://www.ema.europa.eu/en. European Medicines Agency. URL https://www.ema.europa.eu/en.

https://www.fda.gov/. Food and Drug Administration. URL https://www.fda.gov/.

Timothy R Hughes, Matthew J Marton, Allan R Jones, Christopher J Roberts, Roland Stoughton, Christopher D Armour, Holly A Bennett, Ernest Coffey, Hongyue Dai, Yudong D He, Matthew J Kidd, Amy M King, Michael R Meyer, David Slade, Pek Y Lum, Sergey B Stepaniants, Daniel D Shoemaker, Daniel Gachotte, Kalpana Chakraburtty, Julian Simon, Martin Bard, and Stephen H Friend. Functional Discovery via a Compendium of Expression Profiles. *Cell*, 102(1):109–126, jul 2000. ISSN 0092-8674. doi: 10.1016/S0092-8674(00)00015-5. URL https://www.sciencedirect.com/science/article/pii/S0092867400000155?via{%}3Dihub.

Tahir Hussain, Muhsin Jamal, Tayyab ur Rehman, and Saadia Andleeb. Dengue: pathogenesis, prevention and treatment – A mini review. *Advancements in Life Sciences*, 2(3):110–114, 2015.

Gareth Iacobucci. WHO recommends additional tests for Sanofi's dengue vaccine after safety concerns. *Bmj*, 1765(April):k1765, 2018. ISSN 0959-8138. doi: 10.1136/bmj.k1765. URL http://www.bmj.com/lookup/doi/10.1136/bmj.k1765.

Minoru Kanehisa, Yoko Sato, Miho Furumichi, Kanae Morishima, and Mao Tanabe. New approach for understanding genome variations in KEGG. *Nucleic Acids Research*, 47 (October 2018):590–595, 2018. ISSN 0305-1048. doi: 10.1093/nar/gky962. URL https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky962/5128935.

Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A. Shoemaker, Paul A. Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, and Evan E. Bolton. PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Research*, 47(D1):D1102–D1109, 2019. ISSN 13624962. doi: 10.1093/nar/gky1033.

Philip D. Kiser, Marcin Golczak, and Krzysztof Palczewski. Chemistry of the retinoid (visual) cycle. *Chemical Reviews*, 114(1):194–232, 2014. ISSN 00092665. doi: 10.1021/cr400107q.

Moritz U.G. Kraemer, Marianne E. Sinka, Kirsten A. Duda, Adrian Q.N. Mylne, Freya M. Shearer, Christopher M. Barker, Chester G. Moore, Roberta G. Carvalho, Giovanini E. Coelho, Wim Van Bortel, Guy Hendrickx, Francis Schaffner, Iqbal Rf Elyazar, Hwa Jen Teng, Oliver J. Brady, Jane P. Messina, David M. Pigott, Thomas W. Scott, David L. Smith, G. R. William Wint, Nick Golding, and Simon I. Hay. The global distribution of the arbovirus vectors Aedes aegypti and Ae. Albopictus. *eLife*, 4(JUNE2015):1–18, 2015. ISSN 2050084X. doi: 10.7554/eLife.08347.

Kimberly R Kukurba and Stephen B Montgomery. RNA Sequencing and Analysis. *Cold Spring Harbor protocols*, 2015(11):951–969, apr 2015. ISSN 1559-6095. doi: 10.1101/pdb.top084970. URL https://www.ncbi.nlm.nih.gov/pubmed/25870306https://www.ncbi.nlm.nih.gov/pmc/PMC4863231/.

Marcin Kwissa, Helder I. Nakaya, Nattawat Onlamoon, Jens Wrammert, Francois Villinger, Guey Chuen Perng, Sutee Yoksan, Kovit Pattanapanyasat, Kulkanya Chokephaibulkit, Rafi Ahmed, and Bali Pulendran. Dengue virus infection induces expansion of a CD14+CD16+monocyte population that stimulates plasmablast differentiation. *Cell Host and Microbe*, 16(1):115–127, 2014. ISSN 19346069. doi: 10.1016/j.chom.2014.06.001. URL http://dx.doi.org/10.1016/j.chom.2014.06.001.

Justin Lamb, Emily D. Crawford, David Peck, Joshua W. Modell, Irene C. Blat, Matthew J. Wrobel, Jim Lerner, Jean Philippe Brunet, Aravind Subramanian, Kenneth N. Ross, Michael Reich, Haley Hieronymus, Guo Wei, Scott A. Armstrong, Stephen J. Haggarty, Paul A. Clemons, Ru Wei, Steven A. Carr, Eric S. Lander, and Todd R. Golub. The connectivity map: Supplement. *Science*, 313(5795):1929–1935, 2006. ISSN 00368075. doi: 10.1126/science.1132939.

Ben Langmead. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics*, pages 1–24, 2011. doi: 10.1002/0471250953.bi1107s32.Aligning.

Richard M. Leggett, Ricardo H. Ramirez-Gonzalez, Bernardo J. Clavijo, Darren Waite, and Robert P. Davey. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers in Genetics*, 4(DEC):1–5, 2013. ISSN 16648021. doi: 10.3389/fgene.2013.00288.

Jing Liu-Helmersson, Mikkel Quam, Annelies Wilder-Smith, Hans Stenlund, Kristie Ebi, Eduardo Massad, and Joacim Rocklöv. Climate Change and Aedes Vectors: 21st Century Projections for Dengue Transmission in Europe. *EBioMedicine*, 7:267–277, 2016. ISSN 23523964. doi: 10.1016/j.ebiom.2016.03.046. URL http://dx.doi.org/10.1016/j.ebiom.2016.03.046.

William Loging, Lee Harland, and Bryn Williams-Jones. High-throughput electronic biology: Mining information for drug discovery. *Nature Reviews Drug Discovery*, 6(3):220–230, 2007. ISSN 14741776. doi: 10.1038/nrd2265.

P'ng Loke, Samantha N. Hammond, Jacqueline M. Leung, Charles C. Kim, Sajeev Batra, Crisanta Rocha, Angel Balmaseda, and Eva Harris. Gene expression patterns of dengue virus-infected children from nicaragua reveal a distinct signature of increased metabolism. *PLoS Neglected Tropical Diseases*, 4(6), 2010. ISSN 1935-2735. doi: 10.1371/journal.pntd.0000710.

Hoang Truong Long, Martin L Hibberd, Tran Tinh Hien, Nguyen Minh Dung, and Tran Van. Europe PMC Funders Group Patterns of Gene Transcript Abundance in the Blood of Children with Severe or Uncomplicated Dengue Highlight Differences in Disease Evolution and Host Response to Dengue Virus Infection. 199(4):537–546, 2015. doi: 10.1086/596507.Patterns.

Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):1–21, 2014. ISSN 1474760X. doi: 10.1186/s13059-014-0550-8.

Rohan Lowe, Neil Shirley, Mark Bleackley, Stephen Dolan, and Thomas Shafee. Transcriptomics technologies. *PLoS Computational Biology*, 13(5):1–23, 2017. ISSN 15537358. doi: 10.1371/journal.pcbi.1005457.

Gathsaurie Neelika Malavige, Ananda Wijewickrama, Samitha Fernando, Chandima Jeewandara, Anushka Ginneliya, Supun Samarasekara, Praveen Madushanka, Chameera Punchihewa, Shiran Paranavitane, Damayanthi Idampitiya, Chandanie Wanigatunga, Harsha Dissanayake, Shamini Prathapan, Laksiri Gomes, Siti A.B. Aman, Ashley St

John, and Graham S. Ogg. A preliminary study on efficacy of rupatadine for the treatment of acute dengue infection. *Scientific Reports*, 8(1):1–14, 2018. ISSN 20452322. doi: 10.1038/s41598-018-22285-x.

Byron E E Martina, Penelope Koraka, and Albert D M E Osterhaus. Dengue virus pathogenesis: An integrated view. *Clinical Microbiology Reviews*, 22(4):564–581, 2009. ISSN 08938512. doi: 10.1128/CMR.00035-09.

Marlén Martínez-Gutierrez, Jaime E. Castellanos, and Juan C. Gallego-Gómez. Statins reduce dengue virus production via decreased virion assembly. *Intervirology*, 54(4):202–216, 2011. ISSN 03005526. doi: 10.1159/000321892.

Marlen Martinez-Gutierrez, Luis A. Correa-Londoño, Jaime E. Castellanos, Juan C. Gallego-Gómez, and Jorge E. Osorio. Lovastatin delays infection and increases survival rates in AG129 mice infected with dengue virus serotype 2. *PLoS ONE*, 9(2), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0087412.

Anthony R. Mawson. Retinoids, race and the pathogenesis of dengue hemorrhagic fever. *Medical Hypotheses*, 81(6):1069–1074, 2013. ISSN 03069877. doi: 10.1016/j.mehy.2013.08.004. URL http://dx.doi.org/10.1016/j.mehy.2013.08.004.

Emma L. Mohr, Laura K. McMullan, Michael K. Lo, Jessica R. Spengler, Éric Bergeron, César G. Albariño, Punya Shrivastava-Ranjan, Cheng Feng Chiang, Stuart T. Nichol, Christina F. Spiropoulou, and Mike Flint. Inhibitors of cellular kinases with broad-spectrum antiviral activity for hemorrhagic fever viruses. *Antiviral Research*, 120:40–47, 2015. ISSN 18729096. doi: 10.1016/j.antiviral.2015.05.003. URL http://dx.doi.org/10.1016/j.antiviral.2015.05.003.

David Murphy, Pedro Reche, and Darren R Flower. Selection-based Design of in silico Dengue Epitope Ensemble Vaccines. *Chemical Biology & Drug Design*, (April 2018): 21–28, 2018. ISSN 17470277. doi: 10.1111/cbdd.13357. URL http://doi.wiley.com/10.1111/cbdd.13357.

Eduardo J.M. Nascimento, Ulisses Braga-Neto, Carlos E. Calzavara-Silva, Ana L.V. Gomes, Frederico G.C. Abath, Carlos A.A. Brito, Marli T. Cordeiro, Ana M. Silva, Cecilia Magalhães, Raoni Andrade, Laura H.V.G. Gil, and Ernesto T.A. Marques. Gene expression profiling during early acute febrile stage of dengue infection can predict the disease outcome. *PLoS ONE*, 4(11), 2009. ISSN 19326203. doi: 10.1371/journal.pone.0007892.

Marisa Oliveira, Worachart Lert-itthiporn, Bruno Cavadas, Verónica Fernandes, Ampaiwan Chuansumrit, Orlando Anunciação, Isabelle Casademont, Fanny Koeth, Marina Penova, Kanchana Tangnararatchakit, Chiea Chuen Khor, Richard Paul, Prida Malasit, Fumihiko

Matsuda, Etienne Simon-Lorière, Prapat Suriyaphol, Luisa Pereira, and Anavaj Sakuntabhai. Joint ancestry and association test indicate two distinct pathogenic pathways involved in classical dengue fever and dengue shock syndrome. *PLoS Neglected Tropical Diseases*, 12(2):1–20, 2018. ISSN 19352735. doi: 10.1371/journal.pntd.0006202.

Valery O. Polyanovsky, Mikhail A. Roytberg, and Vladimir G. Tumanyan. Comparative analysis of the quality of a global algorithm and a local algorithm for alignment of two sequences. *Algorithms for Molecular Biology*, 6(1):25, 2011. ISSN 17487188. doi: 10.1186/1748-7188-6-25. URL http://www.almob.org/content/6/1/25.

Stephen J. Popper, Aubree Gordon, Minghsun Liu, Angel Balmaseda, Eva Harris, and David A. Relman. Temporal Dynamics of the Transcriptional Response to Dengue Virus Infection in Nicaraguan Children. *PLoS Neglected Tropical Diseases*, 6(12):1–14, 2012. ISSN 19352727. doi: 10.1371/journal.pntd.0001966.

Senaka Rajapakse, Chaturaka Rodrigo, and Anoja Rajapakse. Treatment of dengue fever. *Infection and Drug Resistance*, 5(1):103–112, 2012. ISSN 11786973. doi: 10.2147/IDR.S22613.

Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47, 2015. ISSN 13624962. doi: 10.1093/nar/gkv007.

Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2009. ISSN 14602059. doi: 10.1093/bioinformatics/btp616.

Raul Rodriguez-Esteban and Xiaoyu Jiang. Differential gene expression in disease: A comparison between high-throughput studies and the literature. *BMC Medical Genomics*, 10 (1):1–10, 2017. ISSN 17558794. doi: 10.1186/s12920-017-0293-y.

Jayanta Samanta and Vishal Sharma. Dengue and its effects on liver. *World journal of clinical cases*, 3(2):125–131, feb 2015. ISSN 2307-8960. doi: 10.12998/wjcc.v3.i2.125. URL https://www.ncbi.nlm.nih.gov/pubmed/25685758https://www.ncbi.nlm.nih.gov/pmc/PMC4317605/.

October M. Sessions, Ying Tan, Kenneth C. Goh, Yujing Liu, Patrick Tan, Steve Rozen, and Eng Eong Ooi. Host Cell Transcriptome Profile during Wild-Type and Attenuated Dengue Virus Infection. *PLoS Neglected Tropical Diseases*, 7(3):1–12, 2013. ISSN 19352735. doi: 10.1371/journal.pntd.0002107.

Suzanne Moore Sheperd. Dengue, 2017. URL https://emedicine.medscape.com/article/215840-overview.

Beatriz Sierra, Petr Triska, Pedro Soares, Gissel Garcia, Ana B. Perez, Eglys Aguirre, Marisa Oliveira, Bruno Cavadas, Béatrice Regnault, Mayling Alvarez, Didye Ruiz, David C. Samuels, Anavaj Sakuntabhai, Luisa Pereira, and Maria G. Guzman. OSBPL10, RXRA and lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed CUBANS. *PLoS Pathogens*, 13(2):1–21, 2017. ISSN 15537374. doi: 10.1371/journal.ppat.1006220.

J. L. Smith, D. A. Stein, D. Shum, M. A. Fischer, C. Radu, B. Bhinder, H. Djaballah, J. A. Nelson, K. Fruh, and A. J. Hirsch. Inhibition of Dengue Virus Replication by a Class of Small-Molecule Compounds That Antagonize Dopamine Receptor D4 and Downstream Mitogen-Activated Protein Kinase Signaling. *Journal of Virology*, 88(10):5533–5542, 2014. ISSN 0022-538X. doi: 10.1128/jvi.00365-14.

Drs Sue Stehman, Mary C Smith, R.A. Sani, G.D. Gray, R.L. Baker, Peter M Schantz, Roy P E Yanong, I Pengenalan Tempat, World Health Organization, Norman Gratz, Sunaryo, Puji Astuti, Dyah Widiastuti, and GoI, MoHFW. A global brief on vector-borne diseases. *Symposium of Goat Health*, 11(September): 1–9, 2002. ISSN http://www.who.int/iris/handle/10665/79059. doi: WHO/DCO/WHD/2014.1. URL http://apps.who.int/iris/bitstream/10665/111008/1/WHO{_}DCO{_}WHD{_}2014.1{_}eng.pdf.

Jose A. Suaya, Donald S. Shepard, João B. Siqueira, Celina T. Martelli, Lucy C.S. Lum, Lian Huat Tan, Sukhontha Kongsin, Sukhum Jiamton, Fàtima Garrido, Romeo Montoya, Bias Armien, Rekol Huy, Leticia Castillo, Mariana Caram, Binod K. Sah, Rana Sughayyar, Karen R. Tyo, and Scott B. Halstead. Cost of dengue cases in eight countries in the Americas and asia: A prospective study. *American Journal of Tropical Medicine and Hygiene*, 80(5):846–855, 2009. ISSN 00029637. doi: 10.4269/ajtmh.2009.80.846.

Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, and Jill P Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550, oct 2005. ISSN 0027-8424 (Print). doi: 10.1073/pnas.0506580102.

Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu, Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, David L Lahr, Jodi E Hirschman, Zihan Liu, Melanie Donahue, Bina Julian, Mariya Khan, David Wadden, Ian C Smith, Daniel Lam, Arthur Liberzon, Courtney Toder, Mukta Bagul, Marek Orzechowski, Oana M Enache, Federica Piccioni, Sarah A Johnson, Nicholas J Lyons, Alice H Berger, Alykhan F Shamji, Angela N Brooks, Anita

Vrcic, Corey Flynn, Jacqueline Rosains, David Y Takeda, Roger Hu, Desiree Davison, Justin Lamb, Kristin Ardlie, Larson Hogstrom, Peyton Greenside, Nathanael S Gray, Paul A Clemons, Serena Silver, Xiaoyun Wu, Wen-Ning Zhao, Willis Read-Button, Xiaohua Wu, Stephen J Haggarty, Lucienne V Ronco, Jesse S Boehm, Stuart L Schreiber, John G Doench, Joshua A Bittker, David E Root, Bang Wong, and Todd R Golub. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, 171(6):1437–1452.e17, 2017. ISSN 1097-4172. doi: 10.1016/j.cell.2017.10.049. URL http://www.ncbi.nlm.nih.gov/pubmed/29195078{%}0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5990023.

Ulrike H. Taron, Moritz Lell, Axel Barlow, and Johanna L.A. Paijmans. Testing of alignment parameters for ancient samples: Evaluating and optimizing mapping parameters for ancient samples using the TAPAS tool. *Genes*, 9(3), 2018. ISSN 20734425. doi: 10.3390/genes9030157.

Hannah Tipney and Lawrence Hunter. An introduction to effective use of enrichment analysis software. *Human genomics*, 4(3):202–206, 2010. ISSN 1479-7364 (Electronic). doi: 10.1186/1479-7364-4-3-202.

Dinh The Trung, Le Thi Thu Thao, Tran Tinh Hien, Nguyen The Hung, Nguyen Ngoc Vinh, Pham Tran Dieu Hien, Nguyen Tran Chinh, Cameron Simmons, and Bridget Wills. Liver involvement associated with dengue infection in adults in Vietnam. *American Journal of Tropical Medicine and Hygiene*, 83(4):774–780, 2010. ISSN 00029637. doi: 10.4269/ajtmh.2010.10-0090.

Mathias Uhlen, Cheng Zhang, Sunjae Lee, Evelina Sjöstedt, Linn Fagerberg, Gholamreza Bidkhori, Rui Benfeitas, Muhammad Arif, Zhengtao Liu, Fredrik Edfors, Kemal Sanli, Kalle Von Feilitzen, Per Oksvold, Emma Lundberg, Sophia Hober, Peter Nilsson, Johanna Mattsson, Jochen M. Schwenk, Hans Brunnström, Bengt Glimelius, Tobias Sjöblom, Per Henrik Edqvist, Dijana Djureinovic, Patrick Micke, Cecilia Lindskog, Adil Mardinoglu, and Fredrik Ponten. A pathology atlas of the human cancer transcriptome. *Science*, 357(6352), 2017. ISSN 10959203. doi: 10.1126/science.aan2507.

Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1):57–63, jan 2009. ISSN 1471-0064. doi: 10.1038/nrg2484. URL https://www.ncbi.nlm.nih.gov/pubmed/19015660https://www.ncbi.nlm.nih.gov/pmc/PMC2949280/.

Kosala Gad Weerakoon, Senanayake Am Kularatne, Deepthika H. Edussuriya, Sarachchandra Ka Kodikara, Laxman Pg Gunatilake, Vasanti G. Pinto, Ashoka B. Seneviratne, and Sunethra Gunasena. Histopathological diagnosis of myocarditis in a dengue out-

break in Sri Lanka, 2009. *BMC Research Notes*, 4(1):268, 2011. ISSN 17560500. doi: 10.1186/1756-0500-4-268. URL http://www.biomedcentral.com/1756-0500/4/268.

James Whitehorn, Chau Van Vinh Nguyen, Lam Phung Khanh, Duong Thi Hue Kien, Nguyen Than Ha Quyen, Nguyen Thi Thanh Tran, Nguyen Thuy Hang, Nguyen Thanh Truong, Luong Thi Hue Tai, Nguyen Thi Cam Huong, Vo Thanh Nhon, Ta Van Tram, Jeremy Farrar, Marcel Wolbers, Cameron P. Simmons, and Bridget Wills. Lovastatin for the Treatment of Adult Patients with Dengue: A Randomized, Double-Blind, Placebo-Controlled Trial. *Clinical Infectious Diseases*, 62(4):468–476, 2015. ISSN 15376591. doi: 10.1093/cid/civ949.

Annelies Wilder-Smith, Murray, and Mikkel Quam. Epidemiology of dengue: past, present and future prospects. *Clinical Epidemiology*, 5:299, 2013. ISSN 1179-1349. doi: 10.2147/CLEP.S34440. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3753061/pdf/clep-5-299.pdf{%}0Ahttp://www.dovepress.com/epidemiology-of-dengue-past-present-and-future-prospects-peer-reviewed-article-CLEP.

David S. Wishart, Yannick D. Feunang, An C. Guo, Elvis J. Lo, Ana Marcu, Jason R. Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam MacIejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, DIana Le, Allison Pon, Craig Knox, and Michael Wilson. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1):D1074–D1082, 2018. ISSN 13624962. doi: 10.1093/nar/gkx1037.

World Health Organization. Dengue - WHO. URL https://www.who.int/denguecontrol/en/.

A

SUPPORT MATERIAL

A.1 SUPPLEMENTARY TABLES

Supplementary Table 1 (Excel file "Supplementary Tables 1-6")

Table with gene differential expression for the GSE18090, contains information regarding probe ID, adjusted p-value, p-value, logarithmic fold change, gene symbol and gene title. Genes with pink shade were considered as differentially expressed.

Supplementary Table 2 (Excel file "Supplementary Tables 1-6")

Table with gene differential expression for the GSE38246, contains information regarding probe ID, adjusted p-value, p-value, logarithmic fold change and open reading frame. Genes with pink shade were considered as differentially expressed.

Supplementary Table 3 (Excel file "Supplementary Tables 1-6")

Table with gene differential expression for the GSE51808, contains information regarding probe ID, adjusted p-value, p-value, logarithmic fold change, gene symbol and gene title. Genes with pink shade were considered as differentially expressed.

Supplementary Table 4 (Excel file "Supplementary Tables 1-6")

Table with gene differential expression for Spleen samples, contains information regarding gene ID, logarithmic fold change and p-value. Genes with pink shade were considered as differentially expressed.

Supplementary Table 5 (Excel file "Supplementary Tables 1-6")

Table with gene differential expression for Hepatic samples, contains information regarding gene ID, logarithmic fold change and p-value. Genes with pink shade were considered as differentially expressed.

Supplementary Table 6 (Excel file "Supplementary Tables 1-6")

Table with gene differential expression for Encephalon samples, contains information regarding gene ID, logarithmic fold change and p-value. Genes with pink shade were considered as differentially expressed.

Supplementary Table 7 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the KEGG annotations in GSE18090. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 8 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the CC annotations in GSE18090. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 9 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the MF annotations in GSE18090. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 10 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the BP annotations in GSE18090. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 11 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the KEGG annotations in GSE51808. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 12 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the CC annotations in GSE51808. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 13 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the MF annotations in GSE51808. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 14 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the BP annotations in GSE51808. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 15 (Excel file "Supplementary Tables 7-30")

Table with results from the enrichment analysis performed on GSEA for the KEGG annotations in GSE38246. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 16 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the CC annotations in GSE38246. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 17 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the MF annotations in GSE38246. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 18 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the BP annotations in GSE38246. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 19 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the KEGG annotations in Hepatic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 20 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the CC annotations in Hepatic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 21 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the MF annotations in Hepatic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 22 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the BP annotations in Hepatic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 23 (Excel file "Supplementary Tables 7-30")

Table with results from the enrichment analysis performed on GSEA for the KEGG annotations in Spleen samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 24 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the CC annotations in Spleen samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 25 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the MF annotations in Spleen samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 26 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the BP annotations in Spleen samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 27 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the KEGG annotations in Encephalic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 28 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the CC annotations in Encephalic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 29 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the MF annotations in Encephalic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 30 (Excel file "Supplementary Tables 7-30")
Table with results from the enrichment analysis performed on GSEA for the BP annotations in Encephalic samples. Table has information on the enriched gene sets as well as the statistics on each of the annotations.

Supplementary Table 31 (Excel file "Supplementary Tables 31-37")

Table with results from the drug discovery step, performed on CMap for the microarray data. Table contains score, name of the drug and mechanism of action for the outputted drugs.

Supplementary Table 32 (Excel file "Supplementary Tables 31-37")
Table with results from the drug discovery step, performed on CMap for the hepatic data. Table contains score, name of the drug and mechanism of action for the outputted drugs.

Supplementary Table 33 (Excel file "Supplementary Tables 31-37")
Table with results from the drug discovery step, performed on CMap for the hepatic data limited to the hepatic cell line on CMap. Table contains score, name of the drug and mechanism of action for the outputted drugs.

Supplementary Table 34 (Excel file "Supplementary Tables 31-37")
Table with results from the drug discovery step, performed on CMap for the spleen data. Table contains score, name of the drug and mechanism of action for the outputted drugs.

Supplementary Table 35 (Excel file "Supplementary Tables 31-37")
Table with results from the drug discovery step, performed on CMap for the spleen data limited to the hepatic cell line. Table contains score, name of the drug and mechanism of action for the outputted drugs.

Supplementary Table 36 (Excel file "Supplementary Tables 31-37")
Table with results from the drug discovery step, performed on CMap for the encephalon data. Table contains score, name of the drug and mechanism of action for the outputted drugs.

Supplementary Table 37 (Excel file "Supplementary Tables 31-37")
Table with results from the drug discovery step, performed on CMap for the encephalon data limited to the hepatic cell line. Table contains score, name of the drug and mechanism of action for the outputted drugs.

A.2 SUPPLEMENTARY FIGURES

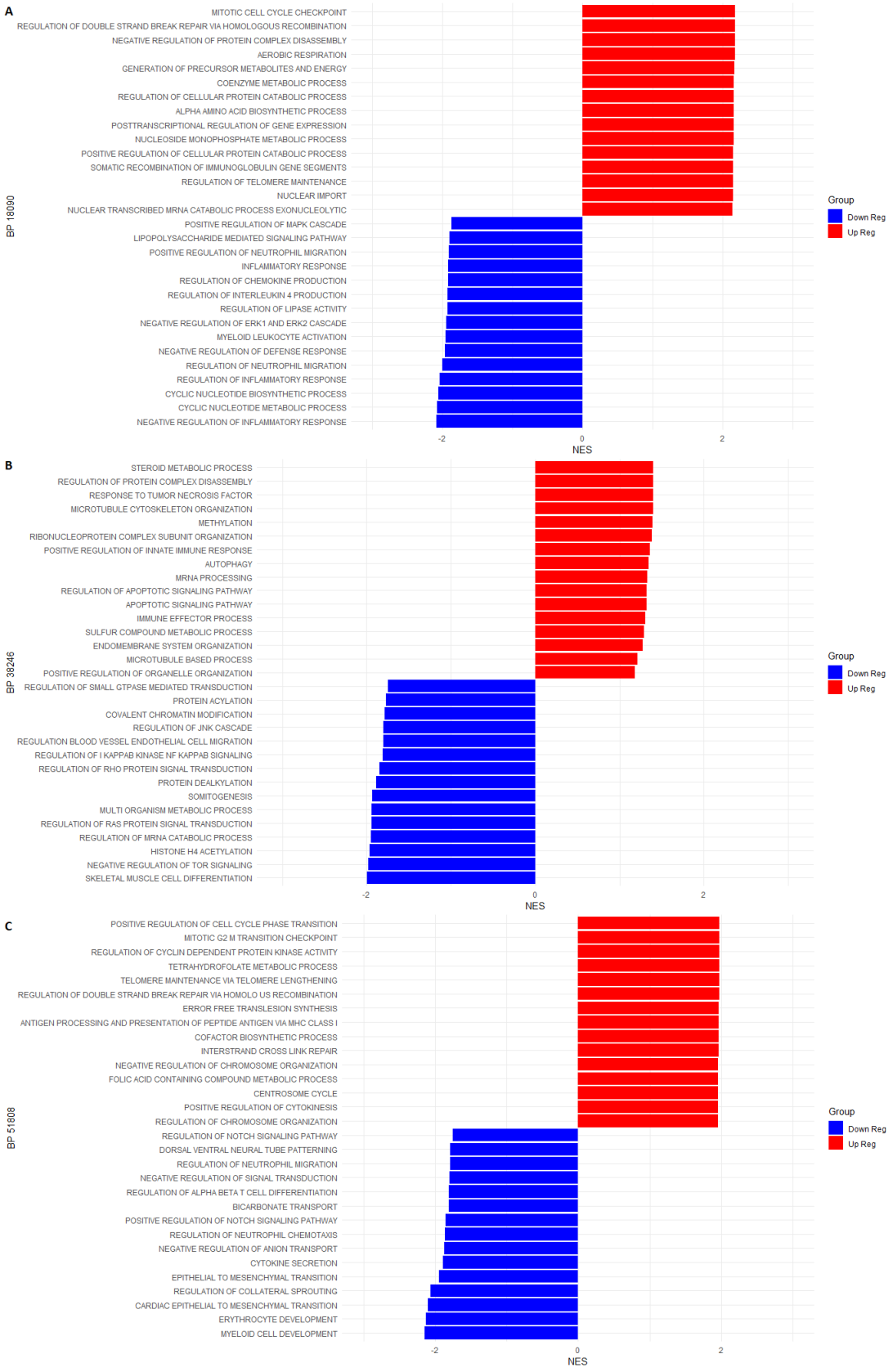


Figure S1.: Top BP pathways (cutoff value; -log p $\geq$ 2) in the (A) GSE18090, (B) GSE38246 and (C) GSE51808. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls, while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.

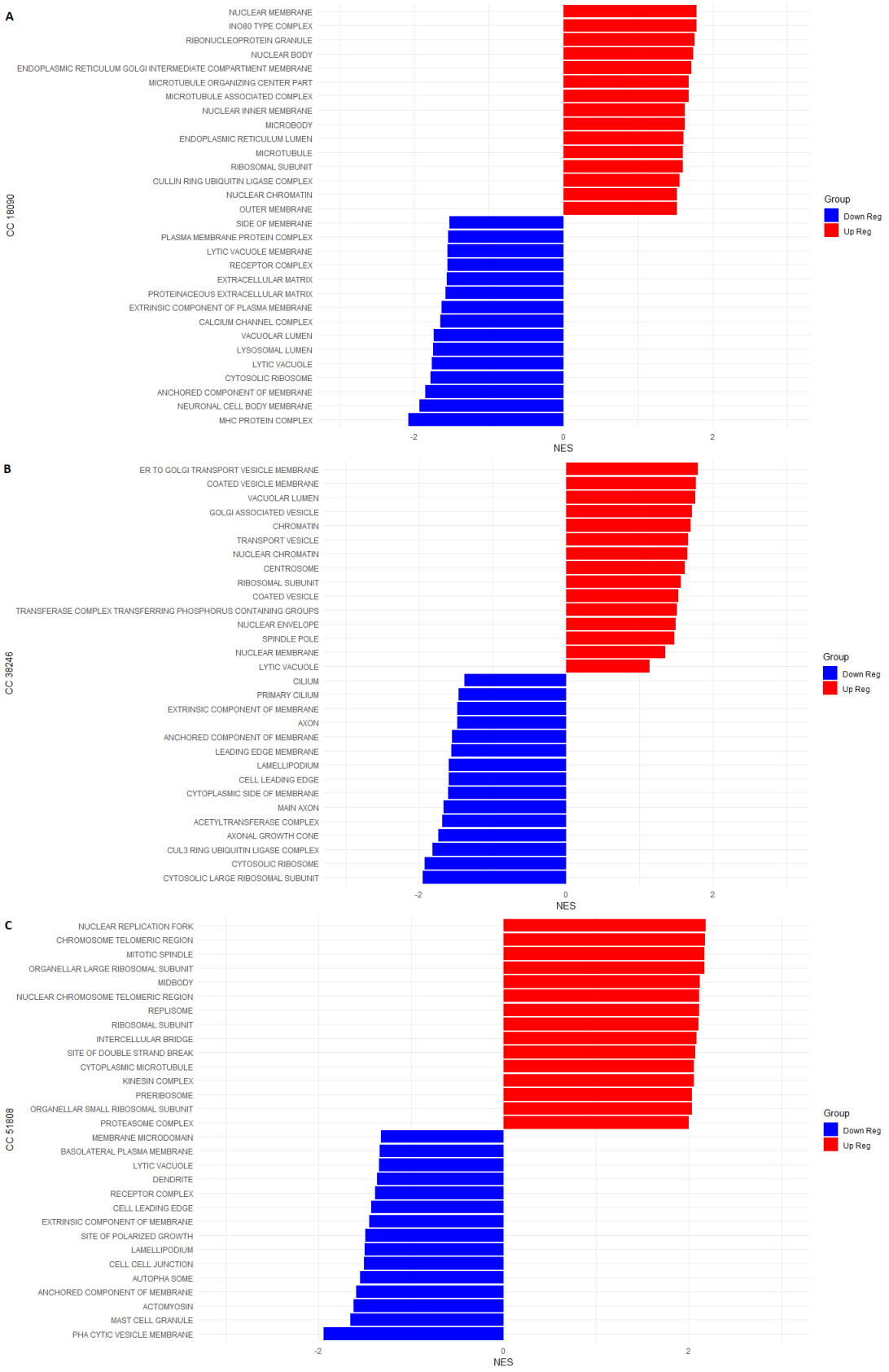


Figure S2.: Top CC pathways (cutoff value; -log p $\geq$ 2) in the (A) GSE18090, (B) GSE38246 and (C) GSE51808. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls, while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.

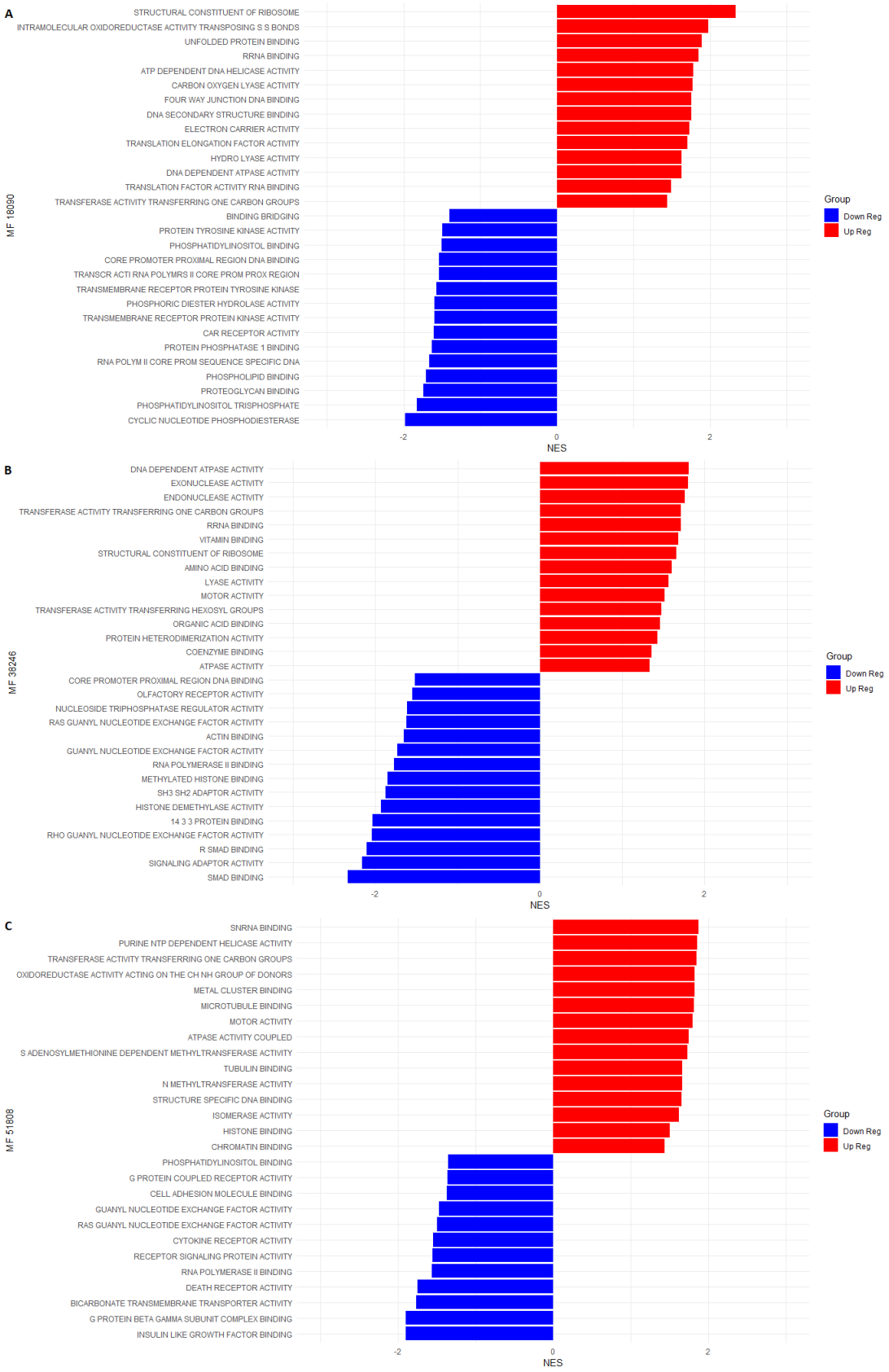


Figure S3.: Top MF pathways (cutoff value; -log p $\geq$ 2) in the (A) GSE18090, (B) GSE38246 and (C) GSE51808. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls, while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.

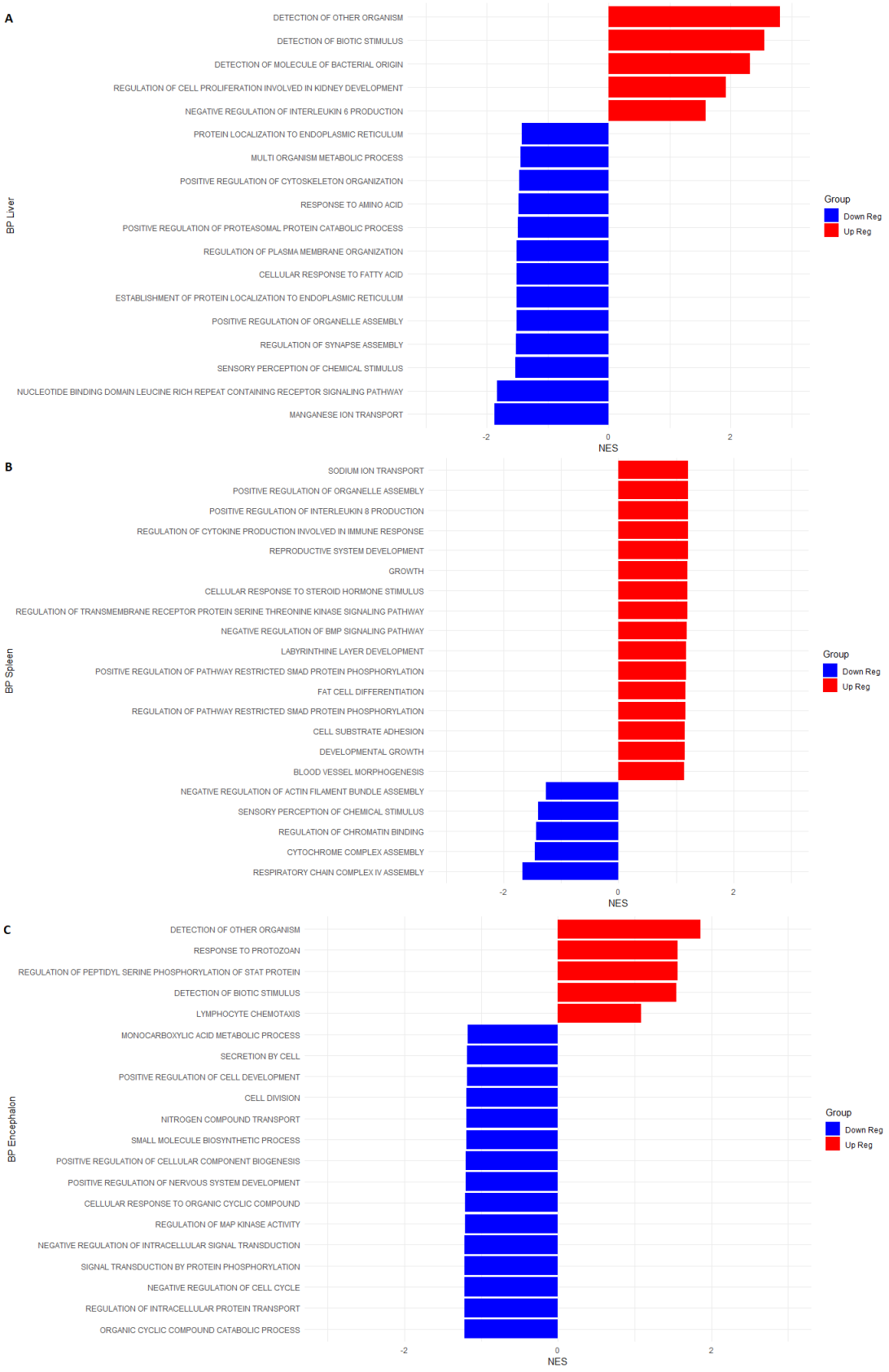


Figure S4.: Top BP pathways (cutoff value; -log p $\geq$ 2) in the (A) Liver, (B) Spleen and (C) Encephalon. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls, while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.

Figure S5.: Top CC pathways (cutoff value; -log p $\geq$ 2) in the (A) Liver, (B) Spleen and (C) Encephalon. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls, while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.

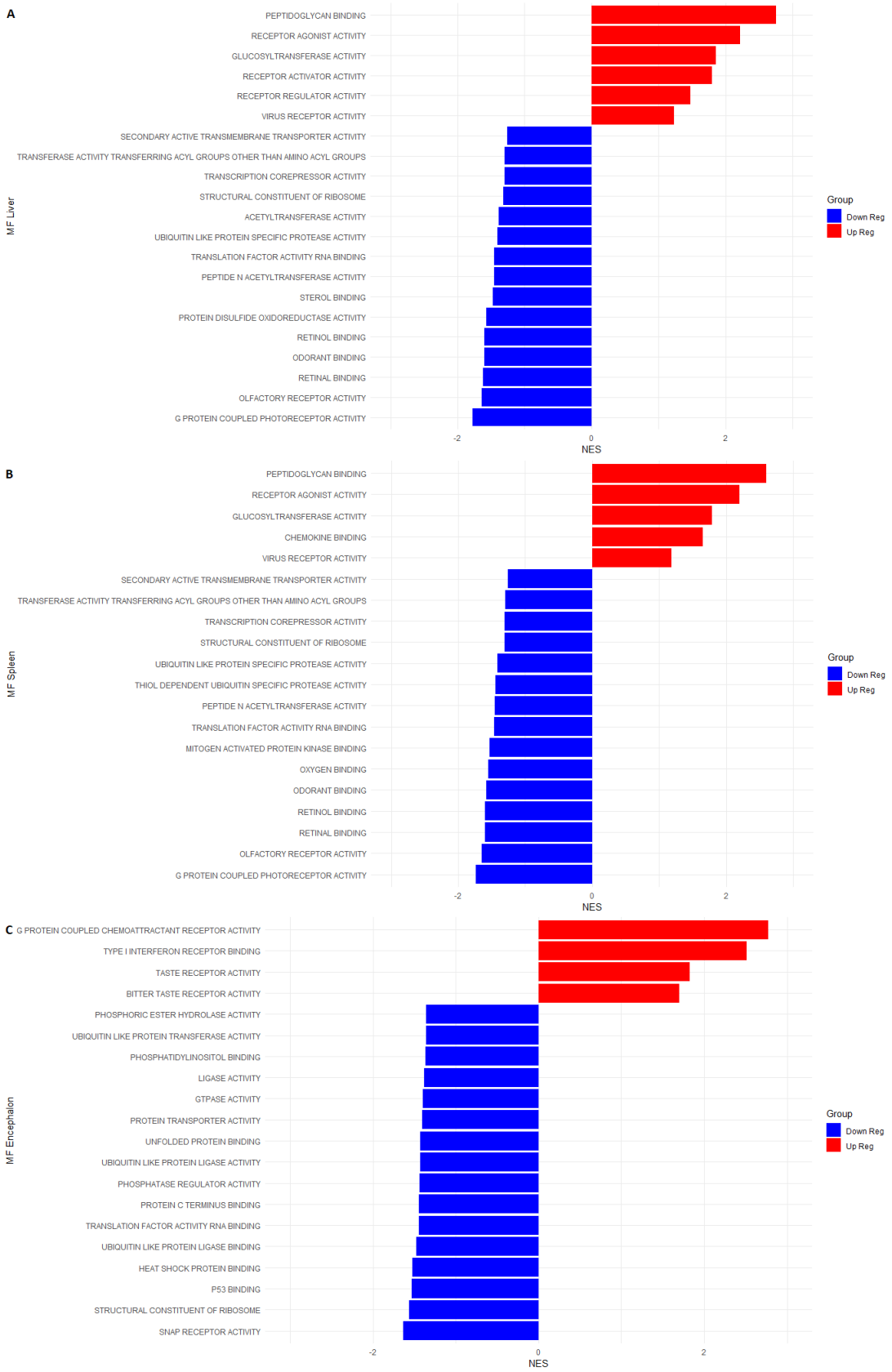


Figure S6.: Top MF pathways (cutoff value; -log p $\geq$ 2) in the (A) Liver, (B) Spleen and (C) Encephalon. Positive NES (in red) mean upregulated pathways in the infected individuals versus controls, while negative NES values (in blue) mean downregulated pathways in the infected individuals versus controls.